



# **FUNCTIONAL AND COMPARATIVE GENOMICS OF SACCHAROMYCES AND NON-SACCHAROMYCES YEASTS: POTENTIAL FOR INDUSTRIAL AND FOOD BIOTECHNOLOGY**

EDITED BY: Isabel Sá-Correia and Ed Louis

PUBLISHED IN: *Frontiers in Genetics* and *Frontiers in Microbiology*



# frontiers

## Frontiers eBook Copyright Statement

The copyright in the text of individual articles in this eBook is the property of their respective authors or their respective institutions or funders. The copyright in graphics and images within each article may be subject to copyright of other parties. In both cases this is subject to a license granted to Frontiers.

The compilation of articles constituting this eBook is the property of Frontiers.

Each article within this eBook, and the eBook itself, are published under the most recent version of the Creative Commons CC-BY licence.

The version current at the date of publication of this eBook is CC-BY 4.0. If the CC-BY licence is updated, the licence granted by Frontiers is automatically updated to the new version.

When exercising any right under the CC-BY licence, Frontiers must be attributed as the original publisher of the article or eBook, as applicable.

Authors have the responsibility of ensuring that any graphics or other materials which are the property of others may be included in the CC-BY licence, but this should be checked before relying on the CC-BY licence to reproduce those materials. Any copyright notices relating to those materials must be complied with.

Copyright and source acknowledgement notices may not be removed and must be displayed in any copy, derivative work or partial copy which includes the elements in question.

All copyright, and all rights therein, are protected by national and international copyright laws. The above represents a summary only. For further information please read Frontiers' Conditions for Website Use and Copyright Statement, and the applicable CC-BY licence.

ISSN 1664-8714

ISBN 978-2-88963-522-1

DOI 10.3389/978-2-88963-522-1

## About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

## Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

## Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.

Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

## What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: [researchtopics@frontiersin.org](mailto:researchtopics@frontiersin.org)



# FUNCTIONAL AND COMPARATIVE GENOMICS OF SACCHAROMYCES AND NON-SACCHAROMYCES YEASTS: POTENTIAL FOR INDUSTRIAL AND FOOD BIOTECHNOLOGY

Topic Editors:

**Isabel Sá-Correia**, Instituto Superior Técnico, Universidade de Lisboa, Portugal  
**Ed Louis**, University of Leicester, United Kingdom

Since 1996, when the first *Saccharomyces cerevisiae* genome sequence was released, a wealth of genomic data has been made available for numerous *S. cerevisiae* strains, its close relatives, and non-conventional yeast species isolates of diverse origins. Several annotated genomes of interspecific hybrids, both within the *Saccharomyces* clade and outside, are now also available. This genomic information, together with functional genomics and genome engineering tools, is providing a holistic assessment of the complex cellular responses to environmental challenges, elucidating the processes underlying evolution, speciation, hybridization, domestication, and uncovering crucial aspects of yeasts' physiological genomics to guide their biotechnological exploitation.

*S. cerevisiae* has been used for millennia in the production of food and beverages and research over the last century and a half has generated a great deal of knowledge of this species. Despite all this, *S. cerevisiae* is not the best for all uses and many non-conventional yeast species have highly desirable traits that *S. cerevisiae* does not have. These include tolerance to different stresses (e.g. acetic acid tolerance in *Zygosaccharomyces bailii*, osmotolerance in *Z. rouxii*, and thermotolerance in *Kluyveromyces marxianus* and *Ogataea (Hansenula) polymorpha*), the capacity of assimilation of diverse carbon sources (e.g. high native capacity to metabolize xylose and potential for the valorization of agroforest residues by *Scheffersomyces (Pichia) stipites*), as well as, high protein secretion, fermentation efficiency and production of desirable flavors, capacity to favor respiration over fermentation, high lipid biosynthesis and accumulation, and efficient production of chemicals other than ethanol amongst many. Several non-*Saccharomyces* species have already been developed as eukaryotic hosts and cell factories. Others are highly relevant as food spoilers or for desirable flavor producers. Therefore, non-conventional yeasts are now attracting increasing attention with their diversity and complexity being tackled by basic research for biotechnological applications.

The interest in the exploitation of non-conventional yeasts is very high and a number of tools, such as cloning vectors, promoters, terminators, and efficient genome editing tools, have been developed to facilitate their genetic engineering. Functional and Comparative Genomics of non-conventional yeasts is elucidating the evolution of genome functions and metabolic and ecological diversity, relating their physiology to genomic features and opening the door to the application of metabolic engineering and synthetic biology to yeasts of biotechnological potential. We are entering the era of the non-conventional yeasts, increasing the exploitation of yeast biodiversity and metabolic capabilities in science and industry. In this collection the industrial properties of *S. cerevisiae*, in particular uses, are explored along with its

closely related species and interspecific hybrids. This is followed by comparisons between *S. cerevisiae* and non-conventional yeasts in specific applications and then the properties of various non-conventional yeasts and their hybrids.

**Citation:** Sá-Correia, I., Louis, E., eds. (2020). Functional and Comparative Genomics of *Saccharomyces* and non-*Saccharomyces* Yeasts: Potential for Industrial and Food Biotechnology. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88963-522-1

# Table of Contents

## S. CEREVISIAE

- 06** *Whole-Genome Analysis of Three Yeast Strains Used for Production of Sherry-Like Wines Revealed Genetic Traits Specific to Flor Yeasts*  
Mikhail A. Eldarov, Alexey V. Beletsky, Tatiana N. Tanashchuk, Svetlana A. Kishkovskaya, Nikolai V. Ravin and Andrey V. Mardanov
- 19** *Traditional Norwegian Kveik are a Genetically Distinct Group of Domesticated *Saccharomyces cerevisiae* Brewing Yeasts*  
Richard Preiss, Caroline Tyrawa, Kristoffer Krogerus, Lars Marius Garshol and George van der Merwe
- 37** *Aneuploidy and Ethanol Tolerance in *Saccharomyces cerevisiae**  
Miguel Morard, Laura G. Macías, Ana C. Adam, María Lairón-Peris, Roberto Pérez-Torrado, Christina Toft and Eladio Barrio
- 49** *Characterization of a New *Saccharomyces cerevisiae* Isolated From Hibiscus Flower and Its Mutant With L-Leucine Accumulation for Awamori Brewing*  
Takayuki Abe, Yoichi Toyokawa, Yukiko Sugimoto, Haruna Azuma, Keiko Tsukahara, Ryo Nasuno, Daisuke Watanabe, Masatoshi Tsukahara and Hiroshi Takagi
- 58** *Controlled Reduction of Genomic Heterozygosity in an Industrial Yeast Strain Reveals Wide Cryptic Phenotypic Variation*  
Nadia M. V. Sampaio, Ruth A. Watson and Juan Lucas Argueso
- 72** *Quantitative Trait Nucleotides Impacting the Technological Performances of Industrial *Saccharomyces cerevisiae* Strains*  
Emilien Peltier, Anne Friedrich, Joseph Schacherer and Philippe Marullo

## S. CEREVISIAE AND RELATIVES AND HYBRIDS

- 96** *A Unique *Saccharomyces cerevisiae* × *Saccharomyces uvarum* Hybrid Isolated From Norwegian Farmhouse Beer: Characterization and Reconstruction*  
Kristoffer Krogerus, Richard Preiss and Brian Gibson
- 111** *The Paralogous Genes PDR18 and SNQ2, Encoding Multidrug Resistance ABC Transporters, Derive From a Recent Duplication Event, PDR18 Being Specific to the *Saccharomyces* Genus*  
Cláudia P. Godinho, Paulo J. Dias, Elise Ponçot and Isabel Sá-Correia
- 128** *The Genetic Makeup and Expression of the Glycolytic and Fermentative Pathways are Highly Conserved Within the *Saccharomyces* Genus*  
Francine J. Boonekamp, Sofia Dashko, Marcel van den Broek, Thies Gehrman, Jean-Marc Daran and Pascale Daran-Lapujade
- 145** *Comparative Genomics Between *Saccharomyces kudriavzevii* and *S. cerevisiae* Applied to Identify Mechanisms Involved in Adaptation*  
Laura G. Macías, Miguel Morard, Christina Toft and Eladio Barrio
- 156** *A Quasi-Domestic Relic Hybrid Population of *Saccharomyces cerevisiae* × *S. paradoxus* Adapted to Olive Brine*  
Ana Pontes, Neža Čadež, Paula Gonçalves and José Paulo Sampaio

## S. CEREVISIAE AND COMPARISONS

- 170** *Adaptive Response and Tolerance to Acetic Acid in Saccharomyces cerevisiae and Zygosaccharomyces bailii: A Physiological Genomics Perspective*

Margarida Palma, Joana F. Guerreiro and Isabel Sá-Correia

- 186** *Analysis of the NCR Mechanisms in Hanseniaspora vineae and Saccharomyces cerevisiae During Winemaking*

Jessica Lleixà, Valentina Martín, Facundo Giorrello, Maria C. Portillo, Francisco Carrau, Gemma Beltran and Albert Mas

## OTHER SPECIES AND HYBRIDS

- 195** *Expansion of a Telomeric FLO/ALS-Like Sequence Gene Family in Saccharomycopsis fermentans*

Beatrice Bernardi, Yeseren Kayacan and Jürgen Wendland

- 208** *Multi-Omics Analysis of Fatty Alcohol Production in Engineered Yeasts Saccharomyces cerevisiae and Yarrowia lipolytica*

Jonathan Dahlin, Carina Holkenbrink, Eko Roy Marella, Guokun Wang, Ulf Liebal, Christian Lieven, Dieter Weber, Douglas McCloskey, Birgitta E. Ebert, Markus J. Herrgård, Lars Mathias Blank and Irina Borodina

- 223** *Ploidy Variation in Kluyveromyces marxianus Separates Dairy and Non-dairy Isolates*

Raúl A. Ortiz-Merino, Javier A. Varela, Aisling Y. Coughlan, Hisashi Hoshida, Wendel B. da Silveira, Caroline Wilde, Niels G. A. Kuijpers, Jan-Maarten Geertman, Kenneth H. Wolfe and John P. Morrissey

- 239** *Interplay of Chimeric Mating-Type Loci Impairs Fertility Rescue and Accounts for Intra-Strain Variability in Zygosaccharomyces rouxii Interspecies Hybrid ATCC42981*

Melissa Bizzarri, Stefano Cassanelli, Laura Bartolini, Leszek P. Pryszcz, Michala Dušková, Hana Sychrová and Lisa Solieri



# Whole-Genome Analysis of Three Yeast Strains Used for Production of Sherry-Like Wines Revealed Genetic Traits Specific to Flor Yeasts

Mikhail A. Eldarov<sup>1</sup>, Alexey V. Beletsky<sup>1</sup>, Tatiana N. Tanashchuk<sup>2</sup>,  
Svetlana A. Kishkovskaya<sup>2</sup>, Nikolai V. Ravin<sup>1</sup> and Andrey V. Mardanov<sup>1\*</sup>

<sup>1</sup> Institute of Bioengineering, Research Center of Biotechnology of the Russian Academy of Sciences, Moscow, Russia,

<sup>2</sup> All-Russian National Research Institute of Viticulture and Winemaking "Magarach" of the Russian Academy of Sciences, Yalta, Russia

## OPEN ACCESS

### Edited by:

Ed Louis,  
University of Leicester,  
United Kingdom

### Reviewed by:

Liti Gianni,  
Institute of Research on Cancer  
and Aging in Nice, France  
Nuno Pereira Mira,  
Instituto de Bioengenharia e  
Biotecnologias, Portugal

### \*Correspondence:

Andrey V. Mardanov  
mardanov@biengi.ac.ru

### Specialty section:

This article was submitted to  
Evolutionary and Genomic  
Microbiology,  
a section of the journal  
Frontiers in Microbiology

Received: 30 November 2017

Accepted: 25 April 2018

Published: 15 May 2018

### Citation:

Eldarov MA, Beletsky AV,  
Tanashchuk TN, Kishkovskaya SA,  
Ravin NV and Mardanov AV (2018)  
Whole-Genome Analysis of Three  
Yeast Strains Used for Production  
of Sherry-Like Wines Revealed  
Genetic Traits Specific to Flor Yeasts.  
Front. Microbiol. 9:965.  
doi: 10.3389/fmicb.2018.00965

Flor yeast strains represent a specialized group of *Saccharomyces cerevisiae* yeasts used for biological wine aging. We have sequenced the genomes of three flor strains originated from different geographic regions and used for production of sherry-like wines in Russia. According to the obtained phylogeny of 118 yeast strains, flor strains form very tight cluster adjacent to the main wine clade. SNP analysis versus available genomes of wine and flor strains revealed 2,270 genetic variants in 1,337 loci specific to flor strains. Gene ontology analysis in combination with gene content evaluation revealed a complex landscape of possibly adaptive genetic changes in flor yeast, related to genes associated with cell morphology, mitotic cell cycle, ion homeostasis, DNA repair, carbohydrate metabolism, lipid metabolism, and cell wall biogenesis. Pangenomic analysis discovered the presence of several well-known "non-reference" loci of potential industrial importance. Events of gene loss included deletions of asparaginase genes, maltose utilization locus, and *FRE-FIT* locus involved in iron transport. The latter in combination with a flor-yeast-specific mutation in the Aft1 transcription factor gene is likely to be responsible for the discovered phenotype of increased iron sensitivity and improved iron uptake of analyzed strains. Expansion of the coding region of the *FLO11* flocullin gene and alteration of the balance between members of the *FLO* gene family are likely to positively affect the well-known propensity of flor strains for velum formation. Our study provides new insights in the nature of genetic variation in flor yeast strains and demonstrates that different adaptive properties of flor yeast strains could have evolved through different mechanisms of genetic variation.

**Keywords:** *Saccharomyces cerevisiae*, flor yeast, sherry, genetic diversity, comparative genomics, biofilm, SNP

## INTRODUCTION

Flor yeast strains represent a specialized group of yeasts used for centuries in various countries for biological wine aging (Alexandre, 2013; Legras et al., 2016). The physiological and biochemical properties of flor yeast strains associated with their application in specific winemaking processes are quite distinct from wine starter yeast strains and are relevant to the technological peculiarities of sherry-type wine formation (reviewed in Alexandre, 2013; Eldarov et al., 2016). One of the key



prominent features of flor yeast is their capability to form a biofilm on the surface of fortified wine (Martínez et al., 1997). This ability to float is critical for flor yeast metabolic changes associated with conditions of biological wine aging and their resistance to harsh winemaking conditions. In the course of sherry wine formation, wine composition changes force flor yeasts to shift their metabolism toward oxidation of non-fermentable carbon sources leading to important changes in wine chemical composition and production of specific aromatic and flavor compounds (Peinado and Mauricio, 2009). Stressful conditions of sherry-wine formation include elevated ethanol and acetaldehyde concentration, increased oxidative damage, poor nitrogen sources, etc. Velum formation by flor yeast is generally considered as an adaptive mechanism ensuring oxygen access and resistance to harsh environmental conditions.

Taxonomic studies showed that yeast present in the velum on the surface of French and Spain sherry wines predominantly belong to *Saccharomyces cerevisiae* (Charpentier et al., 2009). They differ from wine yeast by the presence of specific 24 bp deletion or C insertion in the ITS1 region (Charpentier et al., 2009). Many flor yeasts also possess a specific deletion in the promoter of *FLO11* gene – a key cell-surface adhesin responsible for yeast cell aggregation and biofilm formation (Fidalgo et al., 2006; Voordeckers et al., 2012; Holmes et al., 2013; Legras et al., 2014). This deletion, affecting the ICR1 non-coding RNA and stimulating *FLO11* transcription, is frequent in Spanish, Italian, Hungarian, and French flor strains (Legras et al., 2014). There is a significant degree of strain variation in *FLO11*-dependent phenotypes, resulting both from variations in *FLO11* promoter and coding sequences, and *FLO11* mRNA levels (Zara et al., 2009; Barrales et al., 2012; Barua et al., 2016). Increase of the gene length is another type of *FLO11* polymorphism leading to enhancing hydrophobicity of respective yeast strains (Fidalgo et al., 2008).

These observations, however, touched only limited aspects of the specific traits of flor yeast strains, that, as other quantitative traits, are by no doubt determined by coordinated genetic and gene expression changes of numerous genes involved in cell–cell adhesion, stress resistance, nitrogen and carbon and lipid metabolism, production of aromatic compounds, etc. (Rossignol et al., 2003; Walker et al., 2014). The identification of genomic and proteomic changes specific to flor yeast was the subject of several recent studies. Microsatellite genotyping of flor yeast strains isolated in France, Italy, Spain, and Hungary have shown that most strains belong to the same genetic group (Charpentier et al., 2009). Using comparative genome hybridization, it was shown that flor strains are mostly diploid and do not have large segmental amplifications (Legras et al., 2014). Several papers report comprehensive proteome analysis of a flor yeast with regard to detecting proteins related to carbon uptake, TCA cycle, cell wall biosynthesis, mitochondrial function, and metabolism of glycerol, ethanol, and aromatic compounds (Moreno-García et al., 2015, 2017).

Due to enormous progress in next generation sequencing (NGS) methods, comparative genomics became a powerful instrument to study the origin, diversity, population structure, and natural history of *S. cerevisiae* and related yeast (Marsit and Dequin, 2015; Borneman et al., 2016; Gallone et al., 2016).

Sequencing of wine yeast genomes is the main contemporary tool to elucidate the nature of causative genetic differences underpinning the observed phenotypic variation of yeast strains, to compare the molecular genetic data with industrial characteristics of yeast strains, to study the mechanisms of yeast genome evolution under conditions of artificial selection (Bergström et al., 2014).

In a recent comparative genomic study numerous genomic loci, differentiating wine and flor yeast have been identified and phylogenetic origin of flor yeast was revealed (Coi et al., 2017). Many candidate genomic regions and regulatory networks responsible for adaptation to biological aging conditions were thus identified, providing evidence for adaptive evolution of flor yeast as a result of domestication. Importantly, genomic data confirmed that flor yeast represents a unique lineage that emerged from the wine clade through a relatively recent bottleneck event (Charpentier et al., 2009; Coi et al., 2017). Thus, a comprehensive set of statistic and genetic methods could be applied to search for genomic signatures indicating possible positive selection. Dozens of candidate genes with potentially impacting substitutions were identified, including those important for pseudohyphal growth (*IRA1*, *SFG1*, *HMS2*, *IME4*, *FLO11*, and *RGA2*) carbon metabolism (*HXT3*, *HXT6,7*, and *MDH2*), response to osmotic stress (*SLN1* and *SFL1*), zinc ion transport (*ZRT1*), and other processes and functions (Coi et al., 2017). The phenotypic relevance of several of identified alleles for flor yeast physiology was demonstrated using previously developed set of haploid flor strains (Coi et al., 2016).

Here, we describe the genome sequencing and comparative genomic analysis of the three *S. cerevisiae* strains used for the industrial production of sherry-type wines in Russia. We describe gene content, structural rearrangements, events of gene loss, and contribution of “non-reference” genomic material to genomic makeup of analyzed strains. By combining SNP data for our strains with those from Genowine project (BioProject PRJEB6529), we identified additional genomic regions possibly affected by positive selection. Corresponding genes with flor yeast specific alleles encode proteins involved in cell adhesion, DNA repair, carbohydrate metabolism, ion homeostasis, response to osmotic stress, lipid metabolism, cell wall biogenesis, etc. Preliminary phenotypic analysis of affected genomic loci involved in iron metabolism is provided.

## MATERIALS AND METHODS

### Strains and Reference Sequences

Three flor yeast strains from the Magarach Collection of Microorganisms for Winemaking (Research Institute of Viticulture and Winemaking of the Russian Academy of Sciences) were used for genome sequencing: I-30, I-329, and I-566 (Kishkovskaia et al., 2017). The strains are available from the authors. The R64 2-1 release of the reference *S. cerevisiae* S288c genome was downloaded from Saccharomyces Genome Database (SGD)<sup>1</sup> and used as reference throughout this work.

<sup>1</sup>[https://downloads.yeastgenome.org/sequence/S288C\\_reference/](https://downloads.yeastgenome.org/sequence/S288C_reference/)

The list of strains used for comparative genomic analysis is provided in Supplementary Table S1.

## DNA Isolation, Genome Sequencing, and Assembly

Cells from frozen glycerol stocks were grown on YPD plates at room temperature. Single colony was grown in 50 ml YPD at 20°C for 24 h, and cells were collected, washed in TE, and freeze-dried. Genomic DNA was prepared from freeze-dried cells with CTAB extraction method (Sreenivasaprasad, 2000) and further column purified with QIAGEN Genomic-tip 500/G kit. Final DNA concentrations were measured using Qubit Quant-iT dsDNA HS Assay kit (Thermo Fisher Scientific, United States).

The genome sequence of *S. cerevisiae* I-566 was obtained using Illumina HiSeq2500 technology. The sequencing of a TrueSeq DNA library generated 14,221,481 single-end reads (250 nt). Sequencing primers were removed using Cutadapt (Martin, 2011) and low-quality read regions were trimmed using Sickle<sup>2</sup>. Illumina reads were *de novo* assembled using SPAdes 3.7.1 (Bankevich et al., 2012). Contigs shorter than 200 bp were discarded.

Genomes of two other strains were obtained using a combination of Illumina HiSeq2500 and PacBio RSII technologies. Strains I-30 and I-329 were sequenced using PacBio P6C4 chemistry using eight and nine SMRT cells, respectively. A total of 122,857 and 191,070 reads with an average length of 5,596 and 3,655 bp were obtained. In addition, 14,185,876 and 13,371,670 single-end reads (250 nt) were obtained upon sequencing of a TrueSeq DNA libraries using Illumina HiSeq2500. A hybrid Illumina and PacBio assembly was done using SPAdes 3.7.1 (Bankevich et al., 2012).

Protein-coding genes were predicted using Augustus 3.0.3 (Stanke and Morgenstern, 2005) trained on *S. cerevisiae* S288C dataset. Annotation of protein-coding genes was performed using BLASTP search against *S. cerevisiae* S288C proteins and a non-redundant protein sequence database. tRNA genes were predicted using tRNAscan-SE (Lowe and Chan, 2016), and rRNA genes were identified by BLASTN search against S288C rRNA genes.

For comparative genomic analysis, we also used Illumina reads previously obtained for 21 flor and wine yeast strains (Supplementary Table S1). Illumina reads were downloaded from Sequencing reads archive database and then *de novo* assembled into contigs using SPAdes 3.7.1 (Bankevich et al., 2012). Contigs shorter than 200 bp were discarded.

## Variation Identification and Genome Diversity Analysis

Illumina reads were mapped to *S. cerevisiae* strain S288C reference genome using Bowtie 2 (Langmead and Salzberg, 2012). Freebayes (Garrison and Marth, 2012) was used to find genetic variants, including SNPs, in all mapped samples.

To detect genetic variants specific for flor strains, we used a custom perl script to filter Freebayes output file. According to the

filter, each sample must have a minimum 20x mapping depth in the variant position, all flor strains must support the same variant, with 90% read frequency support in each flor strain, and all wine strains can support any other allele different from the flor-specific variant, with a minimum 90% read frequency support. In total, 2,270 flor-specific genetic variants were detected using this filter (Supplementary Table S2) for the set of strains phylogenetically classified to “flor” and “wine” clades (Supplementary Table S1).

The variants were then analyzed for their non-synonymous effect on *S. cerevisiae* S288C ORFs using the Variant Annotation Integrator tool at the UCSC genome browser (Hinrichs et al., 2016). The non-synonymous to synonymous substitution rate or dN/dS ratio (Zhang et al., 2006) was calculated from the table of obtained variant calling datasets for flor yeast strain-specific SNPs and InDels.

## Phylogenetic Analysis

To analyze the phylogenetic position of selected flor strains within global yeast phylogeny, we inferred phylogenies based on multiple alignments of 16 conserved chromosomal regions suggested by Strobe et al. (2015). Corresponding gene segments were extracted from the genome assemblies of strains listed in Supplementary Table S1 (except for strains WLP862 and AWRI1796) using BLAST, concatenated, and added to the collection of 218 kb sequences of 95 natural, industrial, and clinical strains downloaded from <https://github.com/daskelly/yeast100genomes/>. Multiple alignment was performed with MAFFT (Kato and Standley, 2014) in fftnsi mode. Neighbor-joining tree was also constructed with MAFFT and visualized with Figtree 1.4.3 (Rambaut, 2012).

For the SNP tree, SNPs were filtered where each sample has a minimum 0.9 frequency of the major allele and a minimum 20x depth. SNPs where all major alleles for all samples are the same were excluded from tree building. Using these filters, a total of 14,069 sites were defined and concatenated into the alignment acceptable for the tree construction using a custom perl script. A maximum-likelihood tree was build using PhyML (Stamatakis, 2014). Raw sequence data and genome assemblies for flor and wine yeast strains listed in Supplementary Table S1 were used for construction of SNP-based phylogenetic tree.

## Genes of *S. cerevisiae* S288C Missing in the Analyzed Yeast Strains

Illumina sequencing reads obtained for strains I-30, I-329, and I-566 were mapped to the reference genome using Bowtie 2 and the coverage percent of each gene was calculated using Bedtools. Gene was considered as being missing when the coverage was less than 50%. In addition, we checked the absence of the “missing” genes in *de novo* assemblies by mapping contigs to the reference genome.

## Non-reference Genes Present in the Analyzed Yeast Strains

For pangenomic analysis of the presence-absence variation of key industry-related non-reference genomic segment, we used a collection of 26 sequences suggested in the recent extensive

<sup>2</sup><https://github.com/najoshi/sickle>

comparative genomics study of wine yeast strains (Borneman et al., 2016). Illumina sequencing reads obtained for strains I-30, I-329, and I-566 were mapped to these sequences using Bowtie 2 and the coverage percent of each gene was calculated using Bedtools.

All genes annotated in *de novo* assemblies of I-30, I-329, and I-566 genomes were compared with *S. cerevisiae* S288C genes using BLASTN search. The gene was considered as “new” in the absence of a hit with more than 70% identity over more than 80% of the gene length.

## Gene Ontology (GO) Enrichment Analysis and List Comparison

Gene sets and ORFs identified as bearing mutations or copy number alterations specific for flor yeast strains were analyzed with YeastMine tools (Balakrishnan et al., 2012) at SGD. For cases when gene ontology (GO) analysis did not show statistically significant enrichment ( $p < 0.05$ , Holm–Bonferroni corrected; background: SGD default) we performed GO slim term mapping and compared frequencies of the most represented terms in obtained lists versus default background.

## Other Analysis Tools

Routine sequence visualization and manipulation of nucleotide sequences was performed with Ugene (Okonechnikov et al., 2012). For drawing Venn diagrams depicting similarities and differences between various gene lists, we used the tool developed by Ghent University.<sup>3</sup>

## Nucleotide Sequence Accession Number

This BioProject has been deposited in GenBank under accession number PRJNA414946. The sequences obtained in this project have been deposited in the NCBI Sequence Read Archive under the accession numbers SRR6333650, SRR6333651, and SRR6333652. The annotated genome sequences of strains I-30, I-329, and I-566 have been deposited in the GenBank database under accession numbers PTEP000000000, PTER000000000, and PTEQ000000000, respectively.

# RESULTS

## Strains' Origin

The “Magarach” Collection of the Microorganisms for Winemaking was started more than 60 years ago and at present harbors several hundred strains of wine-making microflora of yeast origin. Several yeast strains belonging to the group of flor yeast were either isolated from different wineries of the former Soviet Union and other countries or obtained from other collections (Kishkovskaia et al., 2017). Some strains were subjected to mutagenesis and selection for increased ethanol tolerance and velum formation properties. The biochemical, physiological, genetic, and winemaking properties of 16 flor yeast strains were re-evaluated in our recent study (Kishkovskaia et al., 2017). Three strains with superior sherry-making properties

and shown to be genetically distinct according to microsatellite markers, ITS, and interdelta genotyping were subjected to *de novo* whole genome sequencing. Strains I-566 and I-30 were isolated from wineries producing sherry-like wines in Armenia and Crimea, respectively. Strain I-329 was obtained by N. F. Sayenko from a Spanish sherry winery more than 70 years ago and then was improved using selection methods in 2004. Strains I-329 and I-566 carry a 24 nt deletion in the ITS1 region found in Spanish sherry yeast strains, while in strain I-30, this region contains the C insertion characteristic of French Jura flor strains (Charpentier et al., 2009). Winemaking-relevant characteristics of these strains were reported in Kishkovskaia et al. (2017).

## Genome Sequencing, Assembly, and Annotation

All three Magarach flor yeast genomes were sequenced using Illumina NGS platform at about 200X coverage. In addition, about 60x coverage by PacBio long reads was obtained for strains I-30 and I-329. Final assemblies had total sizes in the range of 11,50–11,59 Mbp, consisting of 71–562 contigs with the N50 contig length between 58 and 511 kb (Table 1). As expected, the use of PacBio long reads considerably improved the assembly. Complete mitochondrial genomes were assembled as circular contigs in all three strains (Mardanov et al., 2017). On average, about 5,300 protein-coding genes and 300 tRNA genes were predicted in the nuclear genomes of strains I-30, I-329, and I-566 (Table 1).

## Phylogenetic Relationships of Wine and Flor Yeast Strains

Flor yeast strains from different countries are known to share unique origin based on microsatellite typing and population analysis (Legras et al., 2014). To assess the phylogenetic position

**TABLE 1 |** Statistics of sequencing, *de novo* assembly, and annotation of nuclear genomes.

Strain	I-30	I-329	I-566
Coverage by Illumina HiSeq-2500	264X	234X	219X
Number of Illumina reads (after filtration)	14,051,242	13,236,656	13,682,586
Average length of Illumina reads, nt	218	205	184
Coverage by PacBio RSII	59X	60X	–
Number of PacBio reads	122,857	191,070	–
Average length of PacBio reads, nt	5,596	3,655	–
Total contigs	155	71	562
Number of contigs larger than 500 bp	80	45	443
Contig N50, bp	470,360	511,336	58,253
Largest contig, bp	870,688	1,035,721	252,238
Total assembly length, bp	11,587,783	11,587,876	11,502,012
Predicted protein-coding genes	5,323	5,323	5,290
Predicted tRNA genes	285	288	284
Ty elements	472	474	594

<sup>3</sup><http://bioinformatics.psb.ugent.be/webtools/Venn/>



of Magarach flor strains within the global yeast phylogeny, we used the large available *S. cerevisiae* phylogenetic tree constructed on the set of 16 conserved regions from 95 yeast strains (Strope et al., 2015). Corresponding sequences were extracted from genome assemblies of I-30, I-329, and I-566 strains, as well as from 20 other flor and wine strains from Genowine project and the collection of the Australian Wine Research Institute (Supplementary Table S1). One more strain from the Magarach collection, I-328 (Mardanov et al., 2018), was also included in the analysis.

According to the obtained phylogeny of 118 yeast strains, all flor strains except F12-3B (see below) form very tight cluster adjacent to the main wine/European clade (Figure 1). In this cluster, strain I-329 and Spanish flor strains (FS2D, F25, 7-7) form a separate branch, and another branch comprises strains I-566 and I-30.

These data were further refined using the whole-genome SNP-based approach similar to the one described by Coi et al. (2017). According to the obtained tree, our flor strains definitely belong to the “flor group” (Figure 2). They are phylogenetically related to the flor strains 7-7 (Spain), F25 (Spain), FS2D (Sardinia), TS12-A7 (Hungary), and the strain AWRI723 (Australia). The later strain was also found in the flor cluster on a phylogenetic tree constructed using the set of 16 conserved regions (Figure 1). On the contrary, strain F12-3B previously described as flor strain appeared to be closer to wine group on both phylogenetic trees. Strain I-328 from the Magarach collection, previously described as flor strain (Kishkovskaia et al., 2017), is phylogenetically related to the wine group.

## Gene Loss and Gain in Flor Yeast Relative to S288C

Events of gene deletion and acquisition are rather frequent in natural yeast populations and among industrial and commercial strains (Dujon, 2010; Borneman et al., 2011, 2016; Gallone et al., 2016; Marsit et al., 2017). The redundant nature of yeast genome suggests that many genes can be lost without dramatic effects on strain viability and fitness (Dean et al., 2008; DeLuna et al., 2008), but the real evolutionary implications are unclear (Sliwa and Korona, 2005). On the other hand, there are many well-documented events of gene acquisition by wine yeast species through horizontal gene transfer or introgression from other yeast or bacterial species (Galeote et al., 2010; Bergström et al., 2014). The transferred segments encode functions with a clear impact on wine fermentation such as stress resistance and improved utilization of carbon and nitrogen sources, justifying important role of this type of diversification in yeast evolution (Marsit and Dequin, 2015; Marsit et al., 2015).

The degree of gene loss in the three Magarach flor strains as determined using mapping of reads on the genome of the reference strain S288C, as well as by analysis of *de novo* assemblies, appeared to be rather low. A total of 92 genes present in strain S288C were missing in all three sequenced strains (Supplementary Table S3). No genes absent in only one or two strains were identified. These lost genes predominantly encoded either Ty transposon proteins (65), or putative proteins with

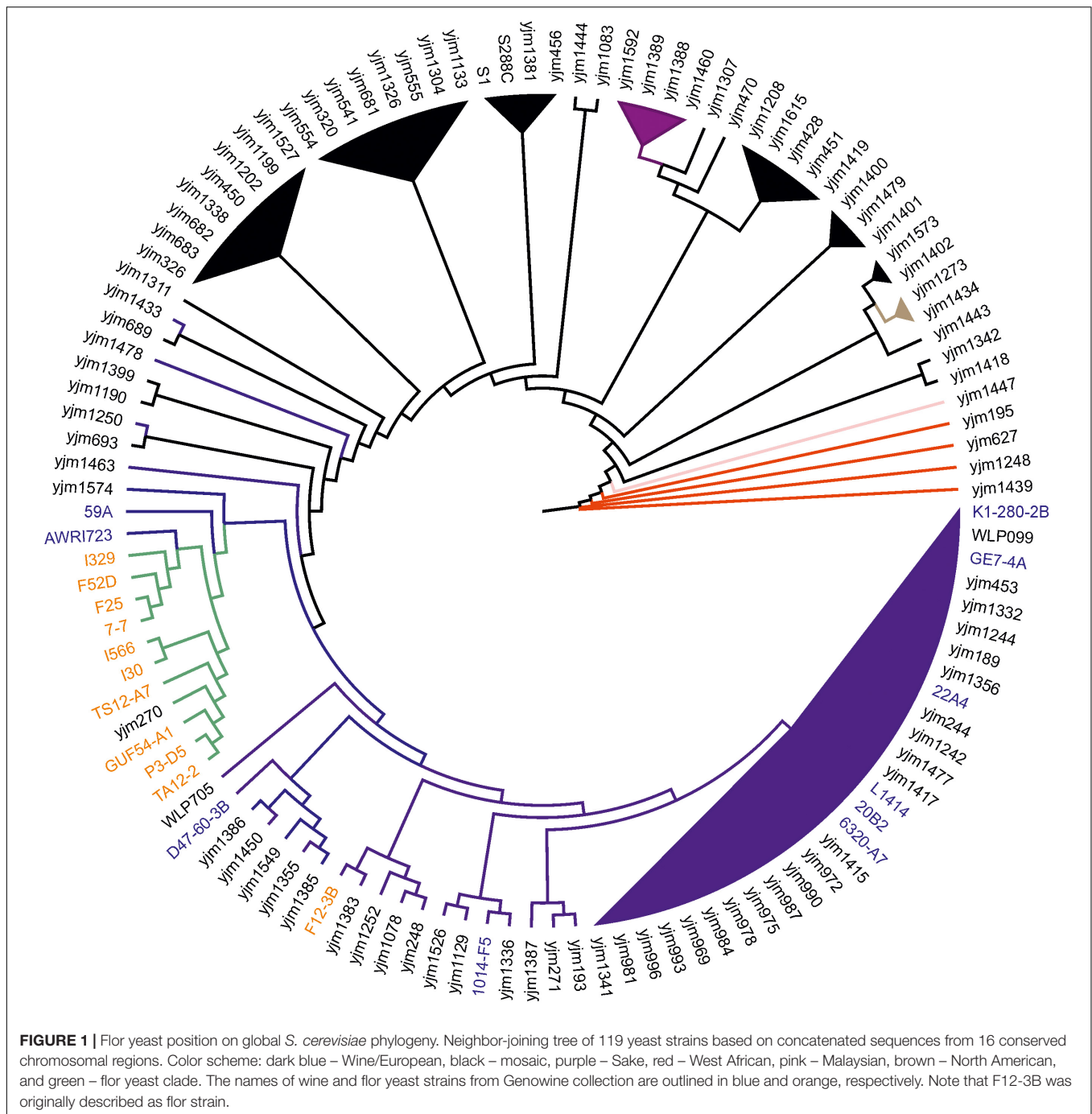
unknown functions (17). The effects of the loss of 10 genes with known functions may be significant. They are located in three genomic loci.

Thus, we observed extended deletions of genes responsible for iron uptake at the subtelomeric region of chromosome XV and nearly located DNA photolyase PHR1, the asparaginase genes near rDNA array on chromosome XII, and *MAL* genes (transcriptional factor *MAL13* and maltose transporter *MAL11*) on chromosome VII. These deletions may obviously affect carbon metabolism, amino acid metabolism, and iron homeostasis.

Comparative genomic analysis of numerous wild, commercial, industrial, and clinical isolates of *S. cerevisiae* has revealed extended regions of genetic material, scattered across distinct chromosomal regions, apparently absent from the reference S288c genome (Novo et al., 2009; Dunn et al., 2012; Song et al., 2015; Borneman et al., 2016; McIlwain et al., 2016). Many of these strain-specific loci encode functions beneficial for particular industry-related traits. Well-known examples of clustered loci of industrial importance are the RTM1 cluster, important for membrane phospholipid homeostasis at high ethanol concentrations, the “wine circle” (Borneman et al., 2011), or region B, regions A and C (Novo et al., 2009) identified in wine strain EC1118, the heat-resistant toxin *KHRI* (Goto et al., 1990), the *MPR1* gene encoding L-azetidine-2-carboxylic acid acetyltransferase conferring ethanol and cold resistance, and oxidative stress tolerance (Takagi et al., 2000). A useful compendium of these non-reference sequences was developed by Borneman et al. (2016) and we used this resource to identify non-reference sequences in our three flor yeast genomes (Supplementary Table S3).

The nuclear genomes of Magarach strains contained about 108–126 kb absent in the reference genome. All three strains lacked the so-called region A previously identified in EC1118 genome. Region B was found only in strain I-30 where it comprises five genes: transcription factor, 5-oxoprolinase, nicotinic acid transporter, flocullin-like protein, and a hypothetical protein. Region C encodes, among other genes, *FOT* oligopeptide transporters beneficial for utilization of “non-conventional” nitrogen sources. Many flor yeast strains contain this region, but region C is absent from the three our strains. Not surprisingly, the three analyzed genomes also lacked the RTM-cluster, which is known to be advantageous for beer and bioethanol strains, grown on molasses.

Potentially important for flor yeast physiology and metabolism is the presence in all three genomes of the *MPR1* gene and two other regions found in wine yeast strains (Argueso et al., 2009; Akao et al., 2011). The 5 kb segment encoding the ortholog of GPI-anchored cell-wall protein AWA1 from sake strain may positively affect surface adhesion of flor yeast cells (Shimoi et al., 2002). All three Magarach strains contained AWA1-like genes most similar to ones from wine strains YJM1341 and YJM1415. The 19 kb cluster from bioethanol strain JAY291 is known to encode a paralog of the HXT4 high-affinity glucose transporter and alpha-glucosidase MAL32, both advantageous under conditions of sugar limitation (Akao et al., 2011). These two genes are present in each of Magarach strains. In contrast to these full-length clusters, other sequences listed in



Supplementary Table S3 are either missing or are represented by significantly truncated fragments. The potential role of KHR1 toxin (present in I-30 and I-566), the EC1118 1M36 cluster harboring one hypothetical protein gene (present in all three strains), and the endogenous 2 mcm plasmid (present in I-30 and I-566) is unclear.

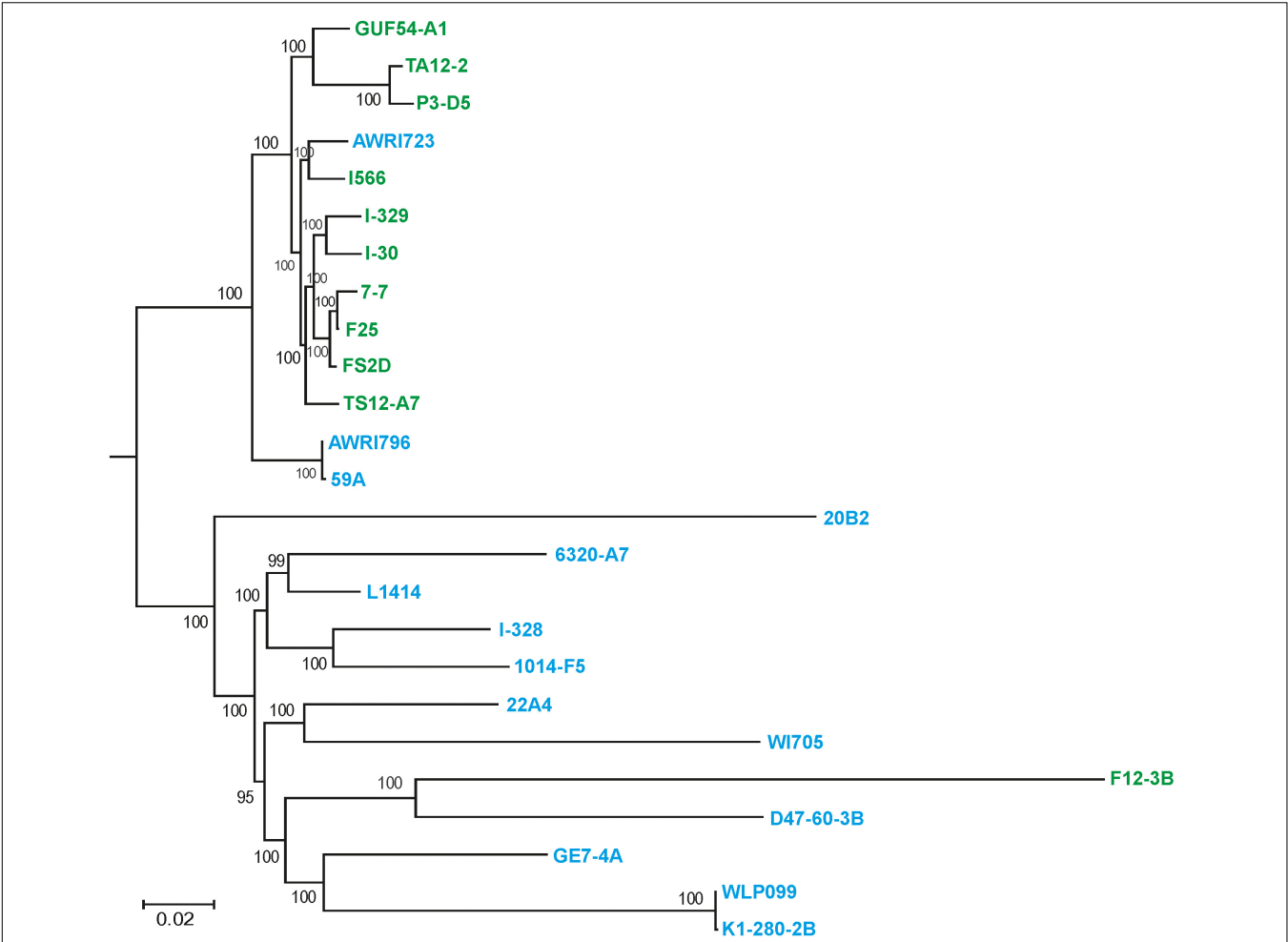
The search for non-reference genes in *de novo* assemblies revealed one to three new genes in each strain in addition to genes located in above-mentioned regions (Supplementary Table S3). All of them encode hypothetical proteins with

unknown functions. Interestingly, all three strains contained a gene which predicted product is identical to 246-aa protein R103\_P20001 from *S. cerevisiae* R103. Highly similar genes were present in several other wine yeast strains (JAY291, FostersB, YJM789, FostersO, Lalvin QA23, VIN7, and VL3).

### Flor-Yeast-Specific Sequence Variations

Using variant calling, we have identified two types of variations – SNP and InDel in three Magarach flor yeast genomes, accounting in each case to more than 45,000 variable site relative to the





**FIGURE 2 |** Flor and wine yeast phylogeny based on SNP analysis. A maximum-likelihood phylogenetic tree of 27 flor, wine, and lab *S. cerevisiae* strains inferred from SNP data. Numbers at nodes represent the bootstrap support values. The names of flor strains are in green, wine strains are in blue, and lab strains are in black.

reference S288C genome (Table 2). In order to narrow down this set and to find flor yeast specific mutations (FYSMs), we have compared obtained SNP sites to draft genomes of wine and flor yeast strains listed in Supplementary Table S1 and phylogenetically assigned to “wine” and “flor” clades as described in Figure 2. In total, we found 2,270 high-quality biallelic flor yeast specific SNV (both SNP and InDels) in 1,337 genomic loci (Supplementary Table S2) and subjected this set to different types of analyses. First, we analyzed the distribution of variable sites across the chromosomes and found significant SNV enrichment in some “hot spots,” including subtelomeric regions of several chromosomes in accordance with well-known view of these structures as “hotbeds” of genome variation in yeast (Supplementary Figure S1). Using SNPeff, we classified mutations functionally in different subcategories (Table 2). These new gene sets, in particular, genes with missense mutations and with mutations in promoter regions, were subjected to GO enrichment analysis to identify GO terms that are under- or over-represented compared to reference genome.

**TABLE 2 |** SNP categories in flor strains.

Total SNP and InDels	Number
I-30	46,756
I-329	47,438
I-566	45,656
Flor yeast specific variants	2,270
Missense	982
Synonymous	583
Frameshift	8
Upstream	549
Downstream	121
Intron	4
Intergenic	8
Stop and splice	15

The ratio between missense and synonymous mutations in coding regions was high (dN/dS = 1.68), and thus we first looked for GO terms enriched in the set of genes with

missense mutation in coding regions likely to be under positive selection. The GO analysis of obtained list of 670 unique genes revealed significant alterations in “cell component,” “biological process,” and “molecular function” categories relative to the reference genome (Supplementary Table S4). In particular, in “cell component” category such terms as “intracellular membrane bound organelle” and “protein complex” were enriched. In “molecular function” category, various terms such as “ATP binding” and “ATP ase activity” were enriched. In “biological process” category, we found enrichment for the following terms: “regulation of cellular process,” “response to stimulus,” “cellular component organization,” “developmental process,” “aromatic compound biosynthetic process,” and others (Supplementary Table S4). This analysis points to importance of process related to integrity of intracellular organelles, ion, and protein homeostasis for flor yeast specific physiological and biochemical features. Notably, in this list, we found 20 genes for stress-responsive transcription factors involved in reprogramming of non-fermentative metabolism, *ACC1*, *CAT8*, *LN3*, *ERT1*, *GCN4*, *GSY2*, *HAP1*, *LST8*, *MSN4*, *NTH1*, *PFK2*, *PHO85*, *PSK1*, *RIM15*, *SUT1*, *TCO89*, *TOR2*, *TPK2*, *TPK3*, and *YAK1* (Soontornngun, 2017).

In order to select ORFs likely to be under stronger positive selection, we have further divided the set of ORF with dN/dS > 1 according to the number of sites per gene. We have ranged the genes with missense mutations according to the number of SNP per gene and those with two or more missense SNP were considered as “highly polymorphic.” For this group of 106 genes (Supplementary Table S5), we performed GO slim mapping and detected prevalence for GO slim terms in all three categories. In the “biological process” category, genes involved in “response to chemical,” “transcription from RNA polymerase II promoter,” “ion transport,” “mitotic cell cycle,” “signaling,” “cellular response to DNA damage stimulus,” “transmembrane transport,” “carbohydrate metabolic process,” “DNA repair,” and others were over-represented (Supplementary Table S5). In the “molecular function” group, the following GO terms were enriched: “hydrolase activity,” “transferase activity,” “ATPase activity,” “transmembrane transporter activity,” “DNA binding,” “enzyme regulator activity,” “helicase activity,” etc. Such GO terms as “cellular bud,” “plasma membrane,” “site of polarized growth,” and others were prevalent in “cell component category.”

The small group of 25 genes with “deleterious mutations” (stop-codon lost or gained, frameshift, and altered splicing site) included proteins involved in transcription regulation and signaling, and unknown genes with unclear role for flor yeast specific adaptation (Supplementary Table S5).

Mutation in the upstream and downstream regions may positively or negatively affect gene expression. We focused on upstream mutations and selected a group of 106 genes with two or more SNPs in promoter regions and performed GO enrichment analysis. We found enrichment for terms related to cellular ion homeostasis, reflecting possible positive selection (Supplementary Table S6). Pathways’ enrichment analysis detected enrichment of gene related to acetoin biosynthesis, pentose phosphate pathway, and amino acid

catabolism, all possibly related to flor yeast specific biochemical features.

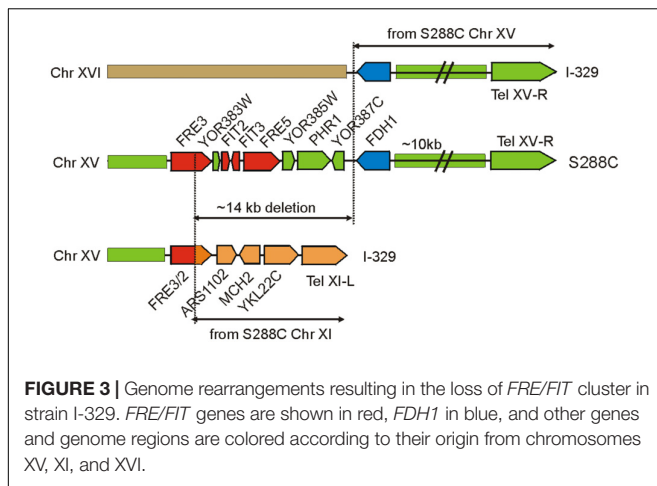
The group of 25 genes with three or more SNVs in the promoter regions (Supplementary Table S7) included those related to carbon metabolism (*PDC1* and *TKL1*) and utilization of unconventional nitrogen sources (*SRY1*), aquaporin *AQY2*, and several proteins that may affect metal ion transport (ferric reductase *FRE6* and zinc transporter *YKE4*), RNA processing (*YRA1* and *MTR2*), and BET3 component of the transport protein particle. Changes in regulation of genes relevant to mitochondrial function (*SDH6*, *SMF1*, *HMX1*, and *FRE6*) may be important for flor yeast under conditions of oxidative metabolism (Supplementary Table S7).

Finally, we ranged all polymorphic genes by total number of SNP per gene (upstream, downstream, synonymous, and missense) to identify those that are most polymorphic and selected among them those with dN/dS > 1. This selection yielded a rather interesting group of 39 extremely polymorphic genes (five or more sites per gene) with functions possibly directly related to flor yeast fitness (Supplementary Table S8). Besides already identified genes with upstream mutations, we found several genes with functions related to flor yeast morphology, in particular septin ring formation (*RGA2*, *VHS2*, and *YCK2*) and intracellular trafficking (*VPS13*, *COS9*, and *SEC24*), that may contribute directly or indirectly to enhanced ability of flor yeast for biofilm formation. Modification of *DNA2* gene involved in DNA replication, double-stranded break repair, and telomere maintenance may enhance the resistance of flor yeast to mutagenic action of high ethanol and acetaldehyde concentrations. Several genes encode proteins with unknown functions and their significance for flor yeast specific properties remains to be elucidated.

## Structural Variations in Flocculins

The key role of *FLO11* in determining the ability of flor yeast for biofilm formation is well established (Fidalgo et al., 2006; Ishigami et al., 2006; Zara et al., 2009). The two sequenced strains, I-30 and I-329, carry a characteristic *FLO11* promoter deletion, known to positively affect *FLO11* transcription (Fidalgo et al., 2006). The coding regions of *FLO11* on our strains were extended due to accumulation of tandem repeats in the central domain (Supplementary Figure S2) that was shown to yield more hydrophobic Flo11p variant and increase the ability of yeast cells to float (Fidalgo et al., 2006).

The opposite trends were observed for three other adhesin genes, *FLO1*, *FLO5*, and *FLO9*. Full-size genes for the largest flocculin Flo1p (1537 a.a. long in strain S288C) were not found in all three flor strains; only genes able to encode 390 a.a. long protein were present. On the contrary, nearly full size *FLO5* genes were found in all Magarach strains. *FLO9* genes were also found, but the number of tandem repeats in the central domain was reduced relative to the reference gene. This balance change between the two groups of Flo proteins in flor yeast strains indicates a possible positive selection in favor of increased *FLO11* expression leading to improved velum formation.



## Phenotypic Assessment of Variations in Iron Uptake Genes

The three sequenced Magarach flor strains possess two structural variations with a potential strong impact of iron uptake and homeostasis – the 14 kb deletion in the right subtelomeric region of chromosome XV (Figure 3) and a flor-yeast-specific deleterious mutation in the gene encoding Aft1 transcription factor, leading to stop-codon insertion at position 648, eliminating 42 C-terminal amino acid residues (Supplementary Figure S3).

Mapping of contigs obtained for Magarach strains to the reference genome revealed that this *FRE/FIT* deletion likely resulted from recombination between subtelomeric regions of chromosomes XV and XI (Figure 3). The left subtelomeric region of chromosome XI contained gene *FRE2* exhibiting high sequence similarity to *FRE3* in the *FRE/FIT* cluster on chromosome XV; recombination between these sequences produced “hybrid” *FRE2/FRE3* gene followed by genes, initially located between *FRE2* and the left telomere of chromosome XI. The 14-kbp *FRE/FIT* region appeared to be lost, while telomere-proximal region with *FDH1* gene was translocated to the chromosome XVI (Figure 3).

The *FRE* and *FIT* proteins are known to cooperate in iron uptake (Outten and Albetel, 2013). Fit2p and Fit3p are GPI-anchored cell-wall mannoproteins facilitating iron uptake through increasing the amount of iron associated with the cell wall and periplasm (Protchenko et al., 2001). Fre2p and Fre5p are plasma membrane reductases that facilitate uptake of siderophore-bound iron. Aft1 upregulates expression of iron uptake genes when iron is scarce and in combination with Yap5 transcription factor is essential to maintain iron homeostasis in yeast (Martínez-Pastor et al., 2017). The Q648X mutation removes C-terminal region with potential sumoylation and CK2 phosphorylation sites, leaving intact the Q-rich domain potentially involved in transcriptional activation (Supplementary Figure S3). The combination of these strong structural variations was found in other flor strains and this prompted us to directly assess its phenotypic effects through comparison of flor and lab yeast strains.

Iron is vital for aerobic flor yeast metabolism under conditions of biological wine aging, but excess iron may be detrimental due to accumulation of toxic reactive oxygen species, damaging cellular macromolecules (Bresgen and Eckl, 2015). There is a significant variation in iron uptake capabilities in natural yeast isolates leading to separation of “iron-sensitive” or “iron-resistant” groups depending on strain response to excess iron in the medium (Martínez-Garay et al., 2016). To assess the net effect of indicated structural variations on iron homeostasis and uptake of flor yeast strains, we performed growth assays similar to those described before (Martínez-Garay et al., 2016).

All three flor strains were more sensitive to excess iron in the medium compared to lab strain (Supplementary Figure S4). Growth on solid medium was inhibited at ferric iron concentration above 3 mM; in liquid medium, the retardation of cell division was observed if concentration of ferrous iron was above 1 mM and became more pronounced at 4 mM (Supplementary Figure S4). In accordance with this iron-sensitive phenotype, flor yeast strains displayed increased coloration on the plates with 2 mM ferric iron and 1% methylene blue indicating more oxidized cellular redox state in the presence of iron (Supplementary Figure S4).

Increased iron sensitivity and iron-dependent methylene blue oxidation are considered to be indicative of improved iron uptake (Martínez-Garay et al., 2016), which prompted us to propose that flor yeast strains are more proficient in iron uptake. This assumption was tested in iron accumulation assays for I-329 strain and control BY4743 strain grown at different conditions. The intracellular iron content in the iron-sensitive strain I-329 was higher under both low-iron (0.1 mM) and high-iron (4 mM) conditions, indicating its iron uptake proficiency (Supplementary Figure S4). Since no other genetic alterations in known iron uptake and homeostasis system were detected in three sequenced strains, we attribute this property to combined effect of *AFT1* mutation and *FRE/FIT* cluster deletion.

## DISCUSSION

Flor yeast strains are highly specialized microbial agents used for production of biological aged wines through sophisticated winemaking process (Alexandre, 2013). The important properties of flor yeast, such as high tolerance to harsh environment conditions, capability for velum formation and production of specific flavor compounds are likely to have evolved through centuries of “unconscious” human selection and domestication (Legras et al., 2007, 2014). Understanding the nature of the genetic variations specifying the particular phenotypic properties of flor yeast is of major importance for the study of molecular mechanisms of yeast adaptation to industrial processes and specific ecological niches and identification of flor yeast specific genes and alleles.

Our comparative genomic approaches have revealed complex landscape of genetic variation in three newly sequenced flor strains represented by SNPs, InDels, events of gene loss and gain. Subsequent GO analysis uncovered differential contribution of different forms of genetic variation to the build-up of the

flor yeast genomes. The polymorphism in the genes involved in yeast morphology, carbohydrate metabolism, ion homeostasis, response to osmotic stress, lipid metabolism, DNA repair, cell wall biogenesis, etc., in sherry strains is mainly due to SNP/InDel accumulation. On the other hand, the genes for FLO adhesins were the subject of significant structural variation that could explain the increased biofilm-formation capacity of flor yeast.

It is necessary to note the difference of our results from the results of the recent study of genomic signatures of flor yeast adaptation reported by Genowine researchers (Coi et al., 2017), although the set of strains and assemblies essentially overlapped. Our criteria for selection of flor-yeast-specific mutations were different both in terms of dataset analysis methods and the selection of affected regions. For instance, we have included mutations in regulatory regions in the set of compared SNPs. Such mutations, as recently was shown, may affect gene expression both positively or negatively not only by affecting transcription binding sites and their spacing in promoters, but also via DNA “zip codes” responsible for interaction between promoters and nuclear memory (Brickner et al., 2015) and mRNA stability sites (Shalem et al., 2015).

Superposition of our set of 670 genes with FYSM and the dN/dS ratio > 1 with the FYSM genes likely to be under positive selection identified in Genowine study showed an overlap of 89 protein-coding genes (Supplementary Table S9). This list is enriched for proteins located at the cell periphery (23 proteins), involves several genes implicated before in regulation of ethanol tolerance (*YDR274*, *FTR1*, *CCS1*, and *BRE5*), signaling (*IRA1* and *TCO89*), DNA repair (*DNA2* and *DDC1*), and transporters (*PMA1*, *TPO5*, and *QDR2*).

Irrespective of the differences in algorithms and approaches applied to select for FYSM genes, this comparison shows the difference in attestation of analyzed strains to flor or wine groups. For instance, the F12-3B strain originally classified as “flor yeast strain” (BioSample: SAMEA2612327) according to SNP-based and 16 conserved regions-based phylogenetic trees belongs to the “wine” clade, while the strain AWRI1723 (BioSample: SAMN 04286124) belongs to the “flor” clade. Wine strains 59A and AWRI 1796 are also phylogenetically closer to the flor group (Figure 2). Of course, strains phylogenetically related to wine group may perform well in biological aging due to some specific set of mutations. It is also possible that some strains originally described as “wine” but phylogenetically related to the flor clade could perform wine aging as well. Obviously, more extensive comparative genomic and post-genomic analysis of flor yeast strains is required to clarify these issues.

Only a limited number of gene acquisition and loss events were observed in three Magarach flor strains. Only two genes, missing in the reference strain S288c, were found in all three studied flor strains. The first is the *MPRI* gene coding for *N*-acetyltransferase that is involved in oxidative stress tolerance via proline metabolism (Nishimura et al., 2010). Its presence is apparently beneficial for flor strains thriving under aerobic conditions. The second gene encodes a protein with unknown function. Both genes are not unique for flor strains and were found in a number of wine yeasts. The gene loss events are mostly related to genes encoding transposon-related and hypothetical

proteins, but deletions of three larger genomic loci were detected as well. Deletions of the *MAL1* locus located in the subtelomeric region of chromosome VII are rather often event in natural population and may impose no obvious phenotypic effect since five nearly identical *MAL* loci have been identified in *S. cerevisiae* (Charron et al., 1989; Naumov et al., 1994). Deletion of the asparaginase gene cluster is also quite often and is not expected to be clearly related to conditions of biological wine aging. The third deletion, targeting the *FRE-FIT* cluster, could be more important.

We took an advantage of the two potentially strongly impacting FYS-genetic variation that could be directly assessed through comparison of wild type flor and lab strains, the deletion of *FRE-FIT* cluster and mutation in *AFT1* transcription factor. Our phenotypic analysis has shown that analyzed flor strains are more sensitive to iron toxicity that is likely to be related to their increased capacity for iron uptake. This assumption was proved in our iron accumulation assays.

The adaptive significance of this trait of course requires additional evaluation. Since *FRE-FIT* genes are dispensable for iron uptake in the absence of siderophore-bound iron (Protchenko et al., 2001), their deletion may be neutral for flor yeasts growing in sterilized wine materials in course of sherry wine making. However, it is also possible that such deletion in combination with flor-yeast-specific *Aft1* allele is advantageous to improve iron uptake from wine materials with low iron content.

*Aft1* is a known positive activator of the iron regulon, that besides *FRE1-4* metalloredutase genes and *FIT1-3* iron siderophore transporters includes genes involved in cell-surface high-affinity iron acquisition (*FET3/FTR1* system), multiple genes for proteins involved in iron recycling, intracellular transport, post-transcriptional regulation, etc. (Martínez-Pastor et al., 2017). One may expect that elimination of the *FRE-FIT* genes in flor yeast strains is compensated by activation of *FET3/FTR1* system and alteration in the iron levels between cytosol, vacuoles, and mitochondria. Thus, *Aft1* targets are attractive candidates for more detailed gene expression analysis in flor yeast strains under a variety of conditions and are in focus of our current investigation.

The metal content in wines is of great interest due to influence on wine technology and is determined largely by geographic origin (Galani-Nikolakaki et al., 2002). It is known that in Jerez wines, the iron content is below 0.05 mM (Paneque et al., 2009). This may be important to preserve typicality of at least some varieties of sherry wines. It is known, for instance, that Fino sherry wines undergo browning at iron concentration above 0.05 mM (Benítez et al., 2002). The influence of *FIT* genes deletion on flor yeast cell wall properties should also be evaluated. Individual and combined allele replacements, iron toxicity, biofilm formation, and other assays may be required for this type of research.

We suppose that the results of our analysis, sequence data, and *de novo* assemblies will help to infer the evolutionary history and the adaptive evolution of flor yeasts. They can also be useful for functional analysis of flor yeast, for instance, through application of modern synthetic biology and genome editing tools (Jagtap et al., 2017), recently developed set of haploid flor strains



(Coi et al., 2016) to aid in development of novel flor yeast with improved properties.

## AUTHOR CONTRIBUTIONS

AM, ME, and NR designed the research project and wrote the paper. SK, TT, AB, and AM performed the research. AB, AM, NR, and ME analyzed the data. All authors read and approved the manuscript.

## FUNDING

This work was supported by Russian Science Foundation (Grant No. 16-16-00109).

## REFERENCES

- Akao, T., Yashiro, I., Hosoyama, A., Kitagaki, H., Horikawa, H., Watanabe, D., et al. (2011). Whole-genome sequencing of sake yeast *Saccharomyces cerevisiae* Kyokai no. 7. *DNA Res.* 18, 423–434. doi: 10.1093/dnares/dsr029
- Alexandre, H. (2013). Flor yeasts of *Saccharomyces cerevisiae*-their ecology, genetics and metabolism. *Int. J. Food Microbiol.* 167, 269–275. doi: 10.1016/j.jfoodmicro.2013.08.021
- Argueso, J. L., Carazzolle, M. F., Mieczkowski, P. A., Duarte, F. M., Netto, O. V. C., Missawa, S. K., et al. (2009). Genome structure of a *Saccharomyces cerevisiae* strain widely used in bioethanol production. *Genome Res.* 19, 2258–2270. doi: 10.1101/gr.091777.109
- Balakrishnan, R., Park, J., Karra, K., Hitz, B. C., Binkley, G., Hong, E. L., et al. (2012). YeastMine—an integrated data warehouse for *Saccharomyces cerevisiae* data as a multipurpose tool-kit. *Database* 2012:bar062. doi: 10.1093/database/bar062
- Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., et al. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 19, 455–477. doi: 10.1089/cmb.2012.0021
- Barrales, R. R., Korber, P., Jimenez, J., and Ibeas, J. I. (2012). Chromatin modulation at the FLO11 promoter of *Saccharomyces cerevisiae* by HDAC and Swi/Snf complexes. *Genetics* 191, 791–803. doi: 10.1534/genetics.112.140301
- Barua, S., Li, L., Lipke, P. N., and Dranginis, A. M. (2016). Molecular basis for strain variation in the *Saccharomyces cerevisiae* adhesin Flo11p. *mSphere* 1:e129–16. doi: 10.1128/mSphere.00129-16
- Benítez, P., Castro, R., Antonio Sanchez, P. J., and Barroso, C. G. (2002). Influence of metallic content of fino sherry wine on its susceptibility to browning. *Food Res. Int.* 35, 785–791. doi: 10.1016/S0963-9969(02)00075-3
- Bergström, A., Simpson, J. T., Salinas, F., Barré, B., Parts, L., Zia, A., et al. (2014). A high-definition view of functional genetic variation from natural yeast genomes. *Mol. Biol. Evol.* 31, 872–888. doi: 10.1093/molbev/msu037
- Borneman, A. R., Desany, B. A., Riches, D., Affourtit, J. P., Forgan, A. H., Pretorius, I. S., et al. (2011). Whole-genome comparison reveals novel genetic elements that characterize the genome of industrial strains of *Saccharomyces cerevisiae*. *PLoS Genet.* 7:e1001287. doi: 10.1371/journal.pgen.1001287
- Borneman, A. R., Forgan, A. H., Kolouchova, R., Fraser, J. A., and Schmidt, S. A. (2016). Whole genome comparison reveals high levels of inbreeding and strain redundancy across the spectrum of commercial wine strains of *Saccharomyces cerevisiae*. *G3* 6, 957–971. doi: 10.1534/g3.115.025692
- Bresgen, N., and Eckl, P. M. (2015). Oxidative stress and the homeodynamics of iron metabolism. *Biomolecules* 5, 808–847. doi: 10.3390/biom5020808
- Brickner, D., Coukos, R., and Brickner, J. (2015). INO1 transcriptional memory leads to DNA zip code-dependent interchromosomal clustering. *Microb. Cell* 2, 481–490. doi: 10.15698/mic2015.12.242
- Charpentier, C., Colin, A., Alais, A., and Legras, J. L. (2009). French Jura flor yeasts: genotype and technological diversity. *Antonie Van Leeuwenhoek* 95, 263–273. doi: 10.1007/s10482-009-9309-8

## ACKNOWLEDGMENTS

This work was performed using the scientific equipment of Core Research Facility “Bioengineering.” We are grateful to Darya Avdanina for the help in yeast iron sensitivity and iron uptake assays. We thank the reviewers for their valuable comments, which helped us to improve the manuscript.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2018.00965/full#supplementary-material>

- Charron, M. J., Read, E., Haut, S. R., and Michels, C. A. (1989). Molecular evolution of the telomere-associated MAL loci of *Saccharomyces*. *Genetics* 122, 307–316.
- Coi, A. L., Bigey, F., Mallet, S., Marsit, S., Zara, G., Gladieux, P., et al. (2017). Genomic signatures of adaptation to wine biological ageing conditions in biofilm-forming flor yeasts. *Mol. Ecol.* 26, 2150–2166. doi: 10.1111/mec.14053
- Coi, A. L., Legras, J.-L., Zara, G., Dequin, S., and Budroni, M. (2016). A set of haploid strains available for genetic studies of *Saccharomyces cerevisiae* flor yeasts. *FEMS Yeast Res.* 16:fow066. doi: 10.1093/femsyr/fow066
- Dean, E. J., Davis, J. C., Davis, R. W., and Petrov, D. A. (2008). Pervasive and persistent redundancy among duplicated genes in yeast. *PLoS Genet.* 4:e1000113. doi: 10.1371/journal.pgen.1000113
- DeLuna, A., Vetsigian, K., Shores, N., Hegreness, M., Colón-González, M., Chao, S., et al. (2008). Exposing the fitness contribution of duplicated genes. *Nat. Genet.* 40, 676–681. doi: 10.1038/ng.123
- Dujon, B. (2010). Yeast evolutionary genomics. *Nat. Rev. Genet.* 11, 512–524. doi: 10.1038/nrg2811
- Dunn, B., Richter, C., Kvitek, D. J., Pugh, T., and Sherlock, G. (2012). Analysis of the *Saccharomyces cerevisiae* pan-genome reveals a pool of copy number variants distributed in diverse yeast strains from differing industrial environments. *Genome Res.* 22, 908–924. doi: 10.1101/gr.130310.111
- Eldarov, M. A., Kishkovskaia, S. A., Tanaschuk, T. N., and Mardanov, A. V. (2016). Genomics and biochemistry of *Saccharomyces cerevisiae* wine yeast strains. *Biochemistry* (81), 1650–1668. doi: 10.1134/S0006297916130046
- Fidalgo, M., Barrales, R. R., Ibeas, J. I., and Jimenez, J. (2006). Adaptive evolution by mutations in the FLO11 gene. *Proc. Natl. Acad. Sci. U.S.A.* 103, 11228–11233. doi: 10.1073/pnas.0601713103
- Fidalgo, M., Barrales, R. R., and Jimenez, J. (2008). Coding repeat instability in the FLO11 gene of *Saccharomyces yeasts*. *Yeast* 25, 879–889. doi: 10.1002/yea.1642
- Galani-Nikolaki, S., Kallithrakas-Kontos, N., and Katsanos, A. A. (2002). Trace element analysis of Cretan wines and wine products. *Sci. Total Environ.* 285, 155–163. doi: 10.1016/S0048-9697(01)00912-3
- Galeote, V., Novo, M., Salema-Oom, M., Brion, C., Valério, E., Gonçalves, P., et al. (2010). FSY1, a horizontally transferred gene in the *Saccharomyces cerevisiae* EC1118 wine yeast strain, encodes a high-affinity fructose/H<sup>+</sup> symporter. *Microbiology* 156, 3754–3761. doi: 10.1099/mic.0.041673-0
- Gallone, B., Steensels, J., Prahl, T., Soriaga, L., Saels, V., Herrera-Malaver, B., et al. (2016). Domestication and divergence of *Saccharomyces cerevisiae* beer yeasts. *Cell* 166, 1397–1410.e16. doi: 10.1016/j.cell.2016.08.020
- Garrison, E., and Marth, G. (2012). *Haplotype-Based Variant Detection from Short-Read Sequencing*. Available at: <http://arxiv.org/abs/1207.3907> [accessed August 7, 2017]
- Goto, K., Iwatuki, Y., Kitano, K., Obata, T., and Hara, S. (1990). Cloning and nucleotide sequence of the KHR killer gene of *Saccharomyces cerevisiae*. *Agric. Biol. Chem.* 54, 979–984. doi: 10.1271/bbb1961.54.979
- Hinrichs, A. S., Raney, B. J., Speir, M. L., Rhead, B., Casper, J., Karolchik, D., et al. (2016). UCSC data integrator and variant annotation integrator. *Bioinformatics* 32, 1430–1432. doi: 10.1093/bioinformatics/btv766



- Holmes, D. L., Lancaster, A. K., Lindquist, S., and Halfmann, R. (2013). Heritable remodeling of yeast multicellularity by an environmentally responsive prion. *Cell* 153, 153–165. doi: 10.1016/j.cell.2013.02.026
- Ishigami, M., Nakagawa, Y., Hayakawa, M., and Iimura, Y. (2006). FLO11 is the primary factor in flor formation caused by cell surface hydrophobicity in wild-type flor yeast. *Biosci. Biotechnol. Biochem.* 70, 660–666. doi: 10.1271/bbb.70.660
- Jagtap, U. B., Jadhav, J. P., Bapat, V. A., and Pretorius, I. S. (2017). Synthetic biology stretching the realms of possibility in wine yeast research. *Int. J. Food Microbiol.* 252, 24–34. doi: 10.1016/j.ijfoodmicro.2017.04.006
- Katoh, K., and Standley, D. M. (2014). MAFFT: iterative refinement and additional methods. *Methods Mol. Biol.* 1079, 131–146. doi: 10.1007/978-1-62703-646-7\_8
- Kishkovskaia, S. A., Eldarov, M. A., Dumina, M. V., Tanashchuk, T. N., Ravin, N. V., and Mardanov, A. V. (2017). Flor yeast strains from culture collection: genetic diversity and physiological and biochemical properties. *Appl. Biochem. Microbiol.* 53, 359–367. doi: 10.1134/S0003683817030085
- Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923
- Legras, J.-L., Erny, C., and Charpentier, C. (2014). Population structure and comparative genome hybridization of European flor yeast reveal a unique group of *Saccharomyces cerevisiae* strains with few gene duplications in their genome. *PLoS One* 9:e108089. doi: 10.1371/journal.pone.0108089
- Legras, J.-L., Merdinoglu, D., Cornuet, J.-M., and Karst, F. (2007). Bread, beer and wine: *Saccharomyces cerevisiae* diversity reflects human history. *Mol. Ecol.* 16, 2091–2102. doi: 10.1111/j.1365-294X.2007.03266.x
- Legras, J.-L., Moreno-García, J., Zara, S., Zara, G., García-Martínez, T., Mauricio, J. C., et al. (2016). Flor yeast: new perspectives beyond wine aging. *Front. Microbiol.* 7:503. doi: 10.3389/fmicb.2016.00503
- Lowe, T. M., and Chan, P. P. (2016). tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. *Nucleic Acids Res.* 44, W54–W57. doi: 10.1093/nar/gkw413
- Mardanov, A. V., Beletsky, A. V., Eldarov, M. A., Tanashchuk, T. N., Kishkovskaya, S. A., and Ravin, N. V. (2017). Complete mitochondrial genomes of three *Saccharomyces cerevisiae* flor strains. *Mitochondrial DNA B* 2, 849–850. doi: 10.1080/23802359.2017.1407699
- Mardanov, A. V., Beletsky, A. V., Eldarov, M. A., Tanashchuk, T. N., Kishkovskaya, S. A., and Ravin, N. V. (2018). Draft genome sequence of the wine yeast strain *Saccharomyces cerevisiae* I-328. *Genome Announc.* 6:e1520-17. doi: 10.1128/genomeA.01520-17
- Marsit, S., and Dequin, S. (2015). Diversity and adaptive evolution of *Saccharomyces* wine yeast: a review. *FEMS Yeast Res.* 15:fov067. doi: 10.1093/femsyr/fov067
- Marsit, S., Leducq, J.-B., Durand, É., Marchant, A., Filteau, M., and Landry, C. R. (2017). Evolutionary biology through the lens of budding yeast comparative genomics. *Nat. Rev. Genet.* 18, 581–598. doi: 10.1038/nrg.2017.49
- Marsit, S., Mena, A., Bigey, F., Sauvage, F. X., Couloux, A., Guy, J., et al. (2015). Evolutionary advantage conferred by an eukaryote-to-eukaryote gene transfer event in wine yeasts. *Mol. Biol. Evol.* 32, 1695–1707. doi: 10.1093/molbev/msv057
- Martin, M. (2011). Cutadapt removes adapter sequence from high-throughput sequencing reads. *EMBnet J.* 17, 10–12. doi: 10.14806/ej.17.1.200
- Martínez, P., Pérez Rodríguez, L., and Benítez, T. (1997). Velum formation by flor yeasts isolated from sherry wine. *Am. J. Enol. Vitic.* 48, 55–56.
- Martínez-Garay, C. A., de Llanos, R., Romero, A. M., Martínez-Pastor, M. T., and Puig, S. (2016). Responses of *Saccharomyces cerevisiae* strains from different origins to elevated iron concentrations. *Appl. Environ. Microbiol.* 82, 1906–1916. doi: 10.1128/AEM.03464-15
- Martínez-Pastor, M. T., Perea-García, A., and Puig, S. (2017). Mechanisms of iron sensing and regulation in the yeast *Saccharomyces cerevisiae*. *World J. Microbiol. Biotechnol.* 33:75. doi: 10.1007/s11274-017-2215-8
- McIlwain, S. J., Peris, D., Sardi, M., Moskvina, O. V., Zhan, F., Myers, K. S., et al. (2016). Genome sequence and analysis of a stress-tolerant, wild-derived strain of *Saccharomyces cerevisiae* used in biofuels research. *G3* 6, 1757–1766. doi: 10.1534/g3.116.029389
- Moreno-García, J., García-Martínez, T., Millán, M. C., Mauricio, J. C., and Moreno, J. (2015). Proteins involved in wine aroma compounds metabolism by a *Saccharomyces cerevisiae* flor-velum yeast strain grown in two conditions. *Food Microbiol.* 51, 1–9. doi: 10.1016/j.fm.2015.04.005
- Moreno-García, J., Mauricio, J. C., Moreno, J., and García-Martínez, T. (2017). Differential proteome analysis of a flor yeast strain under biofilm formation. *Int. J. Mol. Sci.* 18:E720. doi: 10.3390/ijms18040720
- Naumov, G. I., Naumova, E. S., and Michels, C. A. (1994). Genetic variation of the repeated MAL loci in natural populations of *Saccharomyces cerevisiae* and *Saccharomyces paradoxus*. *Genetics* 136, 803–812.
- Nishimura, A., Kotani, T., Sasano, Y., and Takagi, H. (2010). An antioxidative mechanism mediated by the yeast N-acetyltransferase Mpr1: oxidative stress-induced arginine synthesis and its physiological role. *FEMS Yeast Res.* 10, 687–698. doi: 10.1111/j.1567-1364.2010.00650.x
- Novo, M., Bigey, F., Beyne, E., Galeote, V., Gavory, F., Mallet, S., et al. (2009). Eukaryote-to-eukaryote gene transfer events revealed by the genome sequence of the wine yeast *Saccharomyces cerevisiae* EC1118. *Proc. Natl. Acad. Sci. U.S.A.* 106, 16333–16338. doi: 10.1073/pnas.0904673106
- Okonechnikov, K., Golosova, O., Fursov, M., and Ugene team (2012). Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics* 28, 1166–1167. doi: 10.1093/bioinformatics/bts091
- Outten, C. E., and Albetel, A.-N. (2013). Iron sensing and regulation in *Saccharomyces cerevisiae*: ironing out the mechanistic details. *Curr. Opin. Microbiol.* 16, 662–668. doi: 10.1016/j.mib.2013.07.020
- Paneque, P., Álvarez-Sotomayor, M. T., and Gómez, I. A. (2009). Metal contents in “Oloroso” sherry wines and their classification according to provenance. *Food Chem.* 117, 302–305. doi: 10.1016/j.foodchem.2009.04.006
- Peinado, R. A., and Mauricio, J. C. (2009). “Biologically aged wines,” in *Wine Chemistry and Biochemistry*, eds M. V. Moreno-Arribas and M. C. Polo (New York, NY: Springer), 81–101. doi: 10.1007/978-0-387-74118-5\_5
- Protchenko, O., Ferea, T., Rashford, J., Tiedeman, J., Brown, P. O., Botstein, D., et al. (2001). Three cell wall mannoproteins facilitate the uptake of iron in *Saccharomyces cerevisiae*. *J. Biol. Chem.* 276, 49244–49250. doi: 10.1074/jbc.M109220200
- Rambaut, A. (2012). *FigTree v1.4.3*. Available at: <http://tree.bio.ed.ac.uk/software/figtree>
- Rossignol, T., Dulau, L., Julien, A., and Blondin, B. (2003). Genome-wide monitoring of wine yeast gene expression during alcoholic fermentation. *Yeast* 20, 1369–1385. doi: 10.1002/yea.1046
- Shalem, O., Sharon, E., Lubliner, S., Regev, I., Lotan-Pompan, M., Yakhini, Z., et al. (2015). Systematic dissection of the sequence determinants of gene 3' end mediated expression control. *PLoS Genet.* 11:e1005147. doi: 10.1371/journal.pgen.1005147
- Shimoi, H., Sakamoto, K., Okuda, M., Atthi, R., Iwashita, K., and Ito, K. (2002). The awa1 gene is required for the foam-forming phenotype and cell surface hydrophobicity of sake yeast. *Appl. Environ. Microbiol.* 68, 2018–2025. doi: 10.1128/AEM.68.4.2018-2025.2002
- Sliwa, P., and Korona, R. (2005). Loss of dispensable genes is not adaptive in yeast. *Proc. Natl. Acad. Sci. U.S.A.* 102, 17670–17674. doi: 10.1073/pnas.0505517102
- Song, G., Dickins, B. J. A., Demeter, J., Engel, S., Dunn, B., and Cherry, J. M. (2015). AGAPE (Automated Genome Analysis Pipeline) for pan-genome analysis of *Saccharomyces cerevisiae*. *PLoS One* 10:e0120671. doi: 10.1371/journal.pone.0120671
- Soontornngun, N. (2017). Reprogramming of nonfermentative metabolism by stress-responsive transcription factors in the yeast *Saccharomyces cerevisiae*. *Curr. Genet.* 63, 1–7. doi: 10.1007/s00294-016-0609-z
- Sreenivasaprasad, S. (2000). “Isolation of fungal nucleic acids,” in *Nucleic Acid Protocols Handbook*, ed. R. Rapley (Totowa, NJ: Humana Press), 37–45. doi: 10.1385/1-59259-038-1:37
- Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313. doi: 10.1093/bioinformatics/btu033
- Stanke, M., and Morgenstern, B. (2005). AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Res.* 33, W465–W467. doi: 10.1093/nar/gki458
- Strope, P. K., Skelly, D. A., Kozmin, S. G., Mahadevan, G., Stone, E. A., Magwene, P. M., et al. (2015). The 100-genomes strains, an *S. cerevisiae* resource that illuminates its natural phenotypic and genotypic variation and emergence as an opportunistic pathogen. *Genome Res.* 25, 762–774. doi: 10.1101/gr.18553.8.114

- Takagi, H., Shichiri, M., Takemura, M., Mohri, M., and Nakamori, S. (2000). *Saccharomyces cerevisiae* sigma 1278b has novel genes of the N-acetyltransferase gene superfamily required for L-proline analogue resistance. *J. Bacteriol.* 182, 4249–4256. doi: 10.1128/JB.182.15.4249-4256.2000
- Voordeckers, K., De Maeyer, D., van der Zande, E., Vincés, M. D., Meert, W., Cloots, L., et al. (2012). Identification of a complex genetic network underlying *Saccharomyces cerevisiae* colony morphology. *Mol. Microbiol.* 86, 225–239. doi: 10.1111/j.1365-2958.2012.08192.x
- Walker, M. E., Nguyen, T. D., Liccioli, T., Schmid, F., Kalatzis, N., Sundstrom, J. F., et al. (2014). Genome-wide identification of the Fermentome; genes required for successful and timely completion of wine-like fermentation by *Saccharomyces cerevisiae*. *BMC Genomics* 15:552. doi: 10.1186/1471-2164-15-552
- Zara, G., Zara, S., Pinna, C., Marceddu, S., and Budroni, M. (2009). FLO11 gene length and transcriptional level affect biofilm-forming ability of wild flor strains of *Saccharomyces cerevisiae*. *Microbiology* 155, 3838–3846. doi: 10.1099/mic.0.028738-0
- Zhang, Z., Li, J., Zhao, X.-Q., Wang, J., Wong, G. K.-S., and Yu, J. (2006). KaKs\_Calculator: calculating Ka and Ks through model selection and model averaging. *Genomics Proteomics Bioinformatics* 4, 259–263. doi: 10.1016/S1672-0229(07)60007-2
- Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.
- Copyright © 2018 Eldarov, Beletsky, Tanashchuk, Kishkovskaya, Ravin and Mardanov. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Traditional Norwegian Kveik Are a Genetically Distinct Group of Domesticated *Saccharomyces cerevisiae* Brewing Yeasts

Richard Preiss<sup>1,2</sup>, Caroline Tyrawa<sup>1</sup>, Kristoffer Krogerus<sup>3,4</sup>, Lars Marius Garshol<sup>5</sup> and George van der Merwe<sup>1\*</sup>

<sup>1</sup> Department of Molecular and Cellular Biology, University of Guelph, Guelph, ON, Canada, <sup>2</sup> Escarpment Laboratories, Guelph, ON, Canada, <sup>3</sup> VTT Technical Research Centre of Finland, Espoo, Finland, <sup>4</sup> Department of Biotechnology and Chemical Technology, School of Chemical Technology, Aalto University, Espoo, Finland, <sup>5</sup> Independent Researcher, Rælingen, Norway

## OPEN ACCESS

### Edited by:

Isabel Sá-Correia,  
Universidade de Lisboa, Portugal

### Reviewed by:

Jean-luc Legras,  
Institut National de la Recherche  
Agronomique (INRA), France  
Jose Sampaio,  
Universidade Nova de Lisboa,  
Portugal

Eladio Barrio,  
Universitat de València, Spain

### \*Correspondence:

George van der Merwe  
gvanderm@uoguelph.ca

### Specialty section:

This article was submitted to  
Evolutionary and Genomic  
Microbiology,  
a section of the journal  
Frontiers in Microbiology

**Received:** 20 June 2018

**Accepted:** 21 August 2018

**Published:** 12 September 2018

### Citation:

Preiss R, Tyrawa C, Krogerus K, Garshol LM and van der Merwe G (2018) Traditional Norwegian Kveik Are a Genetically Distinct Group of Domesticated *Saccharomyces cerevisiae* Brewing Yeasts. *Front. Microbiol.* 9:2137. doi: 10.3389/fmicb.2018.02137

The widespread production of fermented food and beverages has resulted in the domestication of *Saccharomyces cerevisiae* yeasts specifically adapted to beer production. While there is evidence beer yeast domestication was accelerated by industrialization of beer, there also exists a farmhouse brewing culture in western Norway which has passed down yeasts referred to as kveik for generations. This practice has resulted in ale yeasts which are typically highly flocculant, phenolic off flavor negative (POF-), and exhibit a high rate of fermentation, similar to previously characterized lineages of domesticated yeast. Additionally, kveik yeasts are reportedly high-temperature tolerant, likely due to the traditional practice of pitching yeast into warm (>28°C) wort. Here, we characterize kveik yeasts from 9 different Norwegian sources via PCR fingerprinting, whole genome sequencing of selected strains, phenotypic screens, and lab-scale fermentations. Phylogenetic analysis suggests that kveik yeasts form a distinct group among beer yeasts. Additionally, we identify a novel POF- loss-of-function mutation, as well as SNPs and CNVs potentially relevant to the thermotolerance, high ethanol tolerance, and high fermentation rate phenotypes of kveik strains. We also identify domestication markers related to flocculation in kveik. Taken together, the results suggest that Norwegian kveik yeasts are a genetically distinct group of domesticated beer yeasts with properties highly relevant to the brewing sector.

**Keywords:** yeast, domestication, brewing, *Saccharomyces*, fermentation, kveik, ale

## INTRODUCTION

It is clear that human activity resulted in the domestication of *Saccharomyces cerevisiae* yeasts specifically adapted for beer production. Recently, it has been shown that present-day industrial beer yeasts have originated from a handful of domesticated ancestors, with one major clade, “Beer 1,” comprising the majority of German, British, and American ale yeasts, and another clade, “Beer 2,” which does not have geographic structure and are more closely related to wine yeasts (Gallone et al., 2016). In general, it appears that human selection of beer yeasts over the span of centuries has resulted in the evolution of mechanisms to: efficiently ferment wort sugars such as maltose and maltotriose via duplications of *MAL* genes; eliminate the production of phenolic off flavor (POF)

by frequent nonsense mutations in the genes *PADI* and *FDC1*, responsible for production of 4-vinylguaiacol (4-VG), thereby generating POF negative (POF-) strains, and; flocculate efficiently, thereby assisting in the downstream processing of the product (McMurrough et al., 1996; Brown et al., 2010; Steensels and Verstrepen, 2014; Gallone et al., 2016; Gonçalves et al., 2016).

Regardless of the region of origin, beer yeast was likely maintained and domesticated by reuse (repitching) as well as sharing amongst generations of brewers, resulting in many of the domesticated beer yeasts used in the present day (Gibson et al., 2007; Libkind et al., 2011; Steensels et al., 2014; Gallone et al., 2016). It must not be assumed, however, that the domestication of beer yeasts occurred solely within the confines of industrial breweries, as there were farmhouse brewing traditions predating the industrialization of beer across northern Europe (Nordland, 1969; Räsänen, 1975). These brewers used yeast strains they maintained themselves, and the same yeast was generally used for brewing and for baking. However, in Norway and Sweden, beer and unleavened breads predated leavened bread due to a lack of suitable grain (Visted and Stigum, 1971). Improvements in transportation and increasing economic specialization caused traditional farmhouse brewing to decline from the nineteenth century onwards, which coupled with the entry of commercial yeast likely led to the disappearance of many traditional brewing yeasts (Nordland, 1969).

A region where traditional yeast cultures are still being used is western Norway, where a number of farmhouse brewers have maintained the traditional yeasts of this region, some reportedly for hundreds of years (Figure 1; Nordland, 1969). Norwegian farmhouse ale is produced predominantly from malted barley and is typically hopped, and also infused with juniper branches (Nordland, 1969). The farmhouse beers themselves are typically referred to as maltøl or kornøl. Until recently the yeast cultures, referred to as kveik, a dialect term for yeast in this region, were geographically isolated and maintained only locally by traditional farmhouse brewers. It is hypothesized that kveik yeasts are domesticated, as beers produced using these yeasts are reported to be non-phenolic (POF-) and these yeasts are potentially capable of rapidly fermenting malt-derived sugars due to the reported short fermentation times. Also, much like domesticated beer yeasts, kveik yeasts are maintained and reused via serial repitching (Gibson et al., 2007; Garshol, 2014; Stewart, 2015).

However, there are some critical differences in the way kveik is used and maintained that may have influenced its adaptive evolution and consequently impacted the generation of specific phenotypic characteristics. First, kveik has historically been stored dried for extended time periods of up to 1 year or more (Nordland, 1969). Second, kveik is typically inoculated by pitching into barley wort of between 28 and 40°C (Supplementary Table S1), a very high fermentation temperature for beer yeast (Caspeta and Nielsen, 2015). The most common temperature cited in older sources is “milkwarm,” meaning the temperature of milk as it leaves the udder, which is about 35°C (Iacobsen, 1935; Nordland, 1969; Strese and Tollin, 2015). Third, this wort is often of high sugar content (up to ~1.080 SG/19.25°Plato, compared to a typical wort of 1.050

SG or 12.5°Plato), and the brewers prefer a short fermentation time, often of only 1–2 days before transferring to a serving vessel (Nordland, 1969; Garshol, 2014). Traditionally, in the areas from which the studied yeast cultures come, the wort would be made from home-made barley malts, as barley was the main crop in these areas, and also the preferred grain for brewing (Hasund, 1942). The yeast is typically collected from the foam of the fermenting beer, or from the bottom slurry after primary fermentation, and dried until its next usage (Nordland, 1969). If the yeast went bad or was too old, the brewer would borrow yeast from neighbors, often choosing those who were known for having good beer (Nordland, 1969). Taken together, this adaptive environment for kveik yeasts was somewhat different from most industrial ale yeasts, while still favoring the possible development of domesticated traits.

Remarkably, yeast logs, specifically created for the storage of kveik, can be dated at least as far back as A.D. 1621 (Nordland, 1969), suggesting that kveik reuse began well before this date, as presumably the yeast was being reused prior to the development of specialized technology for yeast storage. This lines up with, and potentially predates, recent predictive modeling of the timeline of modern yeast domestication around A.D. 1573–1604 (Gallone et al., 2016). Kveik may therefore be a group of beer yeasts which have been domesticated and maintained by a geographically isolated brewing tradition, parallel to industrial beer production.

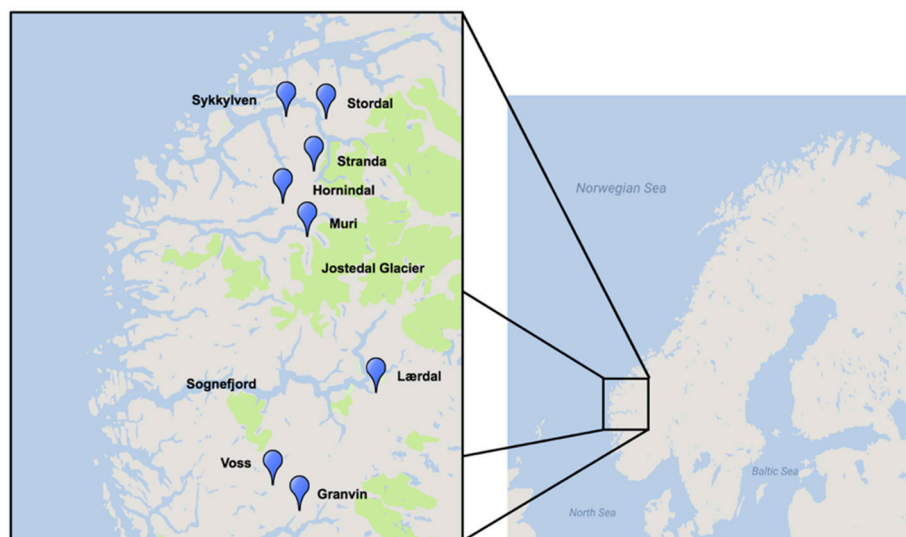
Yet, critically little is understood about kveik yeasts. While some of these yeasts have now been shared globally, there is a lack of empirical phenotypic and genotypic data pertaining to this intriguing group of beer yeasts. Here we report PCR fingerprinting and whole genome sequence data that suggest kveik yeasts form an interrelated group of beer yeasts genetically distinct from known domesticated beer yeasts. Our phenotypic characterizations and whole genome sequencing reveal evidence of domestication and positive characteristics in flavor compound production and stress tolerance that suggests the potential for kveik yeasts in a wide range of industrial applications.

## MATERIALS AND METHODS

### Yeast Strains

A total of 9 samples of Norwegian kveik and one additional Lithuanian farmhouse ale yeast sample were analyzed in the study. Seven kveik were supplied as liquid slurries, and two were supplied as dried yeast samples. The dried samples were rehydrated in sterile water. The liquid yeast slurries were enriched by inoculating 50 µl of the slurry into 5 mL YPD (1% yeast extract; 2% peptone; 2% dextrose). The samples were incubated at 30°C for 24 h with shaking, then streak plated onto Wallerstein Nutrient agar (WLN; Thermo Fisher CM0309), a differential medium for yeasts that distinguishes multiple yeasts from each other within one sample on the basis of uptake of the bromocresol green dye. Yeast colonies were then sub-streaked onto WLN to ensure purity. The resultant strains are summarized in Table 1. Additional control strains for the experiments are listed in Table 1.





**FIGURE 1 |** Geographical distribution of kveik yeast samples sourced for this project. Map was generated using Google Maps and Scribble Maps. Parks, including the Jostedalsbreen (Jostedal glacier) National Park are highlighted in green.

## DNA Extraction

DNA was extracted using an adaptation of a previously described method (Ausubel et al., 2002). Briefly, yeast cells were grown in 3 mL of YPD broth at 30°C, 170 rpm for 24 h, washed with sterile water, and pelleted. The cells were resuspended in 200  $\mu$ L of breaking buffer (2% Triton X-100, 1% SDS, 100 mM NaCl, 10 mM Tris-HCl). 0.3 g of glass beads and 200  $\mu$ L of phenol/chloroform/isoamyl alcohol was added and the samples were vortexed continuously at maximum speed for 3 min to lyse the cells. Following centrifugation, the aqueous layer was transferred to a clean tube and 1 mL of 100% ethanol was added. The supernatant was removed following another centrifugation step. The resulting pellet was resuspended in 400  $\mu$ L of 1X TE buffer and 30  $\mu$ L of 1 mg/mL DNase-free RNase A and incubated at 37°C for 5 min. The pellet was then washed with 1 mL of 100% ethanol and 10  $\mu$ L of 4 M ammonium acetate, followed by another wash with 1 mL of 70% ethanol, and then resuspended in 100  $\mu$ L of sterile ddH<sub>2</sub>O.

## PCR and ITS Sequencing

The internally transcribed spacer (ITS) regions of the yeast strains were amplified using ITS1 and ITS4 primers (Pham et al., 2011). PCR reactions contained 1  $\mu$ L of genomic DNA, 2.5  $\mu$ M of each primer, 0.4 mM dNTPs, 2.5 U of Taq DNA polymerase, and 1X Taq reaction buffer. The amplification reactions were carried out in a BioRad T100 Thermocycler under previously described conditions (Pham et al., 2011). PCR products were visualized on a 1% agarose gel in 1X TAE buffer to confirm successful amplification. The samples were purified using the QIAquick PCR purification kit and sequenced using an Applied Biosystems 3730 DNA analyzer. 4peaks software was used to perform quality control of sequence traces. The resulting sequences were analyzed for species-level homology using NCBI BLAST (blastn suite).

## DNA Fingerprinting

Yeast strains were identified by interdelta PCR fingerprinting using interdelta primers  $\delta 2$  (5'-GTGGATTTTATTCCAACA-3'),  $\delta 12$  (5'-TCAACAATGGAATCCCAAC-3'), and  $\delta 21$  (5'-CATCTTAACACCGTATATGA-3') (Ness et al., 1993; Legras and Karst, 2003). Primer pairs selected for further amplification and analysis were  $\delta 2 + \delta 12$  and  $\delta 12 + \delta 21$ , which both yielded the greatest range of well-resolved bands. PCR was carried out as follows: 4 min at 95°C, then 35 cycles of 30 s at 95°C, 30 s at 46°C, then 90 s at 72°C, followed by a final 10 min step at 72°C (Legras and Karst, 2003). Reaction products were confirmed through electrophoresis on a 1% agarose gel in 1X TAE buffer. PCR samples were then purified using a QIAquick PCR purification kit and analyzed on an Agilent 2100 Bioanalyzer using the Agilent DNA 7500 chip. Banding patterns obtained using Bioanalyzer were analyzed using GelJ software (Heras et al., 2015). Comparisons for each primer set ( $\delta 2 + \delta 12$  and  $\delta 12 + \delta 21$ ) were generated independently using the Comparison feature of the software, clustering the fingerprints using Pearson correlation and UPGMA (Heras et al., 2015). Resultant individual distance matrices were combined using fuse.plot in R (<https://github.com/andrewfletcher/fuse.plot>), which uses the hclust algorithm to format and fuse the matrices and perform hierarchical clustering with UPGMA. The data were visualized using FigTree software (<http://tree.bio.ed.ac.uk/software/figtree/>).

## DNA Content by Flow Cytometry

Flow cytometry was performed on six kveik strains to estimate ploidy essentially as described by Haase and Reed (2002). Cells were grown overnight in YPD medium, and  $\sim 1 \times 10^7$  cells were washed with 1 mL of 50 mM citrate buffer. Cells were then fixed with cold 70% ethanol, and incubated overnight at -20°C. Cells were then washed with 50 mM citrate buffer (pH



**TABLE 1** | Investigated yeast strains, source information, and sequence identification.

Strain name	Source	GenBank Accession #	References
Stordal Ebbegarden 1†	Jens Aage Øvrebust; Stordal, Norway	MG641161	This study
Stordal Ebbegarden 2	Jens Aage Øvrebust; Stordal, Norway	MG641162	This study
Stordal Framgarden 1	Petter B. Øvrebust; Stordal, Norway	MG641163	This study
Stordal Framgarden 2	Petter B. Øvrebust; Stordal, Norway	MG641164	This study
Granvin 1†	Hans Haugse; Granvin, Norway	MG641170	This study
Granvin 2	Hans Haugse; Granvin, Norway	MG641171	This study
Granvin 3	Hans Haugse; Granvin, Norway	MG754414	This study
Granvin 4	Hans Haugse; Granvin, Norway	MG709026	This study
Granvin 5	Hans Haugse; Granvin, Norway	MG709027	This study
Granvin 6	Hans Haugse; Granvin, Norway	MG709028	This study
Granvin 7	Hans Haugse; Granvin, Norway	MG709029	This study
Granvin 8	Hans Haugse; Granvin, Norway	MG709030	This study
Granvin 9	Hans Haugse; Granvin, Norway	MG709031	This study
Hornindal 1†	Terje Raftevd; Hornindal, Norway	MG641172	This study
Hornindal 2†	Terje Raftevd; Hornindal, Norway	MG641173	This study
Hornindal 3	Terje Raftevd; Hornindal, Norway	MG641174	This study
Joniškelis	Julius Simonaitis; Joniškėlis, Lithuania	MG719970	This study
Lærdal 1	Dagfinn Wendelbo; Lærdal, Norway	MG641175	This study
Lærdal 2†	Dagfinn Wendelbo; Lærdal, Norway	MG641176	This study
Muri*	Bjarne Muri; Olden, Norway	MG641177	This study
Stranda	Stein Langlo; Stranda, Norway	MG641165	This study
Sykkylven 1	Sigurd Johan Saure; Sykkylven, Norway	MG641166	This study
Sykkylven 2	Sigurd Johan Saure; Sykkylven, Norway	MG641167	This study
Voss 1†	Sigmund Gjernes; Voss, Norway	MG641168	This study
Voss 2	Sigmund Gjernes; Voss, Norway	MG641169	This study
BBY002 (Vermont Ale)†	Escarpment Laboratories; Canada	–	This study
WLP001†	White Labs; USA	–	This study; Rogers et al. (2016)
WLP002	White Labs; USA	–	This study
WLP007	White Labs; USA	–	This study; Kopecká et al. (2016)
WLP029	White Labs; USA	–	This study
WLP090	White Labs; USA	–	This study
WLP570	White Labs; USA	–	This study; Kopecká et al. (2016)
WLP585	White Labs; USA	–	This study
WLP590	White Labs; USA	–	This study
WLP045	White Labs; USA	–	This study
WLP050	White Labs; USA	–	This study
WY1007	Wyeast; USA	–	This study
WY1272	Wyeast; USA	–	This study
WY1318	Wyeast; USA	–	This study
WY2575	Wyeast; USA	–	This study
RC212	Lallemand; Canada	–	This study
EC1118	Lallemand; Canada	–	This study; Novo et al. (2009)
Idun_1	Idun Industri; Norway	–	This study
Idun_2	Idun Industri; Norway	–	This study
K701	Brewing Society of Japan, Japan	–	This study; Watanabe et al. (2013)
WildThing	Escarpment Laboratories; Canada	–	This study

Sequence identification was performed via ITS1-ITS4 rDNA amplification, sequencing, and BLAST. Strains selected for whole genome sequencing are indicated. \**Saccharomyces cerevisiae*/eubayanus/uvarum. All other strains are *Saccharomyces cerevisiae*. † Strain selected for whole genome sequence analysis.

7.2), resuspended in 50 mM citrate buffer containing 0.25 mg mL<sup>-1</sup> RNase A and incubated overnight at 37°C. 1 mg mL<sup>-1</sup> of Proteinase K was then added, and cells were incubated for 1 h at 50°C. Cells were then stained with SYTOX Green (2 µM; Life Technologies, USA), and their DNA content was determined using a FACSaria IIu cytometer (Becton–Dickinson, USA). DNA contents were estimated by comparing fluorescence intensities with those of *S. cerevisiae* haploid (CEN.PK113-1A) and diploid (CEN.PK) reference strains. One hundred thousand events were collected per sample during flow cytometry. Data was processed with the “flowCore” package (Hahne et al., 2009) in R, while mean peak fluorescence intensities were estimated with the “normalmixEM” function of the “mixtools” package (Benaglia et al., 2009) in R.

## Genome Sequencing and Analysis

The whole genomes of eight strains (six kveik strains and two commercial brewing strains as controls; see **Table 1**) were sequenced by Genome Québec (Montreal, Canada). In brief, DNA was isolated as described above, after which an Illumina TruSeq LT paired-end 150 bp library was prepared for each strain and sequencing was carried out with a HiSeqX instrument. Sequencing reads were quality-analyzed with FastQC (version 0.11.5) (Andrews, 2010) and trimmed and filtered with Trimmomatic (version 0.36; see **Supplementary Table S2** for parameters) (Bolger et al., 2014). Reads were aligned to a *S. cerevisiae* S288c (R64-2-1) reference genome using SpeedSeq (0.1.0) (Chiang et al., 2015). Quality of alignments was assessed with QualiMap (2.2.1) (García-Alcalde et al., 2012). Variant analysis was performed on aligned reads using FreeBayes (1.1.0-46-g8d2b3a0l; see **Supplementary Table S2** for parameters) (Garrison and Marth, 2012). Variants in all strains were called simultaneously (multi-sample). Prior to variant analysis, alignments were filtered to a minimum MAPQ of 50 with SAMtools (1.2; see **Supplementary Table S2** for parameters) (Li et al., 2009). Annotation and effect prediction of the variants was performed with SnpEff (1.2; see **Supplementary Table S2** for parameters) (Cingolani et al., 2012). Copy number variations of chromosomes and genes were estimated based on coverage with Control-FREEC (11.0; see **Supplementary Table S2** for parameters) (Boeva et al., 2012). Statistically significant copy number variations were identified using the Wilcoxon Rank Sum test ( $p < 0.05$ ). The median coverage and heterozygous SNP count over 10,000 bp windows was calculated with BEDTools (2.26.0) (Quinlan and Hall, 2010) and visualized in R.

## Phylogenetic and Population Structure Analysis

Prior to phylogenetic and population structure analysis, consensus genotypes for the sequenced strains were called from the identified variants using BCFtools (1.2) (Li, 2011). Because of the high levels of heterozygosity (>50,000 heterozygous SNPs) in the six kveik strains, haplotype phasing was also attempted using WhatsHap (0.14.1) (Martin et al., 2016). WhatsHap is a read-based phasing tool, that uses mapped sequencing reads spanning at least two heterozygous variants to infer phase. The consensus haplotypes were called from the phased variants using

BCFtools. Genome assemblies of the 157 *S. cerevisiae* strains described in Gallone et al. (2016) were retrieved from NCBI (BioProject PRJNA323691). In addition, the genome assembly of *Saccharomyces paradoxus* CBS432 was retrieved from [https://yix1217.github.io/Yeast\\_PacBio\\_2016/data/](https://yix1217.github.io/Yeast_PacBio_2016/data/) (Yue et al., 2017) to be used as an outgroup. Multiple sequence alignment of the consensus genotypes of the eight sequenced strains and the 158 assemblies was performed with the NASP pipeline (1.0.0) (Roe et al., 2016) using *S. cerevisiae* S288c (R64-2-1) as the reference genome. A matrix of single nucleotide polymorphisms (SNPs) in the 167 strains was extracted from the aligned sequences. The SNPs were annotated with SnpEff (Cingolani et al., 2012) and filtered as follows: only sites that were in the coding sequence of genes, present in all 167 strains and with a minor allele frequency >1% (one strain) were retained. The filtered matrix contained 4161584 SNPs (142120 sites). A maximum likelihood phylogenetic tree was estimated using IQ-TREE (1.5.5; see **Supplementary Table S2** for parameters) (Nguyen et al., 2015). IQ-TREE was run using the “GTR+F+R4” model and 1000 ultrafast bootstrap replicates (Minh et al., 2013). The resulting maximum likelihood tree was visualized in iTOL (Letunic and Bork, 2016) and rooted with *S. paradoxus* CBS432. The above steps from multiple sequence alignment onwards were repeated with the phased consensus haplotypes of the six kveik strains.

The population structure of 165 strains was investigated using the model-based algorithms in STRUCTURE (2.3.4; see **Supplementary Table S2** for parameters) (Pritchard et al., 2000) and fastStructure (1.0; see **Supplementary Table S2** for parameters) (Raj et al., 2014). Both tools were run on multiple threads using structure\_threader (1.2.4; see **Supplementary Table 2** for parameters) (Pina-Martins et al., 2017). The SNP matrix produced from the multiple sequence alignment was filtered using PLINK (1.9; see **Supplementary Table S2** for parameters) (Purcell et al., 2007) by removing sites in linkage disequilibrium (using a 50 SNP window size, 5 SNP step size, and pairwise threshold of 0.5) and with a minor allele frequency <5%. In addition, SNPs from *S. cerevisiae* S288c and *S. paradoxus* CBS432 were excluded from the population structure analysis. The thinned SNP matrix, now consisting of 26583 sites, was used as input to both STRUCTURE and fastStructure, which were run for 1 to 11 ancestral populations ( $K$ ). The SNP matrix is available as **Supplementary Data Sheet 1**. The STRUCTURE algorithm was run in 10 independent replicates for each  $K$  value and with an initial burn-in period of 100,000 iterations, followed by 100,000 iterations of sampling. The number of ancestral populations ( $K$ ) that best represented this dataset was chosen based on the “Evanno method” (Evanno et al., 2005; Earl and vonHoldt, 2012) for the STRUCTURE results with STRUCTURE HARVESTER and by the  $K$  value that maximized marginal likelihood for the fastStructure results (Raj et al., 2014). The STRUCTURE results were finally clustered with the online CLUMPAK server (Kopelman et al., 2015). Results were plotted in “distruct”-type plots in R. Principal component analysis of the thinned SNP matrix produced for population structure analysis was also performed using the SNPRelate package (Zheng et al., 2012). Nucleotide diversities within and between populations

were estimated in R using the PopGenome package (Pfeifer et al., 2014).

## Wort Preparation

Wort used for beer fermentations and yeast propagation was obtained from a commercial brewery, Royal City Brewing (Guelph, ON). The hopped wort was prepared using Canadian 2-row malt to an original gravity of 12.5°Plato (1.050 specific gravity). The wort was sterilized prior to use at 121°C for 20 min, and cooled to the desired fermentation or propagation temperature overnight.

## Propagation and Fermentation

Colonies from WLN plates were inoculated into 5 mL of YPD and grown at 30°C, 170 rpm for 24 h. The YPD cultures were transferred into 50 mL of sterilized wort and grown at 30°C, 170 rpm for 24 h. These cultures were counted using a haemocytometer and inoculated at a rate of  $1.2 \times 10^7$  cells/mL into 50 mL of sterilized wort in glass “spice jars” (glass jars of total volume 100 mL with straight sides) fitted with airlocks. These small-scale fermentations were performed in triplicate at 30°C for 12 days. 30°C was chosen as the fermentation temperature as it is a common temperature in Norwegian farmhouse brewing (Supplementary Table S1). The jars were incubated without shaking to best approximate typical beer fermentation conditions. Fermentation profiles were acquired by weighing the spice jars to measure weight loss, normalizing against water evaporation from the airlocks.

## Beer Metabolite Analysis

Following fermentation, samples were collected and filtered with 0.45 µm syringe filters prior to metabolite analysis. Flavor metabolite analysis was performed using HS-SPME-GC-MS (Rodriguez-Bencomo et al., 2012). Samples contained 2 mL of beer, 0.6 g of NaCl, 10 µL of 3-octanol (0.01 mg/mL), and 10 µL of 3,4-dimethylphenol (0.4 mg/mL). 3-octanol and 3,4-dimethylphenol were used as internal standards. The ethanol and sugar content was measured using HPLC and a refractive index (RI) detector. The samples were analyzed using an Aminex HPX-87H column, using 5 mM sulfuric acid as the mobile phase, under the following conditions: flow rate of 0.6 mL/min, 620 psi, and 60°C. Each sample contained 400 µL of filtered beer and 50 µL of 6% (v/v) isopropanol as the internal standard.

## Phenotypic Assays

To determine temperature tolerance, yeast grown for 24 h at 170 rpm at 30°C in YPD were subcultured into YPD pre-warmed to specified temperatures (30, 40, 42, 43, 45°C) in duplicate to an initial OD<sub>600</sub> of 0.1 and incubated with shaking for 20 h at the indicated temperature. To determine ethanol tolerance, yeast cultures grown for 24 h at 170 rpm at 30°C in YPD were sub-cultured into YPD containing increasing concentrations of ethanol (YPD + EtOH 10, 12, 14, 15, 16%) in duplicate to an initial OD<sub>600</sub> of 0.1 and incubated with shaking for 20 h at the indicated temperature. To assess growth yield for temperature tolerance and ethanol tolerance, the yeast samples were subjected to declumping using phosphoric acid and immediate OD<sub>600</sub>

measurements were taken using a spectrophotometer (Simpson and Hammond, 1989). To determine flocculation, yeast cultures were grown for 24 h at 170 rpm at 30°C in YPD, and then 0.5 mL was inoculated into 5 mL sterilized wort, which was incubated for 24 h at 170 rpm at 30°C. Flocculation was assessed using the spectrophotometric absorbance methodology of ASBC method Yeast-11 (ASBC, 2011). Values are expressed as % flocculance, with <20% representing non-flocculant yeast and >85% representing highly flocculant yeast.

## Statistical Analysis

Statistical analysis was performed on the fermentation, metabolite and phenotypic data with one-way ANOVA and Tukey's test using the “agricolae” package in R (<http://www.r-project.org/>). The results of the statistical tests are available as **Supplementary Data Sheet 2**.

## RESULTS

### Kveik Are a Genetically Distinct Group of Beer Yeasts

In order to determine whether original kveik samples contain multiple yeast strains, the kveik samples were first plated on WLN agar, which is a differential medium allowing for distinguishing of *Saccharomyces* on the basis of differences in colony morphology and uptake of the bromocresol green dye (Hutzler et al., 2015). We found that all but two of the kveik samples contained more than one distinct yeast colony morphology, corresponding to potentially unique strains. The number of strains isolated from individual kveik cultures thus ranged from 1 to 9 and totaled 25 and is summarized in **Table 1**.

Given that anecdotal reports stated kveik yeasts are often flocculent, demonstrate a fast fermentation rate, and are capable of utilizing malt sugars, all of which are hallmarks of domestication (Gallone et al., 2016), we aimed to determine the closest likely relatives of kveik yeasts among known strains of *S. cerevisiae*, and to determine whether kveik yeasts are related to each other. As nearly all domesticated ale yeasts belong to the *S. cerevisiae* species, we hypothesized that the kveik isolates also belong to *S. cerevisiae* (Almeida et al., 2015; Gallone et al., 2016; Gonçalves et al., 2016). We performed ITS sequencing and found that all but one kveik strain was identified (via BLAST search) as *S. cerevisiae* (**Table 1**). We found that the strain originating from Muri is most closely homologous to previously identified *S. cerevisiae/eubayanus/uvarum* triple hybrids, presenting this particular yeast strain as an intriguing potential domesticated hybrid warranting further investigation (**Table 1**).

Since the kveik yeasts appear to be *S. cerevisiae* strains, we next asked how they relate genetically to other *S. cerevisiae* yeasts. In order to answer this question, we performed interdelta PCR using the δ12/21 and δ2/12 primer sets (Legras and Karst, 2003; Hutzler et al., 2015). The δ elements are separated by amplifiable distances in the *S. cerevisiae* genome, and consequently interdelta PCR can be used to amplify interdelta regions, which in turn can be used to rapidly fingerprint yeasts for comparative genetic purposes (Legras and Karst, 2003; Hutzler et al., 2015).

**TABLE 2 |** Estimated ploidy, spore viability, mean sequencing coverage along *S. cerevisiae* S288c reference genome, and number of heterozygous single nucleotide polymorphisms (SNPs) in the six sequenced kveik strains.

Strain	Estimated ploidy	Spore viability (%)	Sequencing coverage (x)	Heterozygous SNPs
Granvin 1	3.93 ( $\pm 0.30$ )	56.5	946	65835
Hornindal 1	3.82 ( $\pm 0.29$ )	59.0	1221	67910
Hornindal 2	4.10 ( $\pm 0.23$ )	53.3	974	61402
Laerdal 2	4.03 ( $\pm 0.22$ )	40.6	472	59090
Stordal	3.92 ( $\pm 0.23$ )	5.9	671	54344
Ebbegarden 1				
Voss 1	3.88 ( $\pm 0.26$ )	63.4	1198	64959

Preliminary trials using the  $\delta 1/2$ ,  $\delta 2/12$ , and  $\delta 12/21$  primer sets showed that the latter two primer sets produced the greatest range of useful bands when separated via agarose gel electrophoresis. We then amplified the  $\delta 2/12$  and  $\delta 12/21$  regions of all the kveik strains and a selection of yeast strains representing “Beer 1” (German, American, UK), “Beer 2” (Belgian Saison), saké, wine, bread, wild, and distilling yeasts. Separation was performed using capillary gel electrophoresis (Agilent Bioanalyzer), which yielded greater accuracy and sensitivity (Hutzler et al., 2015). Analysis of both  $\delta 2/12$  and  $\delta 12/21$  datasets individually revealed that the kveik yeasts formed a subgroup among the other domesticated yeasts, such that the kveik yeasts appeared to be more closely related to each other than to other domesticated yeasts (**Supplementary Figure S1**). We next created a composite analysis of the interdelta datasets, yielding a dendrogram which placed some beer strains close together (**Supplementary Figure S2**). We found that a group of strains from German, British and American origin (WLP029, WLP002, WY1272, WLP007, BBY002) were represented in the dendrogram, and may represent the “Beer 1” clade (Belgian/German, British, American), as identified by Gallone et al. (2016). However, the kveik yeasts formed a group of related yeasts with a likely common ancestor. The kveik yeasts seem to be related to the beer strains more closely than other yeast groups. Furthermore, other yeasts from this study such as the hybrid Muri yeast, a Norwegian bread yeast (Idun) and the Lithuanian yeast strain (Joniškelis) do not appear to fit within the kveik family. Taken together these results suggest that kveik yeasts could represent a genetically distinct group of yeasts. While it does not properly resolve phylogeny due to lack of detail, the interdelta fingerprinting method can be used to assess which kveik yeasts are closely related to each other, and which could be selected for further sequencing analysis such that a representative range of strains are selected.

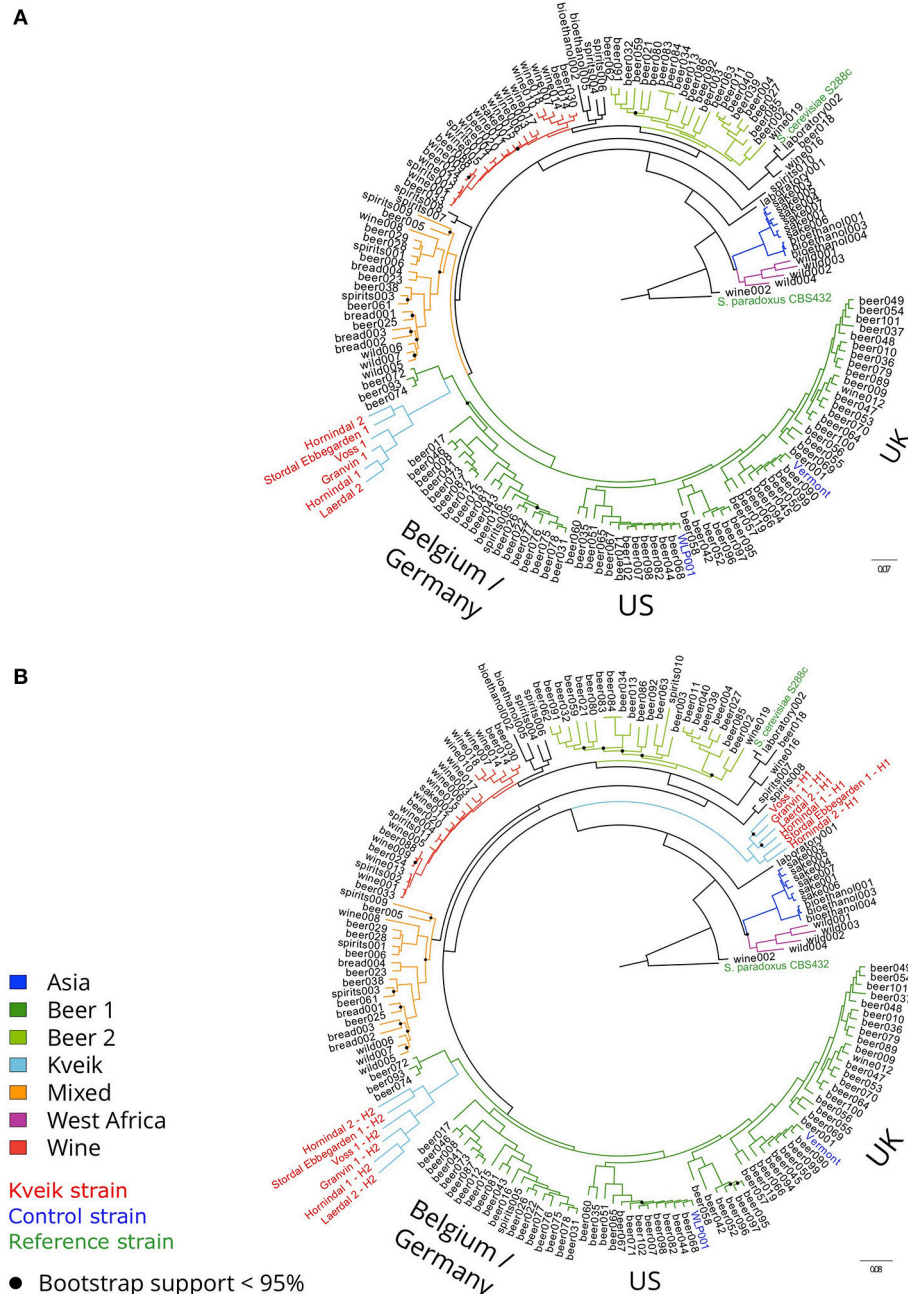
In order to better understand the genomics of kveik in relation to other *S. cerevisiae* yeasts, the whole genomes of six kveik strains (**Table 1**) were sequenced using 150 bp paired-end Illumina technology to an average coverage ranging from  $472\times$  to  $1,221\times$  (**Table 2**). These strains were selected based on the DNA fingerprinting results to represent different subgroups of the kveik family. In addition, two control strains (WLP001 and

Vermont Ale) were sequenced and included in the phylogenetic analysis. Flow cytometry and allele frequency distributions suggested that all six kveik strains were tetraploid (**Table 2**, **Supplementary Figures S3–S5**). However, 4/6 strains did show aneuploidy due to chromosomal CNVs, and of particular note, 3/6 strains containing an additional copy of chromosome IX. The kveik strains also showed high levels of heterozygosity, as the number of heterozygous SNPs ranged from  $\sim 54,000$  to 68,000 (**Table 2**). The heterozygous SNP density was relatively uniform in the strains, with few regions having undergone loss of heterozygosity (**Supplementary Figure S6**).

To examine the genetic relationship between kveik and other domesticated *S. cerevisiae* strains, phylogenetic and population structure analyses were performed together with genome sequences published elsewhere. First, the genome assemblies of the 157 *S. cerevisiae* strains investigated by Gallone et al. (2016) were retrieved from NCBI (PRJNA323691), while consensus genotypes of the six kveik and two control strains were produced from the SNPs and short InDels that were identified. After multiple sequence alignment and SNP identification, a filtered matrix containing 4161584 SNPs across 142120 sites was obtained (the SNP matrix is available as **Supplementary Data Sheet 1**). A maximum-likelihood phylogenetic tree was inferred from these polymorphic sites (**Figure 2A**). The main lineages reported in the original study (Gallone et al., 2016) were successfully reconstructed, and the two control strains clustered in the correct groups (“WLP001” in the “Beer 1–US” group, and “Vermont Ale” in the “Beer 1–UK” group). Consistent with the DNA fingerprinting results, the six kveik strains formed their own subgroup within the “Beer 1” group and appeared genetically distinct from other brewing yeasts, but closest to a group of German wheat beer yeasts known to contain mosaic genomes (beer072, 074, 093). To ensure that the high levels of heterozygosity in the six kveik strains wouldn’t skew the results, read-based phasing of the kveik strain haplotypes was also performed. The analysis was repeated for the two phased haplotypes (**Figure 2B**), and the phylogeny revealed that one haplotype again formed a subgroup within the “Beer 1” group, while the other haplotype formed a unique group between the “Asia” and “Mixed” groups. This is suggestive of a hybrid origin for kveik consisting of both a Beer 1 and an unknown lineage. However, Illumina paired-end data is not ideal for read-based phasing, as many pairs of heterozygous SNPs might not be connected by a read pair. Long read sequencing, e.g., using PacBio or Nanopore technology, could be used to improve the quality and length of the haplotype blocks (Martin et al., 2016). This in turn would allow for a more detailed analysis of the ancestry of the kveik strains.

Population structure analysis was also performed based on the polymorphic sites among the 165 strains. First, the SNP matrix was filtered to remove sites in linkage disequilibrium and with minor allele frequencies  $<5\%$ . The clustering algorithms STRUCTURE and fastSTRUCTURE were then used on the thinned SNP matrix (26583 sites), and the resulting population structure was in agreement with the estimated phylogeny. The number of populations that best represented this dataset was nine ( $K = 9$ ) for STRUCTURE (**Figure 3A**) and ten ( $K = 10$ ) for



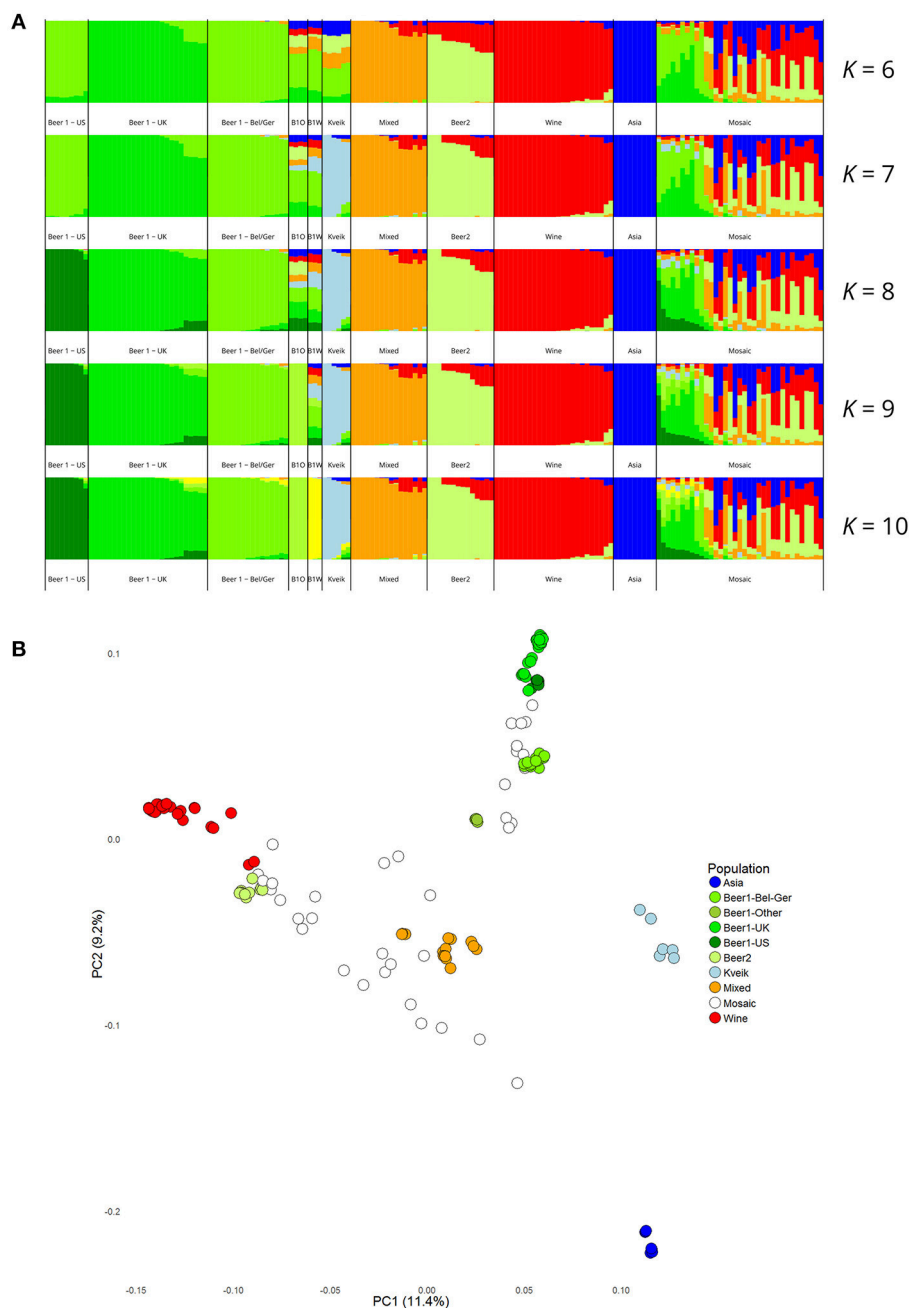


**FIGURE 2 |** Phylogeny of the six sequenced kveik strains compared with two control strains and the 157 *S. cerevisiae* strains sequenced in Gallone et al. (2016). **(A)** Maximum likelihood phylogenetic tree based on SNPs at 142120 sites in 166 *S. cerevisiae* strains (rooted with *S. paradoxus* as outgroup). Black dots on nodes indicate bootstrap support values <95%. Branches are colored according to lineage, and strain names are colored according to type (kveik, red; control, blue; reference, green). Branch lengths represent the number of substitutions per site. **(B)** Maximum likelihood phylogenetic tree produced as in **(A)**, but using the phased haplotypes of the kveik yeasts instead of their consensus genotypes.

fastStructure (**Supplementary Figure S7**). In both cases, the six kveik strains formed their own unique population, while the main populations reported in the Gallone et al. (2016) study were recreated. Even when the number of ancestral populations (*K*) was lowered to 7 or 8, the six kveik strains still formed a unique population (**Figure 3A**, **Supplementary Figure S7A**).

The fastStructure analysis was also repeated to include the phased haplotypes (**Supplementary Figure S7B**), which revealed an admixed ancestry (with contributions from Asia, Beer 1, Mixed, and Wine populations) for one haplotype (H1), and placed the other haplotype (H2) in a population with outliers in the “Beer 1” lineage (beer015, 052, 095–097). The kveik

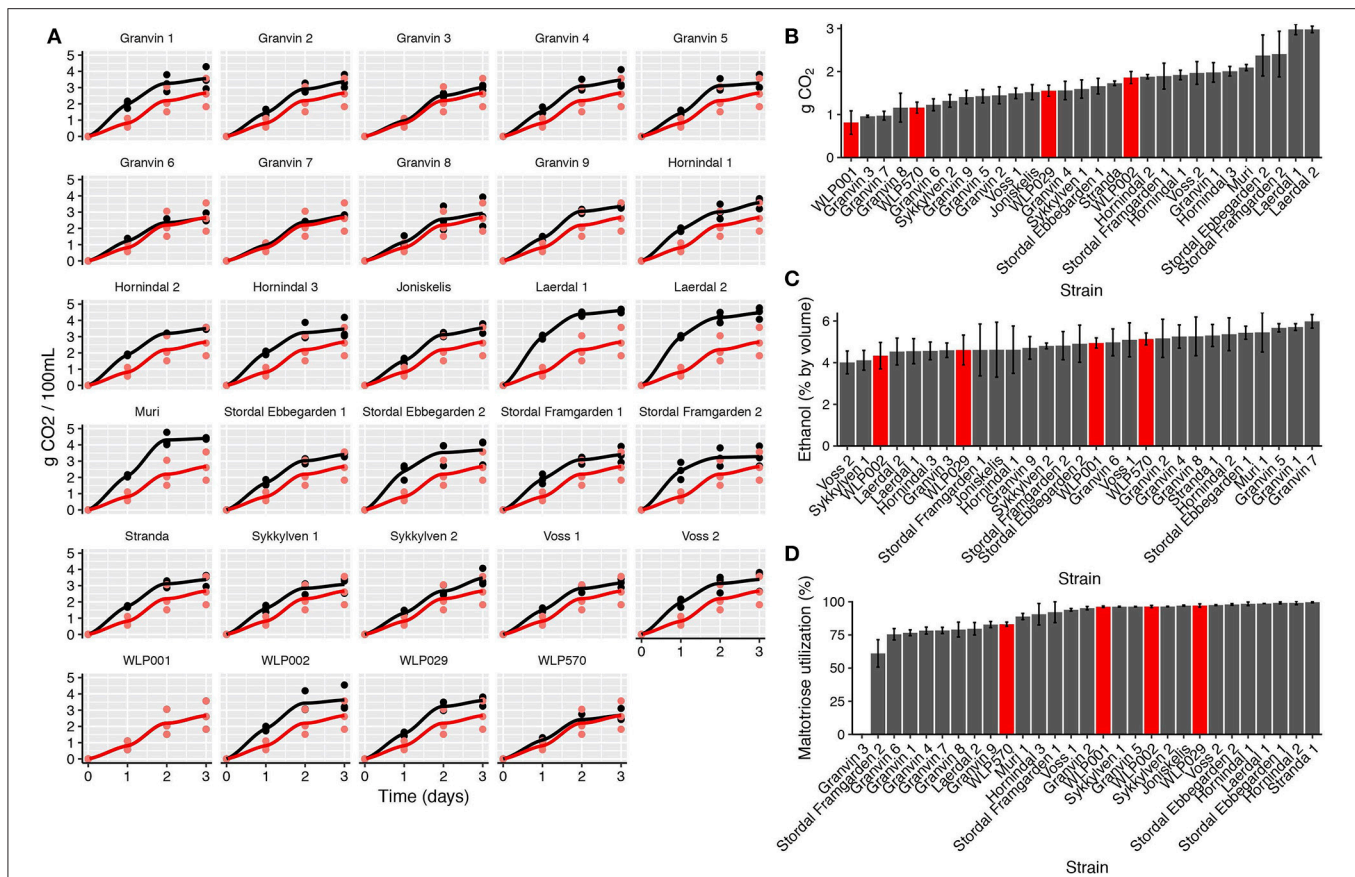




**FIGURE 3 |** Population structure of the six sequenced kveik strains and the 157 *S. cerevisiae* strains sequenced in Gallone et al. (2016). **(A)** Population structure of 163 *S. cerevisiae* strains estimated with *STRUCTURE* based on SNPs at 26583 sites. Each strain along the x-axis is represented by a vertical bar partitioned into colors based on estimated membership fractions to the resolved populations for  $K = 6, 7, 8, 9$ , and  $10$  assumed ancestral populations.  $K = 9$  best explains the data structure according to the “Evanno” method (Evanno et al., 2005). B10: Beer 1–Other. **(B)** Principal component analysis of SNPs at 26583 sites in 163 *S. cerevisiae* strains. Dots are colored by population.

haplotypes appear distinct from the German wheat beer yeasts, and the apparent connection of kveik to these yeasts suggested by the phylogeny is likely a coincidental artifact of both strain groups being mosaic/hybrid in origin. To support the population structure analysis, principal component analysis was performed on the thinned SNP matrix, which again clustered the six kveik

strains separately from the other strains (**Figure 3B**). The per-site nucleotide divergence between the kveik population and the other populations was also higher than those observed between the other beer populations (**Supplementary Table S3**). As suggested by the DNA fingerprinting results, compared to the other beer populations, relatively high nucleotide diversity

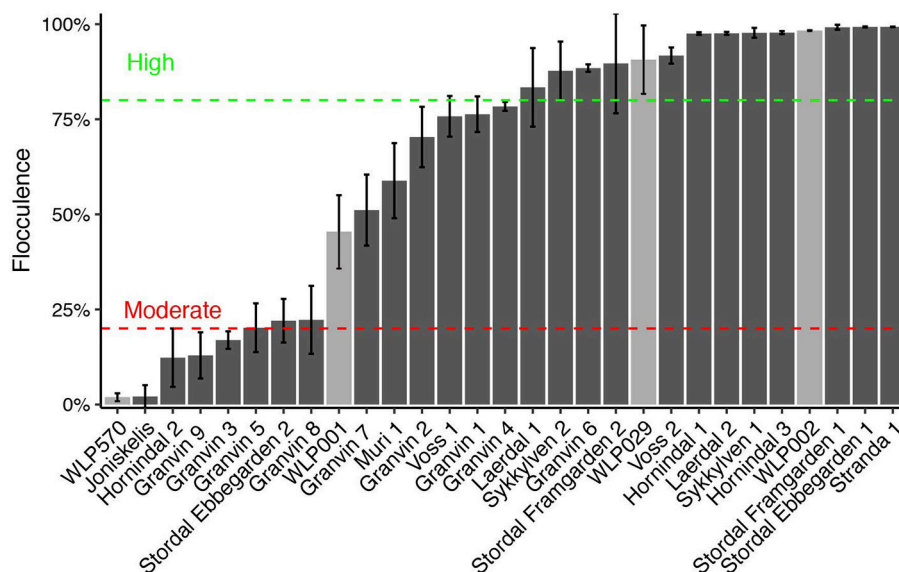


was also observed within the kveik population (**Supplementary Table S4**). Taken together, the results of the phylogenetic and population structure analysis suggest that the kveik strains selected for whole genome sequencing are genetically distinct from other domesticated yeasts.

## Brewing Characteristics, Domestication, and Sporulation Potential in Kveik

We next sought to analyze the brewing-relevant parameters of kveik yeasts in pure culture fermentation. Since Norwegian kveik cultures appear to often contain multiple yeast strains, there is the possibility that strains are interdependent. It is therefore important to determine the fermentation characteristics of individual strains as single culture fermentations would show whether individual kveik strains can adequately ferment beer. An inability to do so would suggest there is an adapted advantage to the multi-strain nature of kveik cultures. Additionally, we aimed to confirm anecdotal reports that these yeasts exhibit short lag phases and display good fermentation kinetics.

We performed test fermentations using the pure culture kveik strains as well as relevant industrial ale yeast controls (WLP001, WLP002, WLP029, WLP570; White Labs). In particular, WLP001 was chosen because it is one of the most popular ale strains for craft beer production. The fermentations were performed at 30°C which has been reported to be a typical temperature for beers fermented using kveik (Garshol, 2015). In order to assess the fermentation rate during the early phases of wort fermentation, we monitored the CO<sub>2</sub> loss in the fermentations via weighing. Using this technique, we observed that the fermentation curves for kveik was often favorable in comparison to the control strain with a shorter fermentation lag time observed in some of the strains (**Figures 4A,B**). Of the control strains, WLP002 produced the most CO<sub>2</sub> after 24 h. We found that 11 of the kveik strains outperformed WLP002 at 24 h, with the best-performing strain (Laerdal 2) producing 70.6% more CO<sub>2</sub> within the first 24 h of fermentation (**Figure 4B**). One-way ANOVA with Tukey's *post-hoc* test was performed and both Laerdal 1 and Laerdal 2 strains were determined to be significantly faster in this period at *P* < 0.05 (**Supplementary Data Sheet 2**).



**FIGURE 5 |** Flocculation capacity of kveik yeasts. Flocculation was assessed using the spectrophotometric absorbance methodology of ASBC Method Yeast-11. Values are expressed as %flocculence, with <20% representing non-flocculent yeasts, between 20 and 80% representing moderately flocculent yeast and >80% representing highly flocculent yeast. Strains are sorted in order of flocculence. Error bars represent SD,  $n = 3$ .

Following the 12-day fermentation and maturation period, we also measured the ethanol concentration of the beers using HPLC. The control ale strains produced ethanol values in the expected ascending order: WLP002 ( $4.33 \pm 0.64\%$ ), WLP029 ( $4.60 \pm 0.72\%$ ), WLP001 ( $4.94 \pm 0.25\%$ ), WLP570 ( $5.14 \pm 0.29\%$ ). We found that the kveik yeasts produced expected ethanol yields within the expected range for beer strains of *S. cerevisiae*, with apparent attenuation ranges spanning 60–90%, and ethanol yield ranging from  $4.01 \pm 0.55$  to  $5.98 \pm 0.32\%$  (Figure 4C). Statistically significant groupings among the ethanol data were not observed (Supplementary Data Sheet 2). The control data combined with the ethanol yield from the kveik yeasts in wort fermentation indicates that the kveik yeasts attenuate wort within the expected range of industrial domesticated ale strains.

Domesticated brewing yeasts are characterized by their ability to efficiently use maltose and maltotriose (Gallone et al., 2016; Gonçalves et al., 2016). These sugars constitute the majority of the fermentable sugars in brewer's wort. As has been observed previously in brewing strains (Gallone et al., 2016; Gonçalves et al., 2016), the six sequenced kveik strains showed considerable copy number variations in genes related to maltose and maltotriose transport (Table 3). Significant amplifications in the entire *MAL3x* locus (containing the *MAL31* permease, *MAL32* maltase and *MAL33* transcription factor) and the putative maltose-responsive transcription factor *YPR196W* were observed in particular. Indeed, we also observed maltotriose utilization across the kveik strains in the wort test fermentations, with exception to the Granvin 3 strain (Figure 4D).

To understand beer flavor contributions by the kveik yeasts, we also analyzed volatile aromatic compounds using HS-SPME-GC-MS (Table 4). Intriguingly, we found that all kveik

**TABLE 3 |** Estimated copy numbers of genes linked to maltose transport in the six sequenced kveik strains.

Strain	<i>MAL1x</i>	<i>MAL3x</i>	<i>IMA2</i>	<i>MPH2</i>	<i>MPH3</i>	<i>YPR196W</i>
Granvin 1	2	14	5	1	4	10
Hornindal 1	5	14	5	1	4	11
Hornindal 2	6	14	6	1	4	13
Laerdal 2	4	11	6	1	4	9
Stordal	5	11	6	0	4	14
Ebbegarden 1						
Voss 1	2	17	7	0	4	15

yeasts belonging to the main kveik genetic lineage (Figure 2, Supplementary Figures 1, 2) produced minimal levels of 4-vinylguaiacol (clove, smoke), suggesting that the kveik family are POF- (Table 4). Indeed, these levels were significantly different from the POF+ control strain (WLP570) in all but one kveik yeast (Muri kveik; Supplementary Data Sheet 2). Non-domesticated *S. cerevisiae* strains tend to have functional *PAD1* and *FDC1* genes, allowing them to decarboxylate hydroxycinnamic acids to vinylphenols (Mukai et al., 2010). Many brewing strains lack the ability to produce such off-flavors, and studies have shown that these strains carry loss-of-function mutations in either *PAD1* and *FDC1* (Mukai et al., 2014; Gallone et al., 2016; Gonçalves et al., 2016). The six kveik strains sequenced here indeed carry loss-of-function mutations in these two genes (Table 5). Three of these mutations, 305G>A in *PAD1*, 460C>T in *FDC1*, and 501insA in *FDC1*, have been observed previously in brewing strains (Mukai et al., 2014; Gallone et al., 2016; Gonçalves et al., 2016), and are widespread among the

**TABLE 4 |** Fermentation flavor metabolites (ppm) produced by kveik yeasts during wort fermentation at 30°C measured using HS-SPME-GC-MS.

	Ethyl Acetate	Ethyl Caproate	Ethyl Caprylate	Ethyl Decanoate	Ethyl Nonanoate	Hexanoic Acid	Isoamyl Acetate	Isoamyl Alcohol	Isobutanol	Phenethyl Acetate	Phenethyl Alcohol	4-Vinyl Guaiacol
Granvin 1	1.715	0.156	2.512	0.494	0.161	0.023	0.674	6.79	1.324	1.052	19.694	0.058
Granvin 2	3.118	0.366	4.555	0.455	0.197	0.01	0.781	7.879	1.527	1.87	21.603	0.012
Granvin 3	1.492	0.122	1.159	0.013	0.143	0.002	0.744	7.506	2.282	0.36	17.216	0.014
Granvin 4	1.195	0.059	0.232	0.012	0.025	0.004	0.467	4.719	1.126	0.257	15.163	0.043
Granvin 5	2.231	0.116	1.666	0.08	0.149	0.008	0.933	9.432	2.175	0.749	28.262	0.016
Granvin 6	3.2	0.365	5.005	0.88	0.238	0.02	0.905	9.046	1.9	1.36	24.966	0.016
Granvin 7	1.564	0.128	1.712	0.056	0.155	0.001	0.7	7.049	2.022	0.424	20.577	0.012
Granvin 8	1.229	0.056	0.299	0.026	0.028	0.003	0.538	5.423	1.344	0.298	14.628	0.043
Granvin 9	1.537	0.085	1.188	0.076	0.109	0.003	0.467	4.704	1.065	0.474	13.653	0.037
Hornindal 1	3.408	0.193	3.58	1.39	0.164	0.074	0.539	5.436	0.945	2.074	14.128	0.043
Hornindal 2	2.257	0.084	1.271	0.247	0.091	0.002	0.635	6.421	1.184	0.906	15.291	0.043
Hornindal 3	2.505	0.236	4.151	1.412	0.155	0.203*	0.556	5.659	0.838*	1.498	13.504	0.042
Joniskelis	1.495	0.117	2.301	1.277	0.151	0.055	0.589	5.942	1.018	1.568	17.63	0.223
Laerdal 1	1.838	0.315	4.124	0.891	0.204	0.116	0.453	4.689	0.624*	0.687	13.535	0.069
Laerdal 2	1.849	0.102	1.8	0.554	0.159	0.022	0.672	6.927	1.005	1.04	15.838	0.044
Muri	2.713	0.224	2.005	1.078	0.188	0.011	0.53	5.354	0.892	2.276	14.804	0.31
Stordal Ebbegarden 1	2.103	0.083	0.811	0.272	0.053	0.097	0.475	4.783	0.947	0.794	13.974	0.039
Stordal Ebbegarden 2	2.542	0.089	0.619	0.341	0.041	0.217*	0.677	7.052	1.135	1.074	16.637	0.049
Stordal Framgarden 1	2.395	0.168	2.975	0.772	0.158	0.058	0.55	5.536	0.901	1.635	15.809	0.052
Stordal Framgarden 2	2.654	0.44	4.112	0.753	0.176	0.006	0.593	5.998	0.976	0.864	14.03	0.047
Stranda	2.393	0.168	2.818	1.035	0.157	0.027	0.602	6.086	0.857	1.018	16.056	0.049
Sykkylven 1	2.046	0.101	1.306	0.427	0.08	0.005	0.483	4.883	0.867	0.749	14.28	0.043
Sykkylven 2	1.668	0.102	1.392	0.675	0.079	0.133	0.422	4.257	0.619*	0.622	12.081	0.044
Voss 1	2.156	0.209	3.317	0.618	0.145	0.006	0.463	4.651	0.941	0.825	12.377	0.039
Voss 2	2.364	0.307	3.059	0.347	0.157	0.005	0.519	5.225	1.01	1.148	15.121	0.039
WLP001	2.064	0.192	0.241	0.105	0.196	0.03	0.66	6.654	2.46	1.004	25.918	0.072
WLP002	0.735	0.076	0.537	0.047	0.101	0	0.81	8.168	4.062	0.478	19.481	0.053
WLP029	3.22	0.348	4.142	0.99	0.292	0.002	0.655	6.601	1.962	1.601	21.047	0.013
WLP570	5.734	0.806	8.586	1.583	0.424	0.019	1.395	14.057	2.106	3.529	33.427	0.299
Threshold (ppm)	30	0.21	0.9	0.2	0.85	8	1.2	70	100	3.8	100	0.3

Fermentations were performed in triplicate. Metabolite values are shaded if present in quantities at or above the stated sensory threshold values. Values presented are as mean ppm. Statistical analysis is available via **Supplementary Data Sheet 2**. Values marked with an asterisk are significantly different from all controls ( $P < 0.05$ , one-way ANOVA with Tukey's post-hoc test).

strains belonging to the “Beer 1” population (Gallone et al., 2016). Notably, a 232A>T mutation in *FDC1*, causing a premature stop codon at position 78, was also observed in the Stordal Ebbegarden 1 strain. To our knowledge, this loss-of-function mutation in *FDC1* has not been reported before.

Also, analysis of the volatile ester profiles revealed the kveik yeasts produced above-threshold concentrations of three yeast fatty acid esters: ethyl caproate (pineapple, tropical; threshold 0.21 ppm), ethyl caprylate (tropical, apple, cognac; threshold 0.9 ppm), and ethyl decanoate (apple; threshold 0.2 ppm) (Engan, 1972; Meilgaard, 1982; Verstrepen et al., 2003; Comuzzo et al., 2006). However, significant differences were not observed in the concentrations of these esters relative to the various control strains. Isoamyl acetate (banana; threshold 1.2 ppm) was detected above threshold and significantly higher in WLP570 only (**Supplementary Data Sheet 2**), indicating that this is not a major ester component in the flavor profile of the kveik

yeasts, or for the other industrial beer strains. Interestingly, loss-of-function mutations were identified in acetate ester-relevant genes *ATF1* and *ATF2* among 4/6 of the sequenced kveik strains (**Supplementary Table S5**). However, only one of these mutations was homozygous (“Laerdal 2”; homozygous *ATF1* lost stop codon) and was not linked to lower acetate ester formation in the beer fermentations (**Table 4**). Additionally, isobutanol levels were significantly lower among 3 kveik yeasts in comparison to the control ale strains, suggesting kveik may be capable of lower fusel alcohol production (**Table 4**, **Supplementary Data Sheet 2**).

We also analyzed the spore viability of the 6 sequenced kveik yeasts. Reasonable spore viability (40.6–63.4%) was observed in 5/6 of the strains, with one strain (“Stordal Ebbegarden 1”) showing low spore viability (**Table 2**). Interestingly, all sequenced kveik strains contain a loss of function mutation in *RMRI*, a protein required for meiotic recombination (Jordan et al.,



**TABLE 5 |** Loss-of-function single nucleotide polymorphisms in *PAD1* and *FDC1* in the six sequenced kveik strains.

Strain	<i>PAD1</i>		<i>FDC1</i>	
	305G>A	232A>T	460C>T	501insA
	Trp102*	Lys78*	Gln154*	Trp168fs
Granvin 1	0/0/0/1		0/0/0/1	1/1/1/1
Hornindal 1	1/1/1/1		1/1/1/1	1/1/1/1
Hornindal 2	1/1/1/1		1/1/1/1	1/1/1/1
Laerdal 2	0/1/1/1		0/0/0/1	1/1/1/1
Stordal Ebbegarden 1	0/0/0/1	0/1/1/1	0/0/0/1	0/0/0/1
Voss 1	1/1/1/1		1/1/1/1	1/1/1/1

\*, premature stop codon; ins, insertion; fs, frameshift.

2007). This mutation (726A>T causing lost stop codon) is only homozygous in the “Stordal Ebbegarden 1” strain, which may explain why this strain demonstrated low spore viability.

## Thermotolerance, Ethanol Tolerance, and Flocculation in Kveik

Since the initial fermentation trials demonstrated kveik yeasts are largely POF- and produce desirable fruity ester flavors, we next investigated the stress tolerance and flocculation of these yeasts to better determine their potential utility and to confirm these additional hallmarks of domestication. Given the reports of high-temperature fermentation by traditional Norwegian brewers (Nordland, 1969; Garshol, 2014), we monitored the growth of the kveik yeasts alongside known ale yeasts as control strains (WLP001; American ale, WLP029; German ale, WLP570; Belgian ale, WLP002; British ale) under normal and high temperature growth conditions (30–45°C).

We found that 19/25 kveik strains grew to >1.0 OD<sub>600</sub> at 40°C, while only 1/4 of the control ale strains (WLP570) grew to this optical density at 40°C (Table 6). Furthermore, 11/25 kveik strains grew to >0.4 OD<sub>600</sub> at 42°C, while only one of the control ale strains (WLP570) was able to. Remarkably, 19/25 kveik strains at least doubled its cell density at 43°C with the maximal optical density at this temperature observed for Laerdal 1 (OD<sub>600</sub> 0.44). Interestingly, one of the control strains (WLP570) also showed growth at 43°C (OD<sub>600</sub> 0.39). These data indicate that high temperature tolerance is common among kveik yeasts, and that high temperature tolerance is often limited among the American/British/German ale strains (Gallone et al., 2016). Notably, kveik strains displayed some growth up to 43°C, nearing the theoretical limit, and current technological upper threshold for *S. cerevisiae* cell growth (Caspeta et al., 2013, 2016; Caspeta and Nielsen, 2015). All strains failed to grow at 45°C (data not shown). A number of mutations in yeast have been linked to enhanced thermotolerance. In general, the kveik yeasts fell into statistical groupings between the WLP001/WLP002/WLP029 and WLP570 strains (Supplementary Data Sheet 2). We have observed heterozygous loss-of-function mutations in several genes relevant to thermotolerance, including *KEX1* (cell death protease; 4/6 sequenced kveik strains), *LRG1* (Rho1-specific

GTPase-activating protein and negative regulator of PKC-controlled cell wall integrity pathway; 6/6 sequenced kveik strains), *SWP82* (member of the SWI/SNF chromatin remodeling complex; 1/6 sequenced kveik strains), *RPI1* (modulates cell wall integrity; 6/6 sequenced kveik strains), *IRA1/IRA2* (GTPase-activating proteins and inhibitory regulators of the RAS-cAMP pathway; 6/6 and 1/6 sequenced kveik strains, respectively), and *CDC25* (membrane bound guanine nucleotide exchange factor and activator of RAS-cAMP pathway; 4/6 sequenced kveik strains) (Jones et al., 1991; Lorberg et al., 2001; Puria et al., 2009; Wallace-Salinas et al., 2015; Satomura et al., 2016; Huang et al., 2018; Supplementary Table S5).

We next investigated the ethanol tolerance of kveik yeasts in comparison to the control ale strains with ethanol tolerances available from the supplier (White Labs). Kveik and control strains were inoculated at 0.1 OD<sub>600</sub> into media containing from 10 to 16% ethanol and grown aerobically for 20 h. Our control data were in line with the suppliers’ broadly specified ethanol tolerances, e.g., WLP570 to “High–10 to 15%” and WLP002 to “Medium–5 to 10%” (Table 6). Interestingly WLP570, a Belgian-origin strain, showed high ethanol tolerance with evidence of growth up to 16% ethanol. Compared to the American, British and German-origin strains (WLP001, WLP002, WLP029, respectively) the kveik strains generally showed superior ethanol tolerance. 19/25 kveik strains at least doubled in density during the growth period at 14% ethanol, while 13/25 strains at least doubled in density during the growth period at 16% ethanol. Again, the kveik yeasts often fell into statistical groupings between the WLP001/WLP002/WLP029 and WLP570 strains (Supplementary Data Sheet 2). With exception to a number of strains originating from the Granvin sample, kveik yeasts display high levels of ethanol tolerance, suggesting that ethanol tolerance is generally conserved among kveik yeasts and may be a domestication signature of this yeast group. Supporting the phenotypic data, we observed a number of mutations relevant to ethanol tolerance in the sequenced kveik strains (Supplementary Table S5). Among these are *AGP2* (heterozygous, 6/6 strains), *PCA1* (heterozygous, 6/6 strains), and *VPS70* (heterozygous, 6/6 strains) (Teixeira et al., 2009; Voordeckers et al., 2015).

Flocculation is a hallmark of yeast domestication, as this property enhances the brewer’s ability to harvest yeast via either top or bottom cropping in the fermenter. We assessed the flocculence of the kveik yeasts using the absorbance method of ASBC Yeast-11 Flocculence method of analysis (ASBC, 2011). The control strains produced expected flocculence values: for example, the Belgian strain (WLP570) is non-flocculant (2%) and the British strain (WLP002) is highly flocculant (98%) (Figure 5). We observed high levels of flocculation among the kveik yeasts, but this property was not universal: 12/24 strains had flocculence values >80% (highly flocculant), while others showed very low flocculence (<20%; 4 strains). Interestingly, in most kveik samples containing more than one strain, at least one of the strains showed high flocculation rates above 80% (Figure 5). It is possible that in the original kveik mixed *S. cerevisiae* cultures, the yeasts undergo co-flocculation and consequently some strains never developed or needed this function (Smukalla et al., 2008;



TABLE 6 | Thermotolerance and ethanol tolerance in kveik yeasts.

	Temperature (°C)				Ethanol (% v/v)			
	30	40	42	43	10	12	14	16
WLP570	2.00	1.80	0.51	0.39	1.84	0.50	0.41	0.37
WLP001	1.93	0.14	0.13	0.12	0.80	0.48	0.34	0.14
WLP002	1.90	0.21	0.11	0.11	0.56	0.11	0.11	0.10
WLP029	1.96	0.17	0.11	0.11	0.50	0.40	0.10	0.10
Granvin 1	1.86	<b>1.53</b>	0.42	0.35	<b>1.18</b>	0.42	0.10	0.10
Granvin 2	1.92	<b>1.40</b>	<b>0.36</b>	<b>0.28</b>	<b>1.44</b>	0.55	0.45	0.25
Granvin 3	1.95	<b>1.53</b>	0.45	<b>0.31</b>	0.72	<b>0.27</b>	0.12	0.10
Granvin 4	1.87	<b>1.53</b>	0.23	<b>0.15</b>	0.70	0.38	<b>0.21</b>	0.10
Granvin 5	1.91	0.13	0.10	0.13	0.40	0.11	0.10	0.10
Granvin 6	1.84	1.74	0.41	0.40	<b>1.63</b>	0.46	0.42	0.19
Granvin 7	1.82	<b>0.70</b>	<b>0.31</b>	<b>0.25</b>	0.74	0.33	0.16	0.10
Granvin 8	1.84	0.14	0.13	0.13	0.62	0.21	0.10	0.10
Granvin 9	1.84	<b>0.84</b>	0.44	<b>0.22</b>	0.77	0.25	0.10	0.10
Hornindal 1	1.84	1.76	0.41	0.35	<b>1.39</b>	0.48	<b>0.30</b>	<b>0.27</b>
Hornindal 2	1.88	1.67	<b>0.33</b>	<b>0.26</b>	<b>1.12</b>	0.40	<b>0.32</b>	<b>0.22</b>
Hornindal 3	1.93	<b>1.49</b>	0.22	<b>0.19</b>	<b>1.47</b>	0.48	<b>0.27</b>	<b>0.29</b>
Joniskelis	1.88	<b>1.62</b>	0.56	<b>0.30</b>	1.70	0.62	<b>0.54</b>	0.37
Laerdal 1	1.83	<b>1.70</b>	0.48	0.44	1.79	0.50	0.40	0.33
Laerdal 2	1.86	<b>1.21</b>	0.45	0.33	<b>1.39</b>	0.47	0.39	<b>0.24</b>
Muri 1	1.96	<b>0.51</b>	<b>0.33</b>	0.30	0.93	0.47	0.49	<b>0.21</b>
Stordal Ebbegarden 1	1.81	<b>1.41</b>	<b>0.36</b>	<b>0.29</b>	0.73	0.47	0.47	0.34
Stordal Ebbegarden 2	1.91	<b>0.32</b>	<b>0.25</b>	<b>0.21</b>	0.72	0.39	0.27	0.10
Stordal Framgarden 1	1.97	1.64	<b>0.29</b>	<b>0.25</b>	<b>1.39</b>	0.60	0.41	0.32
Stordal Framgarden 2	1.84	1.72	<b>0.28</b>	<b>0.19</b>	<b>1.47</b>	0.61	0.44	0.33
Stranda	1.86	1.48	0.16	<b>0.18</b>	<b>1.14</b>	0.45	0.33	0.13
Sykkylven 1	1.87	1.78	0.46	<b>0.30</b>	1.70	0.51	<b>0.31</b>	<b>0.20</b>
Sykkylven 2	1.83	1.26	<b>0.26</b>	<b>0.26</b>	<b>1.01</b>	0.50	0.28	0.16
Voss 1	1.83	1.84	0.70	<b>0.30</b>	1.79	0.56	0.39	<b>0.22</b>
Voss 2	1.97	1.82	0.60	<b>0.24</b>	1.79	0.58	0.47	0.19

High temperature and ethanol tolerance assays were performed as described in materials and methods. OD<sub>600</sub> readings were obtained following 20 h of incubation in the specified conditions. Values represent the mean of biological replicates. Statistical analysis is available via **Supplementary Data Sheet 2**. Values marked in bold are significantly different from all controls ( $P < 0.05$ , one-way ANOVA with Tukey's post-hoc test).

Rossouw et al., 2015). Nonetheless, the high incidence of efficient flocculation among kveik yeasts is further support these yeasts have been domesticated. Copy number variations linked to flocculation genes (*FLO*) are common among domesticated yeasts (Dunn et al., 2012; Bergström et al., 2014; Gallone et al., 2016; Steenwyk and Rokas, 2017). Upon examination of the WGS data, we observed a high degree of copy number variation in *FLO* genes in the sequenced kveik strains (Table 7). Notably, the only strain with very low flocculence analyzed with whole genome sequencing (“Hornindal 2”; 12.3%) had a complete deletion of *FLO1*, known to be a critical gene conferring the flocculent phenotype (Vidgren and Londesborough, 2011). The flocculence of this strain was significantly lower ( $P < 0.05$ ) when compared to the Hornindal 1 strain. It is also worth noting that all kveik yeasts sequenced carry a 425A>G SNP in *FLO8* which causes a lost stop codon, restoring the functionality of *FLO8*, which is inactive in the S288c reference strain (Liu et al., 1996).

TABLE 7 | Estimated copy number variation among flocculation (*FLO*) genes in kveik.

Strain	<i>FLO1</i>	<i>FLO5</i>	<i>FLO8</i>	<i>FLO9</i>	<i>FLO10</i>	<i>FLO11</i>
Voss 1	1	0	4	2	0	5
Laerdal 2	3	1	4	2	0	2
Hornindal 1	2	1	4	2	0	5
Stordal Ebbegarden 1	1	1	4	2	0	1
Granvin 1	1	0	4	2	0	5
Hornindal 2	0	1	4	2	0	1

DISCUSSION

Here we present evidence which suggests kveik yeasts obtained from Norwegian farmhouse brewers represent a previously

undiscovered group of genetically distinct and domesticated beer yeasts, and that these yeasts have promising beer production attributes (Almeida et al., 2015; Baker et al., 2015; Gallone et al., 2016; Gonçalves et al., 2016). Our PCR fingerprint data suggested kveik yeast strains form a genetically distinct group of ale yeasts. Moreover, whole genome sequencing analysis of a representative group of 6 strains shows that kveik yeasts form a distinct group likely related to the “Beer 1” clade but with possible mixed ancestry when the separate haplotypes of the kveik yeasts are analyzed separately. The apparent conserved mixed ancestry of kveik is interesting given that mosaic/mixed-origin beer yeasts are not particularly common among either major beer yeast group (Gallone et al., 2016). Importantly, our analysis of Norwegian kveik yeasts suggests that the high-frequency production pressure of industrialization may not be necessary for domestication of brewing yeasts.

Our investigation of the beer production attributes with small-scale fermentation trials, phenotypic screens and genome sequencing revealed the majority of the Norwegian kveik yeasts metabolize wort sugars quickly (with related CNVs in maltose-relevant genes), are POF- (with loss-of-function mutations in *PAD1* and *FDC1*), flocculate efficiently (with CNVs in the *FLO* and related genes), and are highly ethanol tolerant and thermotolerant (typically polygenic traits). The domestication phenotypes and genomic domestication markers in kveik largely line up with those of previously analyzed domesticated beer yeasts (Gallone et al., 2016; Gonçalves et al., 2016). Thus, it appears that kveik have been domesticated in a similar manner to modern industrialized ale yeasts. The increased production rates of early industrial breweries in the seventeenth to eighteenth century was previously proposed to provide the foundation for beer yeast domestication (Gallone et al., 2016). Here we show kveik yeasts, surprisingly, have similar adaptation characteristics to the beer fermentation environment despite presumably being domesticated by farmhouse brewers without the high-frequency production pressure of an industrial brewing environment (Gallone et al., 2016). Thus, it is possible that the high frequency beer production associated with industrialization was not the only mechanism of adaptation resulting in the domesticated beer yeasts used today. Whether or not similar, small scale brewing practices analogous to the Norwegian farmhouse brewing culture, resulted in the domestication of yeast strains in Beer 1 predating industrialization, is currently unknown. As more yeast genomic data become available, it may be possible to identify yeasts which are more closely related to kveik and better understand the timeline of domestication for these yeasts and for other domesticated beer yeasts.

Approximately one third of the kveik yeasts did not flocculate with high efficiency. This may be influenced by the procedure used by farmhouse brewers to harvest yeast for repitching, including harvesting at least some of the top-fermenting yeast cells where the evolutionary pressure to flocculate would be less. It is therefore not surprising that some kveik strains flocculate less efficiently than others. However, kveik may present a new model for understanding yeast co-flocculation given the ability for high flocculation in some but not all members of a mixed yeast culture (e.g., the Hornindal culture) (Nishihara et al., 2000; Stewart, 2015).

Wort fermentations revealed that kveik strains produce a range of fruity esters, with ethyl caproate, ethyl caprylate, ethyl decanoate, and phenethyl acetate present above detection threshold (Table 3), indicating that these yeasts can be used to produce beers with fruity character. How kveik yeasts compare to a broader range of industrial beer yeasts in terms of diversity and intensity of flavor production is currently unknown and is a limitation of the present study. We have shown the kveik ale yeasts have a broad range of wort attenuation values. As these yeasts are POF-, a desirable trait for the majority of beer styles (McMurrough et al., 1996), they also could have broad utility for ale production, with selection by the brewer in accordance with desired attenuation target values and flavor profiles.

Strikingly, our phenotypic screening revealed the favorable thermotolerance and ethanol tolerance of these yeasts in comparison to known domesticated beer yeasts. Long-term heat adaptation is particularly relevant to fermentation processes performed at elevated temperatures, including those used for industrial bioethanol production. Multiple molecular and cellular processes and targets have been identified in the adaptation of yeast to heat. A prior study investigating the adaptation of yeast to ~40°C over a prolonged period of time, identified SNPs in genes related to DNA repair, replication, membrane composition and membrane structure as specific genetic markers of thermotolerance (Caspeta et al., 2013). Similarly, we identify SNPs in: *CDC25*, *IRA1*, and *IRA2*, which are genes that regulate the RAS/cAMP/PKA pathway; *RPI1* and *LRG1*, which impacts cell wall integrity; and *KEX1* and *SWP82* (Puria et al., 2009; Wallace-Salinas et al., 2015; Peeters et al., 2017; Huang et al., 2018). These mutations could aid thermotolerance in kveik and be future routes for development of thermotolerant yeasts. This characteristic also has potential application in brewing, as wort inoculation at higher temperatures (>30°C) without compromise in flavor could help limit the expensive cooling needed to manage wort fermentation temperatures that are typically controlled at 18–22°C for ale fermentations (Hill, 2015).

We also demonstrate that ethanol tolerance, known to be a polygenic and genetically complex trait involving multiple alleles, is a common adaptation of kveik yeasts. While single genetic alterations can incrementally increase ethanol tolerance, it does not approach that of the polygenic/multiallelic phenotype (Lam et al., 2014; Snoek et al., 2016). High ethanol environments generally disrupt cell membrane structure and function, and impact protein folding. Not surprisingly, genes linked to ethanol tolerance are often associated with: stabilizing cell walls and cell membranes; increasing the protein folding capacity; maintaining the electrochemical gradient across the plasma membrane; and maintaining vacuolar function to mention a few (Lam et al., 2014; Snoek et al., 2016). Remarkably, almost one third of the kveik yeasts reported here could grow in the presence of 16% ethanol. Correspondingly, we observed mutations in genes linked to ethanol tolerance among the sequenced kveik strains, comprising *AGP2*, *PCA1*, and *VPS70* (Teixeira et al., 2009; Voordeckers et al., 2015). Interestingly, these mutations were always heterozygous. Given the ethanol and high temperature tolerances of kveik yeasts, it is possible these yeasts could benefit the distillation and bioethanol industries where these traits are desired (Caspeta et al., 2016).

It is now known that a broader selection of traditional Norwegian kveik yeasts are still in existence, and it is possible that other domesticated or “landrace yeasts” may exist in other geographic locations beyond Norway. Whole genome sequencing of additional kveik yeasts could better support geographical subgroups suggested in the present study. Furthermore, further detailed analysis into the individual kveik cultures (for example, screening more colonies) may reveal greater strain diversity than evident here. The apparent mosaic nature of the kveik genomes also warrants further investigation. To elucidate the ancestry of the kveik strains in more detail, one could apply the use of long read sequencing to improve the quality and length of the haplotype blocks during phasing and expanding the genome data set, e.g., with the recently published 1,011 yeast genomes (Peter et al., 2018), used for phylogenetic and population structure analysis. It is possible that through more detailed phenotypic screening and sequencing, particularly using long-read technology, a wider range of such yeasts may result in an expanded understanding of beer yeast domestication given the noted differences between farmhouse ale production (infrequent, non-commercial) vs. industrial ale production (frequent, commercial).

## DATA AVAILABILITY STATEMENT

The whole genome sequence datasets generated for this study can be found in the NCBI BioProject number PRJNA473622 (<https://www.ncbi.nlm.nih.gov/bioproject/PRJNA473622>).

## REFERENCES

- Almeida, P., Barbosa, R., Zalar, P., Imanishi, Y., Shimizu, K., Turchetti, B., et al. (2015). A population genomics insight into the Mediterranean origins of wine yeast domestication. *Mol. Ecol.* 24, 5412–5427. doi: 10.1111/mec.13341
- Andrews, S. (2010). *FastQC: A Quality Control Tool for High Throughput Sequence Data*. Babraham Institute.
- ASBC (2011). *Yeast—11. ASBC Methods of Analysis*, 1–2.
- Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A., et al. (2002). *Short Protocols in Molecular Biology, 5th Edn*. Hoboken, NJ: Wiley.
- Baker, E., Wang, B., Bellora, N., Peris, D., Hulfachor, A. B., Koshalek, J. A., et al. (2015). The genome sequence of *Saccharomyces eubayanus* and the domestication of lager-brewing yeasts. *Mol. Biol. Evol.* 32, 2818–2831. doi: 10.1093/molbev/msv168
- Benaglia, T., Chauveau, D., Hunter, D. R., and Young, D. (2009). mixtools : an R package for analyzing finite mixture models. *J. Stat. Softw.* 32, 1–29. doi: 10.18637/jss.v032.i06
- Bergström, A., Simpson, J. T., Salinas, F., Barré, B., Parts, L., Zia, A., et al. (2014). A high-definition view of functional genetic variation from natural yeast genomes. *Mol. Biol. Evol.* 31, 872–888. doi: 10.1093/molbev/msu037
- Boeva, V., Popova, T., Bleakley, K., Chiche, P., Cappel, J., Schleiermacher, G., et al. (2012). Control-FREEC: a tool for assessing copy number and allelic content using next-generation sequencing data. *Bioinformatics* 28, 423–425. doi: 10.1093/bioinformatics/btr670
- Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170
- Brown, C. A., Murray, A. W., and Verstrepen, K. J. (2010). Rapid expansion and functional divergence of subtelomeric gene families in yeasts. *Curr. Biol.* 20, 895–903. doi: 10.1016/j.cub.2010.04.027

## AUTHOR CONTRIBUTIONS

CT, RP, and KK conducted the experiments described in this study. KK performed bioinformatic analysis of the whole genome sequence data. RP, CT, and GM designed the experiments. LG contributed to introductory materials and supplied yeast cultures. RP, KK, and GM wrote the manuscript. All authors read and approved the final manuscript.

## FUNDING

This research was funded by an NSERC Discovery (#264792-400922) and OMAFRA-University of Guelph Gryphons LAAIR (LAAIR2017-5321) grants.

## ACKNOWLEDGMENTS

We thank Cam Fryer and Royal City Brewing for the donation of beer wort used in this study, and Brian Gibson for critical reading of the manuscript.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2018.02137/full#supplementary-material>

- Caspeta, L., Chen, Y., Ghiaci, P., Feizi, A., Buskov, S., Petranovic, D., et al. (2013). Altered sterol composition renders yeast thermotolerant. *Science* 944, 1–5. doi: 10.1126/science.1258137
- Caspeta, L., Chen, Y., and Nielsen, J. (2016). Thermotolerant yeasts selected by adaptive evolution express heat stress response at 30°C. *Sci. Rep.* 6:27003. doi: 10.1038/srep27003
- Caspeta, L., and Nielsen, J. (2015). Thermotolerant yeast strains adapted by laboratory evolution show trade-off at ancestral temperatures and preadaptation to other stresses. *MBio* 6:e00431-15. doi: 10.1128/mBio.00431-15
- Chiang, C., Layer, R. M., Faust, G. G., Lindberg, M. R., Rose, D. B., Garrison, E. P., et al. (2015). SpeedSeq: ultra-fast personal genome analysis and interpretation. *Nat. Methods* 12, 966–968. doi: 10.1038/nmeth.3505
- Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., et al. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly* 6, 80–92. doi: 10.4161/fly.19695
- Comuzzo, P., Tat, L., Tonizzo, A., and Battistutta, F. (2006). Yeast derivatives (extracts and autolysates) in winemaking: release of volatile compounds and effects on wine aroma volatility. *Food Chem.* 99, 217–230. doi: 10.1016/j.foodchem.2005.06.049
- Dunn, B., Richter, C., Kvitek, D. J., Pugh, T., and Sherlock, G. (2012). Analysis of the *Saccharomyces cerevisiae* Pan genome reveals a pool of copy number variants distributed in diverse yeast strains from differing industrial environments. *Genome Res.* 22, 908–924. doi: 10.1101/gr.130310.111
- Earl, D. A., and vonHoldt, B. M. (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv. Genet. Resour.* 4, 359–361. doi: 10.1007/s12686-011-9548-7
- Engan, S. (1972). Organoleptic threshold values of some alcohols and esters in beer. *J. Inst. Brew.* 78, 33–36. doi: 10.1002/j.2050-0416.1972.tb03425.x

- Evanno, G., Regnaut, S., and Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol. Ecol.* 14, 2611–2620. doi: 10.1111/j.1365-294X.2005.02553.x
- Gallone, B., Steensels, J., Prah, T., Soriaga, L., Saels, V., Herrera-Malaver, B., et al. (2016). Domestication and divergence of *Saccharomyces cerevisiae* beer yeasts. *Cell* 166, 1397.e16–1410.e16. doi: 10.1016/j.cell.2016.08.020
- García-Alcalde, F., Okonechnikov, K., Carbonell, J., Cruz, L. M., Götz, S., Tarazona, S., et al. (2012). Qualimap: evaluating next-generation sequencing alignment data. *Bioinformatics* 28, 2678–2679. doi: 10.1093/bioinformatics/bts503
- Garrison, E., and Marth, G. (2012). Haplotype-based variant detection from short-read sequencing. *arXiv [Preprint]*. arXiv:1207.3907, 9.
- Garshol, L. M. (2014). *Brewing with Kveik*. Available online at: <http://www.garshol.priv.no/blog/291.html>
- Garshol, L. M. (2015). *Brewing Raw Ale in Hornidal*. Available online at: <http://www.garshol.priv.no/blog/342.html>
- Gibson, B. R., Lawrence, S. J., Leclaire, J. P. R., Powell, C. D., and Smart, K. A. (2007). Yeast responses to stresses associated with industrial brewery handling. *FEMS Microbiol. Rev.* 31, 535–569. doi: 10.1111/j.1574-6976.2007.00076.x
- Gonçalves, M., Pontes, A., Almeida, P., Barbosa, R., Serra, M., Libkind, D., et al. (2016). Distinct domestication trajectories in top-fermenting beer yeasts and wine yeasts. *Curr. Biol.* 26, 2750–2761. doi: 10.1016/j.cub.2016.08.040
- Haase, S. B., and Reed, S. I. (2002). Improved flow cytometric analysis of the budding yeast cell cycle. *Cell Cycle* 1, 132–136. doi: 10.4161/cc.1.2.114
- Hahne, F., LeMeur, N., Brinkman, R. R., Ellis, B., Haaland, P., Sarkar, D., et al. (2009). flowCore: a Bioconductor package for high throughput flow cytometry. *BMC Bioinformatics* 10:106. doi: 10.1186/1471-2105-10-106
- Hasund, S. (1942). “Bygget som ‘ølkorn’ i det vestlandske havreområdet,” in *Or Norges Bondesoge, Glytt og granskninger*, Vol. 1 (Oslo), 208–226.
- Heras, J., Dominguez, C., Mata, E., Pascual, V., Lozano, C., Torres, C., et al. (2015). GelJ—a tool for analyzing DNA fingerprint gel images. *BMC Bioinformatics* 16:270. doi: 10.1186/s12859-015-0703-0
- Hill, A. E. (ed.). (2015). “Introduction to brewing microbiology,” in *Brewing Microbiology* (Cambridge: Woodhead Publishing), 27–29. doi: 10.1016/B978-1-78242-331-7.02001-3
- Huang, C., Lu, M., Chang, Y., and Li, W. (2018). Experimental evolution of yeast for high-temperature tolerance. *Mol. Biol. Evol.* 35, 1823–1839. doi: 10.1093/molbev/msy077
- Hutzler, M., Koob, J., Riedl, R., Schneiderbanger, H., Mueller-Auffermann, K., and Jacob, F. (2015). “Yeast identification and characterization,” in *Brewing Microbiology*, ed A. E. Hill (Cambridge: Woodhead Publishing), 65–104. doi: 10.1016/B978-1-78242-331-7.00005-8
- Jacobsen, C. (1935). *Ølbrygning i Gamle Dage, Den gamle by's Aarbog*. Aarhus: Købstadmuseet “Den Gamle By”.
- Jones, S., Vignais, M. L., and Broach, J. R. (1991). The CDC25 protein of *Saccharomyces cerevisiae* promotes exchange of guanine nucleotides bound to ras. *Mol. Cell. Biol.* 11, 2641–2646. doi: 10.1128/MCB.11.5.2641
- Jordan, P. W., Klein, F., and Leach, D. R. F. (2007). Novel roles for selected genes in meiotic DNA processing. *PLoS Genet.* 3, 2368–2380. doi: 10.1371/journal.pgen.0030222
- Kopecká, J., Némec, M., and Matoušková, D. (2016). Comparison of DNA-based techniques for differentiation of production strains of ale and lager brewing yeast. *J. Appl. Microbiol.* 120, 1561–1573. doi: 10.1111/jam.13116
- Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A., and Mayrose, I. (2015). Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. *Mol. Ecol. Resour.* 15, 1179–1191. doi: 10.1111/1755-0998.12387
- Lam, F. H., Ghaderi, A., Fink, G. R., and Stephanopoulos, G. (2014). Engineering alcohol tolerance in yeast. *Science* 346, 71–75. doi: 10.1126/science.1257859
- Legras, J.-L., and Karst, F. (2003). Optimisation of interdelta analysis for *Saccharomyces cerevisiae* strain characterisation. *FEMS Microbiol. Lett.* 221, 249–255. doi: 10.1016/S0378-1097(03)00205-2
- Letunic, I., and Bork, P. (2016). Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* 44, W242–W245. doi: 10.1093/nar/gkw290
- Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* 27, 2987–2993. doi: 10.1093/bioinformatics/btr509
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352
- Libkind, D., Hittinger, C. T., Valério, E., Gonçalves, C., Dover, J., Johnston, M., et al. (2011). Microbe domestication and the identification of the wild genetic stock of lager-brewing yeast. *Proc. Natl. Acad. Sci. U.S.A.* 108, 14539–14544. doi: 10.1073/pnas.1105430108
- Liu, H., Styles, C. A., and Fink, G. R. (1996). *Saccharomyces cerevisiae* S288C has a mutation in FLO8, a gene required for filamentous growth. *Genetics* 144, 967–978.
- Lorberg, A., Schmitz, H. P., Jacoby, J., and Heinisch, J. (2001). Lrg1p functions as a putative GTPase-activating protein in the Pkc1p-mediated cell integrity pathway in *Saccharomyces cerevisiae*. *Mol. Genet. Genomics* 266, 514–526. doi: 10.1007/s004380100580
- Martin, M., Patterson, M., Garg, S., Fischer, S. O., Pisanti, N., Klau, G. W., et al. (2016). WhatsHap: fast and accurate read-based phasing. *bioRxiv[Preprint]*. 1–18. doi: 10.1101/085050
- McMurrough, I., Madigan, D., Donnelly, D., Hurley, J., Doyle, A., Hennigan, G., et al. (1996). Control of ferulic acid and 4-Vinyl guaiacol in brewing. *J. Inst. Brewing* 102, 327–332. doi: 10.1002/j.2050-0416.1996.tb00918.x
- Meilgaard, M. C. (1982). Prediction of flavor differences between beers from their chemical composition. *J. Agric. Food Chem.* 30, 1009–1017. doi: 10.1021/jf00114a002
- Minh, B. Q., Nguyen, M. A. T., and Von Haeseler, A. (2013). Ultrafast approximation for phylogenetic bootstrap. *Mol. Biol. Evol.* 30, 1188–1195. doi: 10.1093/molbev/mst024
- Mukai, N., Masaki, K., Fujii, T., and Iefuji, H. (2014). Single nucleotide polymorphisms of PAD1 and FDC1 show a positive relationship with ferulic acid decarboxylation ability among industrial yeasts used in alcoholic beverage production. *J. Biosci. Bioeng.* 118, 50–55. doi: 10.1016/j.jbiosc.2013.12.017
- Mukai, N., Masaki, K., Fujii, T., Kawamukai, M., and Iefuji, H. (2010). PAD1 and FDC1 are essential for the decarboxylation of phenylacrylic acids in *Saccharomyces cerevisiae*. *J. Biosci. Bioeng.* 109, 564–569. doi: 10.1016/j.jbiosc.2009.11.011
- Ness, F., Lavalley, F., Dubourdieu, D., Aigle, M., and Dulaub, L. (1993). Identification of yeast strains using the polymerase chain reaction. *J. Sci. Food Agric.* 6, 89–94. doi: 10.1002/jsfa.2740620113
- Nguyen, L. T., Schmidt, H. A., Von Haeseler, A., and Minh, B. Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32, 268–274. doi: 10.1093/molbev/msu300
- Nishihara, H., Kio, K., and Imamura, M. (2000). Possible mechanism of co-flocculation between non-flocculent yeasts. *J. Inst. Brewing* 106, 7–10. doi: 10.1002/j.2050-0416.2000.tb00034.x
- Nordland, O. (1969). *Brewing and Beer Traditions in Norway*. Oslo: The Norwegian Research Council for Science and the Humanities.
- Novo, M., Bigey, F., Beyne, E., Galeote, V., Gavory, F., Mallet, S., et al. (2009). Eukaryote-to-eukaryote gene transfer events revealed by the genome sequence of the wine yeast *Saccharomyces cerevisiae* EC1118. *Proc. Natl. Acad. Sci. U.S.A.* 106, 16333–16338. doi: 10.1073/pnas.0904673106
- Peeters, K., Van Leemputte, F., Fischer, B., Bonini, B. M., Quezada, H., Tsytlonok, M., et al. (2017). Fructose-1,6-bisphosphate couples glycolytic flux to activation of Ras. *Nat. Commun.* 8:922. doi: 10.1038/s41467-017-01019-z
- Peter, J., De Chiara, M., Friedrich, A., Yue, J.-X., Pflieger, D., Bergström, A., et al. (2018). Genome evolution across 1,011 *Saccharomyces cerevisiae* isolates. *Nature* 556, 339–344. doi: 10.1038/s41586-018-0030-5
- Pfeifer, B., Wittelsbürger, U., Ramos-Onsins, S. E., and Lercher, M. J. (2014). PopGenome: an efficient swiss army knife for population genomic analyses in R. *Mol. Biol. Evol.* 31, 1929–1936. doi: 10.1093/molbev/msu136
- Pham, T., Wimalasena, T., Box, W. G., Koivuranta, K., Storgårds, E., Smart, K., et al. (2011). Evaluation of ITS PCR and RFLP for differentiation and identification of brewing yeast and brewery “Wild” yeast contaminants. *J. Inst. Brewing* 117, 556–568. doi: 10.1002/j.2050-0416.2011.tb00504.x
- Pina-Martins, F., Silva, D. N., Fino, J., and Paulo, O. S. (2017). Structure\_threader: an improved method for automation and parallelization of programs structure, fastStructure and MaverickK on multicore CPU systems. *Mol. Ecol. Resour.* 17, e268–e274. doi: 10.1111/1755-0998.12702
- Pritchard, J. K., Stephens, M., and Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics* 155, 945–959.



- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81, 559–575. doi: 10.1086/519795
- Puria, R., Mannan, M. A. U., Chopra-Dewasthaly, R., and Ganesan, K. (2009). Critical role of RPI1 in the stress tolerance of yeast during ethanolic fermentation. *FEMS Yeast Res.* 9, 1161–1171. doi: 10.1111/j.1567-1364.2009.00549.x
- Quinlan, A. R., and Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842. doi: 10.1093/bioinformatics/btq033
- Raj, A., Stephens, M., and Pritchard, J. K. (2014). fastSTRUCTURE: variational inference of population structure in large SNP data sets. *Genetics* 197, 573–589. doi: 10.1534/genetics.114.164350
- Räsänen, M. (1975). *Vom Halm zum Fass*. Helsinki: Kansatieteellinen arkisto.
- Rodriguez, M. E., Orozco, H., Cantoral, J. M., Matallana, E., and Aranda, A. (2014). Acetyltransferase SAS2 and sirtuin SIR2, respectively, control flocculation and biofilm formation in wine yeast. *FEMS Yeast Res.* 14, 845–857. doi: 10.1111/1567-1364.12173
- Rodriguez-Bencomo, J. J., Muñoz-González, C., Martín-Álvarez, P. J., Lázaro, E., Mancebo, R., Castañé, X., et al. (2012). Optimization of a HS-SPME-GC-MS procedure for beer volatile profiling using response surface methodology: application to follow aroma stability of beers under different storage conditions. *Food Anal. Methods* 5, 1386–1397. doi: 10.1007/s12161-012-9390-x
- Roe, C., Smith, D. E., Williamson, C. H. D., Aziz, M., Keim, P., Hepp, C. M., et al. (2016). NASP: an accurate, rapid method for the identification of SNPs in WGS datasets that supports flexible input and output formats. *Microb. Genom.* 2:e000074. doi: 10.1099/mgen.0.000074
- Rogers, C. M., Veatch, D., Covey, A., Staton, C., and Bochman, M. L. (2016). Terminal acidic shock inhibits sour beer bottle conditioning by *Saccharomyces cerevisiae*. *Food Microbiol.* 57, 151–158. doi: 10.1016/j.fm.2016.02.012
- Rossouw, D., Bagheri, B., Setati, M. E., and Bauer, F. F. (2015). Co-flocculation of yeast species, a new mechanism to govern population dynamics in microbial ecosystems. *PLoS ONE* 10:e0136249. doi: 10.1371/journal.pone.0136249
- Satomura, A., Miura, N., Kuroda, K., and Ueda, M. (2016). Reconstruction of thermotolerant yeast by one-point mutation identified through whole-genome analyses of adaptively-evolved strains. *Sci. Rep.* 6:23157. doi: 10.1038/srep23157
- Simpson, W. J., and Hammond, J. R. M. (1989). The response of brewing yeasts to acid washing. *J. Inst. Brewing* 95, 347–354. doi: 10.1002/j.2050-0416.1989.tb04642.x
- Smukalla, S., Caldara, M., Pochet, N., Beauvais, A., Guadagnini, S., Yan, C., et al. (2008). FLO1 is a variable green beard gene that drives biofilm-like cooperation in budding yeast. *Cell* 135, 726–737. doi: 10.1016/j.cell.2008.09.037
- Snoek, T., Verstrepen, K. J., and Voordeckers, K. (2016). How do yeast cells become tolerant to high ethanol concentrations? *Curr. Genet.* 62, 475–480. doi: 10.1007/s00294-015-0561-3
- Steensels, J., Snoek, T., Meersman, E., Nicolino, M. P., Voordeckers, K., and Verstrepen, K. J. (2014). Improving industrial yeast strains: exploiting natural and artificial diversity. *FEMS Microbiol. Rev.* 38, 947–995. doi: 10.1111/1574-6976.12073
- Steensels, J., and Verstrepen, K. J. (2014). Taming wild yeast: potential of conventional and nonconventional yeasts in industrial fermentations. *Annu. Rev. Microbiol.* 68, 61–80. doi: 10.1146/annurev-micro-091213-113025
- Steenwyk, J., and Rokas, A. (2017). Extensive copy number variation in fermentation-related genes among *Saccharomyces cerevisiae* wine strains. *G3* 7, 1475–1485. doi: 10.1534/g3.117.040105
- Stewart, G. G. (2015). Seduced by yeast. *J. Am. Soc. Brewing Chem.* 73, 1–21. doi: 10.1094/ASBCJ-2015-0202-01
- Strese, E. M., and Tollin, C. (2015). *Humle - Det Gröna Gullet*. Stockholm: Nordiska Museets Förlag.
- Teixeira, M. C., Raposo, L. R., Mira, N. P., Lourenco, A. B., and Sá-Correia, I. (2009). Genome-wide identification of *Saccharomyces cerevisiae* genes required for maximal tolerance to ethanol. *Appl. Environ. Microbiol.* 75, 5761–5772. doi: 10.1128/AEM.00845-09
- Verstrepen, K. J., Derdelinckx, G., Dufour, J. P., Winderickx, J., Thevelein, J. M., Pretorius, I. S., et al. (2003). Flavor-active esters: adding fruitiness to beer. *J. Biosci. Bioeng.* 96, 110–118. doi: 10.1016/S1389-1723(03)90112-5
- Vidgren, V., and Londesborough, J. (2011). 125th anniversary review: yeast flocculation and sedimentation in brewing. *J. Inst. Brewing* 117, 475–487. doi: 10.1002/j.2050-0416.2011.tb00495.x
- Visted, K., and Stigum, H. (1971). *Vår Gamle Bondekultur, Bind 2*. Oslo: Cappelen.
- Voordeckers, K., Kominek, J., Das, A., Espinosa-Cantú, A., De Maeyer, D., Arslan, A., et al. (2015). Adaptation to high ethanol reveals complex evolutionary pathways. *PLoS Genet.* 11:e1005635. doi: 10.1371/journal.pgen.1005635
- Wallace-Salinas, V., Brink, D. P., Ahrén, D., and Gorwa-Grauslund, M. F. (2015). Cell periphery-related proteins as major genomic targets behind the adaptive evolution of an industrial *Saccharomyces cerevisiae* strain to combined heat and hydrolysate stress. *BMC Genomics* 16:14. doi: 10.1186/s12864-015-1737-4
- Watanabe, D., Hashimoto, N., Mizuno, M., Zhou, Y., Akao, T., and Shimoi, H. (2013). Accelerated alcoholic fermentation caused by defective gene expression related to glucose derepression in *Saccharomyces cerevisiae*. *Biosci. Biotechnol. Biochem.* 77, 2255–2262. doi: 10.1271/bbb.130519
- Yue, J. X., Li, J., Aigrain, L., Hallin, J., Persson, K., Oliver, K., et al. (2017). Contrasting evolutionary genome dynamics between domesticated and wild yeasts. *Nat. Genet.* 49, 913–924. doi: 10.1038/ng.3847
- Zheng, X., Levine, D., Shen, J., Gogarten, S. M., Laurie, C., and Weir, B. S. (2012). A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics* 28, 3326–3328. doi: 10.1093/bioinformatics/bts606

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Preiss, Tyrawa, Krogerus, Garshol and van der Merwe. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.





# Aneuploidy and Ethanol Tolerance in *Saccharomyces cerevisiae*

Miguel Morard<sup>1,2</sup>, Laura G. Macías<sup>1,2</sup>, Ana C. Adam<sup>2</sup>, María Lairón-Peris<sup>2</sup>, Roberto Pérez-Torrado<sup>2</sup>, Christina Toft<sup>1,2†</sup> and Eladio Barrio<sup>1,2\*</sup>

<sup>1</sup> Departament de Genètica, Universitat de València, Valencia, Spain, <sup>2</sup> Departamento de Biotecnología, Instituto de Agroquímica y Tecnología de los Alimentos (IATA), CSIC, Valencia, Spain

## OPEN ACCESS

### Edited by:

Ed Louis,  
University of Leicester,  
United Kingdom

### Reviewed by:

Alfredo Ghezzi,  
University of Puerto Rico, Río Piedras  
Campus, Puerto Rico  
Juan Lucas Argueso,  
Colorado State University,  
United States

### \*Correspondence:

Eladio Barrio  
Eladio.Barrío@uv.es

### † Present address:

Christina Toft,  
Institute of Integrative and Systems  
Biology, I2SysBio, Universitat  
de València and CSIC, Valencia, Spain

### Specialty section:

This article was submitted to  
Evolutionary and Genomic  
Microbiology,  
a section of the journal  
Frontiers in Genetics

**Received:** 01 October 2018

**Accepted:** 28 January 2019

**Published:** 12 February 2019

### Citation:

Morard M, Macías LG, Adam AC,  
Lairón-Peris M, Pérez-Torrado R,  
Toft C and Barrio E (2019) Aneuploidy  
and Ethanol Tolerance  
in *Saccharomyces cerevisiae*.  
Front. Genet. 10:82.  
doi: 10.3389/fgene.2019.00082

Response to environmental stresses is a key factor for microbial organism growth. One of the major stresses for yeasts in fermentative environments is ethanol. *Saccharomyces cerevisiae* is the most tolerant species in its genus, but intraspecific ethanol-tolerance variation exists. Although, much effort has been done in the last years to discover evolutionary paths to improve ethanol tolerance, this phenotype is still hardly understood. Here, we selected five strains with different ethanol tolerances, and used comparative genomics to determine the main factors that can explain these phenotypic differences. Surprisingly, the main genomic feature, shared only by the highest ethanol-tolerant strains, was a polysomic chromosome III. Transcriptomic data point out that chromosome III is important for the ethanol stress response, and this aneuploidy can be an advantage to respond rapidly to ethanol stress. We found that chromosome III copy numbers also explain differences in other strains. We show that removing the extra chromosome III copy in an ethanol-tolerant strain, returning to euploidy, strongly compromises its tolerance. Chromosome III aneuploidy appears frequently in ethanol-tolerance evolution experiments, and here, we show that aneuploidy is also used by natural strains to enhance their ethanol tolerance.

**Keywords:** *Saccharomyces cerevisiae*, wine yeasts, chromosome III, aneuploidy, comparative genomics, ethanol tolerance

## INTRODUCTION

The yeast *Saccharomyces cerevisiae* is among the most beneficial microorganisms for humans, especially industrial strains involved in the production of fermented products, such as bread, beer or wine. *S. cerevisiae*, as well as other *Saccharomyces* species, are characterized by their ability to ferment simple sugars into ethanol, even when oxygen is available for aerobic respiration (Crabtree effect), due to an overflow in the glycolysis pathway (Hagman and Piškur, 2015). Although, alcohol fermentation is energetically less efficient than respiration, it provides a selective advantage to these yeasts to out-compete other microorganisms. This way, sugar resources are consumed faster and the ethanol produced during fermentation, as well as high levels of heat and CO<sub>2</sub>, can be harmful or less tolerated by their competitors. Once competitors are overcome, *Saccharomyces* yeasts can use the accumulated ethanol as a substrate for aerobic respiration in the presence of oxygen. This ecological strategy was named (ethanol) “make-accumulate-consume” (Thomson et al., 2005; Piškur et al., 2006).

With the advent of the human hunter-gatherer societies, *S. cerevisiae*, due to its fermentative capabilities, successfully occupied a new ecological niche in the crushed grape berries, collected by

humans to produce the first fermented beverages. With agriculture, Neolithic societies improved fermentations as a way to preserve their foods and beverages. Since then, human-associated *S. cerevisiae* yeasts have been exposed to selective pressures due to fluctuating stresses occurring during fermentations, such as osmotic stress, ethanol toxicity, anaerobic stress, acid stress, nutrient limitation, etc. (Querol et al., 2003). As a result of this passive domestication, human-associated *S. cerevisiae* yeasts exhibit differential adaptive traits and conform genetically separated populations (Gallone et al., 2016; Duan et al., 2018; Legras et al., 2018; Peter et al., 2018), according to their sources of isolation rather than to their geographic origins.

One of the most important selective pressures imposed to *S. cerevisiae* is ethanol tolerance. High ethanol concentrations also have a strong effect on *S. cerevisiae* yeast growth and metabolic efficiency (Ansanay-Galeote et al., 2001). Ethanol is a small amphipathic alcohol that can cross through cell membranes, increasing their fluidity and permeability, interfering in the folding and activity of proteins, and also affecting intracellular redox balance and pH homeostasis (reviewed in Auesukaree, 2017).

It is sometimes hard to differentiate between tolerance and resistance because they are defined in different ways depending on the research field. Most of the literature related to ethanol stress uses both concepts as synonymous to refer to the ability of yeasts to grow and survive in the presence of ethanol, although “ethanol tolerance” is the most frequently used (Snoek et al., 2016). In an attempt to differentiate these terms in microbiology, Brauner et al. (2016) defined resistance as the ability of a microorganism to grow in the presence of high concentrations of a drug, resulting in a higher minimal inhibitory concentration (MIC), and tolerance as the ability of the cell to survive the transient presence of a drug above the MIC. As ethanol is the main product of the *Saccharomyces* respire-fermentative metabolism, and, as mentioned, the basis of the “make-accumulate-consume” strategy, *Saccharomyces* yeasts acquired mechanisms to survive the transient presence of ethanol, and hence, we consider that the term “ethanol tolerance” would be more appropriate.

Different studies have been devoted to understand the molecular mechanisms responsible of yeast response and tolerance to ethanol (for a review see Snoek et al., 2016). However, ethanol tolerance is a multilocus trait, not well characterized, because genes related to ethanol tolerance are broadly distributed throughout the genome (Giudici et al., 2005). In fact, as many different cellular processes are affected by ethanol, more than 200 genes have been linked to ethanol tolerance. Therefore, although many efforts have been made, mechanisms of ethanol tolerance are not fully understood yet.

In recent years, researchers have looked at adaptation to different stresses (Yona et al., 2012; Voordeckers et al., 2015; Adamczyk et al., 2016), including ethanol, in non-tolerant yeast exposed to gradually increasing stress levels. An interesting outcome of these experiments was the fixation in yeast of different genome rearrangements of adaptive value (Gorter de Vries et al., 2017).

In a previous study, we determined significant differences in ethanol tolerance between natural and fermentative *S. cerevisiae* strains, including strains isolated from different sources, from wine to traditional fermentations of Latin America (Arroyo-López et al., 2010). In the present study, we have sequenced the genomes of the most and least ethanol-tolerant *S. cerevisiae* strains reported in Arroyo-López et al. (2010) study to determine if they differ in their chromosomal constitution.

## MATERIALS AND METHODS

### Strains and Sequencing

The *S. cerevisiae* strains used in this study are those exhibiting extreme differences in their ethanol tolerance (Arroyo-López et al., 2010). Temohaya-MI26 was isolated from the fermentation of Mezcal production in Durango, Mexico, and shows the lowest ethanol tolerance. Wine strain T73 was selected as a commercial dry yeast (Querol et al., 1992). It was isolated from a red wine fermentation in Alicante, Spain, and possesses an intermediate ethanol tolerance. Finally, strains CECT10094 and GBFlor-C are *flor* strains isolated from red Pitarra wine in Extremadura, Spain, and González Byass Sherry wine (Esteve-Zarzoso et al., 2001) in Jerez de la Frontera, Spain, respectively. They both exhibit the highest ethanol tolerances.

Yeast cells were grown in an overnight culture of GPY in 5 ml. Cells were pelleted in a microcentrifuge and suspended in 0.5 ml of 1 M sorbitol-0.1 M EDTA, pH 7.5. Then, they were transferred to a 1.5 ml microcentrifuge tube, with 0.02 ml of a solution of Zymolyase 60 (2.5 mg/ml). A microcentrifuge was used to spin down cells for 1 min, which were suspended in 0.5 ml of 50 mM Tris-HCl-20 mM EDTA, pH 7.4. After suspension, 0.05 ml of 10% sodium dodecyl sulfate was added and the mixture was incubated at 65°C for 30 min. Then, 0.2 ml of 5 M potassium acetate was added and the tubes were placed on ice for 30 min. Then they were centrifuged at maximum speed in a microcentrifuge for 5 min. Supernatant was transferred to a fresh microcentrifuge tube, and the DNA was precipitated by adding one volume of isopropanol. After incubation at room temperature for 5 min, the tubes were centrifuged for 10 min. The DNA was washed with 70% ethanol, vacuum dried, and dissolved in 50 µl of TE (10 mM Tris-HCl, 1 mM EDTA, pH 7.5). T73 was sequenced with a 300-bp paired-end library in an Illumina HiSeq 2500 equipment. Strains CECT10094, GBFlor-C and Temohaya-MI26 were sequenced with paired-end libraries of 100 bp with a mean insert size of 300 bp in an Illumina HiSeq 2500 instrument.

EC1118 sequencing data was downloaded from NCBI with identifiers: SRA, ERS484054; BioSample SAMEA2610549.

*S. cerevisiae* 2-200-2 is a diploid strain with a chromosome III trisomy used in the chromosome III removal experiment. This strain was obtained by Voordeckers et al. (2015) after evolving the haploid S288c-derivative strain FY5 in the presence of increasing levels of ethanol.

### Genomes Assembly and Annotation

Reads were trimmed with Sickle v1.2 (Joshi and Fass, 2011) with a minimum quality value per base of 28 at both ends and filtered at

a minimum read length of 85 bp. A first preassembly step was carried out with Velvet v1.2.03 (Zerbino and Birney, 2008) to determine the best k-mer value for each library. The assembly was done with Sopra v1.4.6 (Dayarian et al., 2010) integrated with Velvet with the k-mer value determined in the previous step. Then, refinement of the results was carried out with SSPACE v2.0 (Boetzer et al., 2011) and GapFiller v1.11 (Boetzer and Pirovano, 2012) to improve scaffold length and remove internal gaps. Several rounds of Sopra/SSPACE/GapFiller were performed until the number of scaffolds could not be reduced. At each step of the process, the scaffolds were aligned against the reference genome of *S. cerevisiae* S288C with Mauve v2.3.1 (Darling et al., 2010). These steps can lead to overfitting and the nature of our sequence data mean we cannot verify any new recombination events, so, they were manually corrected. The final scaffolds were then aligned against the S288C reference genome with MUMmer v3.07 (Kurtz et al., 2004) and ordered into chromosomes with an in-house script.

Genomes annotation was done using two different strategies. First, the annotation from S288C genome was used to transfer to the new genomes by sequence homology using RATT (Otto et al., 2011). Second, an *ab initio* gene prediction was performed using Augustus web server (Stanke and Morgenstern, 2005), to complete the annotation of low homology regions where RATT was not able to transfer annotation. Both results were merged and the annotation was then manually corrected using Artemis (Rutherford et al., 2000) to remove false gene discovery and incorrect RATT transfer were either removed or corrected dependent on the nature of the mistake (e.g., wrong placement, lack of intron, etc.).

## Variants Detection and Chromosome Copy Number Analysis

Mappings against the reference *S. cerevisiae* S288C genome (version R64-2-1) were done using bowtie2 v2.3.0 (Langmead and Salzberg, 2012) with default parameters. Read Depth (RD) for each position was computed with bedtools v2.17.0 (Hung and Weng, 2017). To smooth the representation of RD by chromosome, a sliding windows analysis was performed. Mean mapping reads was calculated for 10kb windows moving by 1,000 nt. Variant calling analysis was performed with breseq v0.27.1 (Barrick et al., 2014) pipeline with polymorphisms mode to enable heterozygotic variants to be called. Minimum polymorphism frequency was set to 0.15 to avoid low frequency variants calling. Variants annotation and manipulation was done with gdools v0.27.1 from breseq package. Variants whose frequency was higher than 0.95 were considered homozygotic and they were considered heterozygotic if it was lower. R and ggplot2 package were used for data representation.

## Phylogenetic Analysis

All gene sequences were extracted from the annotations of the genomes assembled, as well as sequences from 38 strains representatives of different known clades (**Supplementary Table S1**). Orthologous genes among *S. cerevisiae* strains were translated and aligned with MAFFT v7.221 (Katoh and

Standley, 2013). Then the alignments were back translated to nucleotides, and concatenated. Maximum Likelihood phylogeny was performed on the concatenated genes alignment with RAxML v8.1.24 (Stamatakis, 2014) with model GTR- $\Gamma$  and 100 bootstrap replicates. The concatenated-gene ML tree was drawn with R and ggtree package (Yu et al., 2017).

## Determination of the Ploidy by Flow Cytometry

The total DNA content of the strain of interest was estimated by flow cytometry analysis in a BD FACSVerse cytometer following the SYTOX Green method as described in Haase and Reed (2002). Ploidy levels were scored on the basis of fluorescence intensity compared to the reference haploid S288c and diploid FY1679 *S. cerevisiae* strains. The estimated ploidy of the strains was obtained from three independent measurements.

## Expression Analysis

The expression data from a previous work on ethanol response (Navarro-Tapia et al., 2016) was used in this study (GEO accession: GSE44863). In brief, transcriptomic analysis come from a microarray analysis after ethanol shock. Temohaya-MI26 and CECT10094 were subjected to a 10% ethanol treatment and RNA was extracted 1 and 10 h later. As a control, RNA was extracted after 1 and 10 h of growth without ethanol treatment. Samples were hybridized for each condition against a pool of all the samples from all the conditions in the analysis. Expression data is reported as the log2 of the ratio of signal intensities between each condition and the pool. After combining each replicate, the genes were assigned to chromosomes according to their systematic names. Wilcoxon-Test implemented in ggpubr package v0.1.7 was used for the statistical analysis of differences between the expression of the chromosomes.

## Aneuploidy Analysis in *S. cerevisiae* Strains

Ploidies, aneuploidy presence, and 15% ethanol tolerances were extracted from the recent study of 1011 *S. cerevisiae* genomes (Peter et al., 2018). The 1011 strains were grouped by ploidy level and the presence of chromosome III copy number variations. Wilcoxon paired test was used to test differences between euploid diploids and the rest of ploidies and aneuploidies. Relative growth rates are represented as the normalization of the ratio between growth on standard YPD at 30°C and the stress condition (Peter et al., 2018).

## Yeast Chromosome III Removal

A counter selectable marker (Kutyna et al., 2014) was used to remove a single copy of chromosome III in *S. cerevisiae* 2-200-2 strain (Voordeckers et al., 2015) to obtain a derivative strain with one less copy of chromosome III, named as 2-200-2-S4. An integrative cassette targeted to a wide intergenic region (YCR027C-YCR028C) was synthesized from pCORE5 vector using the following primers: CHRIIIdel\_F: CTGTAGCCATATTAAATTCCTTTGTCTCTGGACTCTTTTCG AGCCCCGATTTAGAGCTT and CHRIIIdel\_R: TTAAC

GTTCAGCAGCGTCAGTGAGAACTAAAATCATCCAATCTCGAGGTCGACGGTATCGAT. The 2-200-2 strain was transformed and colonies were selected in GPY with G418. Correct integration was corroborated by PCR using the following primers: test-CHRIII del\_F: TCGACATCATCTGCCCAGAT and test-CHRIII del\_R: ACTTAGGTGGAGGAGCAAG. After overnight growth in GPY (2% glucose, 0.5%, peptone, 0.5% yeast extract), cells were plated on galactose counter selection media (2% galactose, 0.5% peptone, 0.5% yeast extract), and colonies were used to measure chromosome III copy numbers.

## Chromosome III Copy Number Measurements

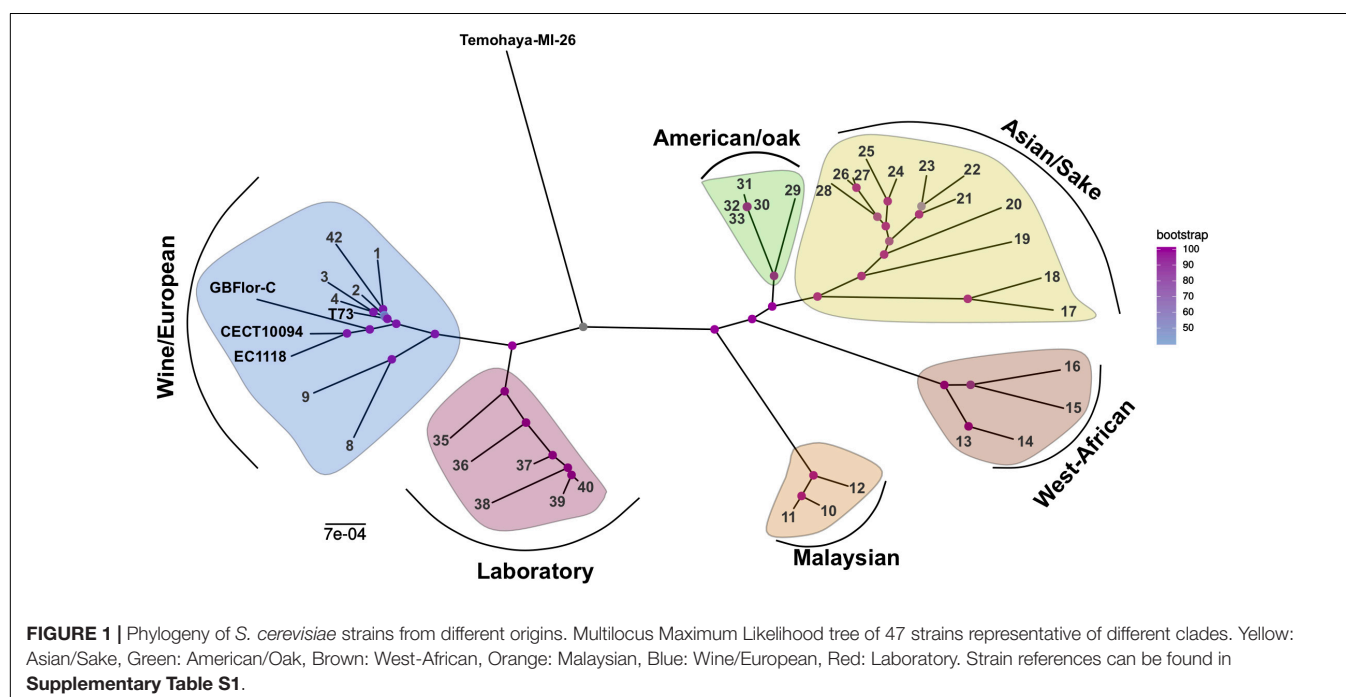
Genomic DNA was isolated and ethanol precipitated from the GPY liquid cell suspension in five independent culture replicates of 2-200-2 and 2-200-2-S4. DNA purity and concentration were determined in a NanoDrop ND-1000 spectrophotometer (Thermo-Scientific), and the integrity of all samples was checked by electrophoresis in agarose gel (1%). The PCR primers used to study the chromosome III copy number were designed from the available genomic sequence of *S. cerevisiae* strain S288C (R64-2-1, *Saccharomyces* genome database<sup>1</sup>). The sequences of PCR primer pairs used in this study are: ARE1-F: CCTCGTGTACCAGATCAAC; ARE1-R: AGGAAGATGGTGCCAATGAT; YCL001W-A-F: TGC TACGGTGGTTCTGCAAG; YCL001W-A-R: ACCACTGTGT CATCCGTTCT; POF1-F: TAATGGAGAGCTTCATGTCCGG; POF1-R: CCCTCAAGGATGTCACTGGC; ACT1\_F: ATGTTTCCAGGTATTGCCG; ACT1\_R: GCCAAAGCGGTGATTTCT; YFR057W-F: ACACCGCCAAGCTTCCAATA; YFR057

W-R: TTGCCACGCAAAGAAAGGAC; ACT1\_F: CATGTTC CCAGGTATTGCCG; ACT1\_R: GCCAAAGCGGTGATTTCT; YFR057W-F: ACACCGCCAAGCTTCCAATA and YFR057W-R: TTGCCACGCAAAGAAAGGAC. Primers were designed to get amplicons of 100–200 bp in size to ensure maximal PCR efficiency, and the accuracy of quantification. PCR amplification was performed in a 10-μL final volume that contained 2.5 μL of the DNA template, 1.5 μL MilliQ water, 0.2 μM of each primer, and 5 μL of LightCycler 480 SYBR Green I Master (Roche). Reactions were performed in 96-well plates in an LightCycler 480 (II) PCR amplification and detection instrument with an initial denaturalization step at 95°C for 5 min, followed by 45 cycles of 95°C for 10 s, either 53 or 54°C for 10 s and 72°C for 4 s. A melting curve analysis was included at the end of each amplification program to confirm the presence of a single PCR product of all the samples with no primer-dimers. The Advanced Relative Quantification program v.1.5.1, implemented in the LightCycler 480, was used to analyze the results, and the efficiency of all the primer pairs was previously determined and included in the analysis. Normalization of the quantification results of genes *ARE1*, *YCL001W-A*, and *POF1* was performed using the levels of genes *ACT1* and *YFR057W* as reference genes.

## Ethanol Tolerance Assays by Drop Test Experiments

Drop test experiments were carried out to assess strains 2-200-2 and 2-200-2-S4 ethanol tolerances. Rectangular GPY plates supplemented with different ethanol percentages (0, 6, 10, 14, 16, and 18%) were prepared. Yeast cells were grown overnight at 28°C in GPY media and diluted to an OD<sub>600</sub> = 0.1 in sterile water. Then, serial dilutions of cells (10<sup>-1</sup> to 10<sup>-3</sup>) were transferred

<sup>1</sup><http://www.yeastgenome.org>





on the plates with replicates and incubated at 28°C for 10 days with the plates wrapped in plastic paraffin film to avoid ethanol evaporation. Each strain was inoculated twice in the same plate but in different positions, and an exact replicate of the plate was done. With this method, four biological replicates of each strain were generated.

## RESULTS

### Phylogenetic Position of the Strains

We used whole genome sequencing of four strains, exhibiting different levels of ethanol tolerance (Arroyo-López et al., 2010), to investigate the relationship between genomic differences and ethanol tolerance. Our assembly and annotation pipeline allowed us to extract about 6,000 coding sequences per genome, of which 2115 concatenated gene sequences in common with other 38 strains (**Supplementary Table S1**), representative of different pure lineages described for *S. cerevisiae* (Liti et al., 2009), were used to reconstruct a multi-locus ML phylogeny (**Figure 1**). When looking at the placement of our strains we observed that Temohaya-MI26 does not appear to cluster within any of the groups we selected, and shows a central position in the tree. This strain may be from a different American population not considered here, like the recently described Ecuadorean population (Peter et al., 2018). The wine strains (T73 and the two *flor* strains) clustered with wine/European strains, but within two sister-clades. More specifically, the *flor* strains GBFlor-C and CECT10094 group with EC1118 and T73 in the other clade which contains wine strains. The position of EC1118 is consistent with previous results (Coi et al., 2017) describing that strain EC1118 clusters with *flor* strains which form a subpopulation among wine/European strains. As EC1118 was closely related to high ethanol-tolerant strains, and showed an intermediate tolerance, we used published genomic data of this strain for further analysis.

### Heterozygosity Levels Differ Between Strains

The frequency of sexual reproduction and outcrossing in *S. cerevisiae* has a high impact on heterozygosity levels, which can indicate differences in life-style between strains. We assessed heterozygosity levels here calculating the number of heterozygous positions in coding regions of the studied strains. High differences are found between strains (**Supplementary Table S2**). Temohaya-MI26 has the lowest heterozygosity with 2433 heterozygous positions in the genome. T73 and GBFlor-C have 4586 and 3094 heterozygous positions, respectively. However, EC1118 and CECT10094 have the highest heterozygosity levels with 12983 and 13789 heterozygous SNPs in the genome, respectively, which represent a mean of two SNPs per gene in their genomes. Interestingly, no relationship is observed between ethanol tolerance and differences in heterozygosity.

In general, heterozygous SNPs were uniformly present along the genome, although several events of loss of heterozygosity (LOH) were observed (**Figure 2** and **Supplementary Figure S1**). These events affected large chromosome portions, mostly including chromosome ends.

To identify possible genes involved in ethanol tolerance, we checked for non-synonymous SNPs fixed only in both highly ethanol-tolerant strains CECT10094 and GBFlor-C. Due to the heterozygosity and the phylogenetic relatedness that these strains have with EC1118, only seven amino-acid changes were fixed and exclusive to both strains (**Table 1**). These are located in proteins encoded by six genes: *CUZ1*, *GCY1*, *RPN7*, *KAR3*, *DPB2*, and *ATG13*. With the exception of *CUZ1* and *GCY1*, these genes were located on the right arm of chromosome XVI, which was affected by a LOH event shared by CECT10094 and GBFlor-C. Interestingly, *CUZ1* and *RPN7* are two genes related with ubiquitin and proteasome pathways, which are important processes in the maintenance of protein homeostasis and the degradation of unfolded proteins. Both processes could be related with the presence of aneuploidies in the studied strains and their ethanol tolerance, as discussed below.

### Highly Ethanol-Tolerant Strains Share Chromosome III Aneuploidy

Most strains of *S. cerevisiae* are diploids, but it has been shown that industrial strains, associated with human-related environments, present different ploidy levels (Gallone et al., 2016; Peter et al., 2018). We assessed our strains' ploidy by flow cytometry, and found that T73, CECT10094, EC1118, and Temohaya-MI26 were diploids. Cytometry average of triplicates compared to a known diploid were respectively:  $2.117 \pm 0.029$ ,  $2.200 \pm 0.030$ ,  $2.196 \pm 0.029$ ,  $2.218 \pm 0.027$  (**Supplementary Table S3**). Contrastingly, GBFlor-C was found to be a triploid strain ( $3.510 \pm 0.055$ ) (**Supplementary Table S3**).

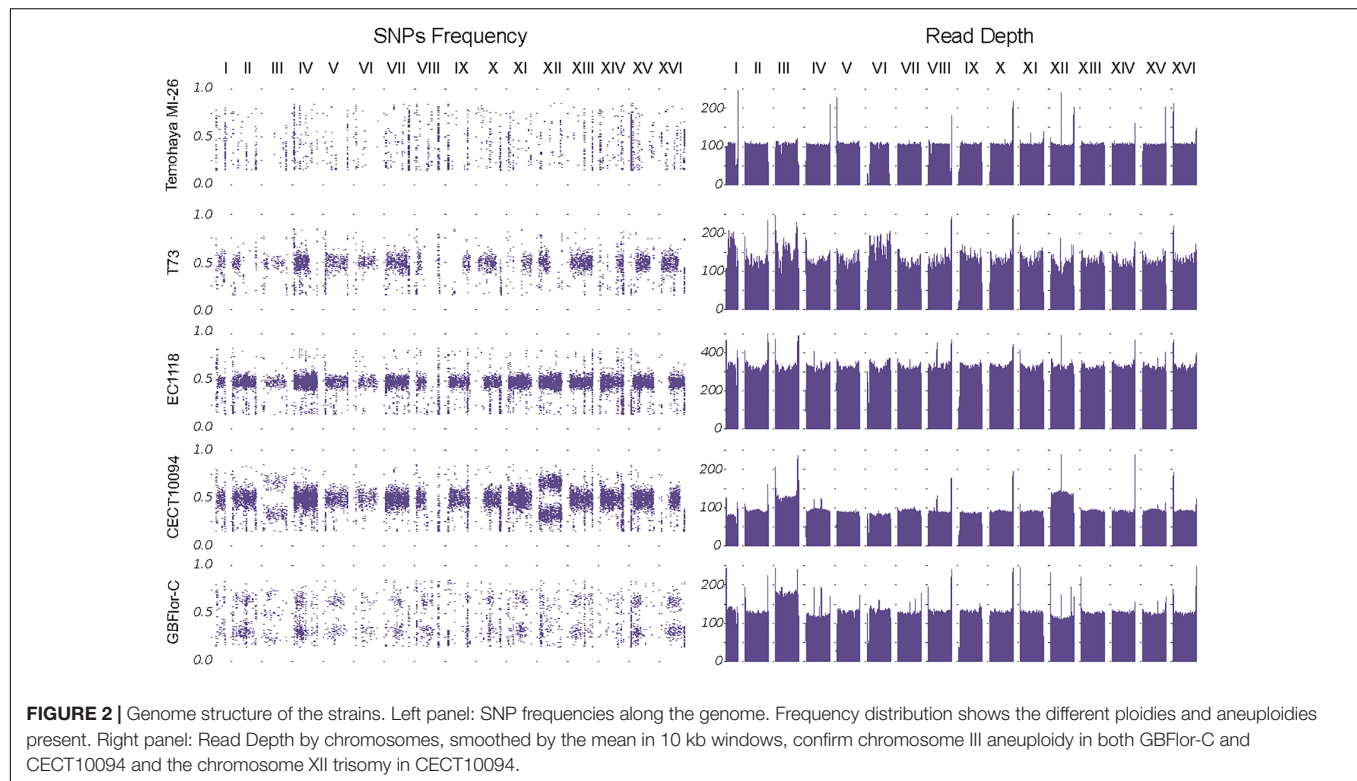
Another method to confirm the ploidy state is to use the heterozygous SNP frequency distribution along the genome (**Figure 2**, left panel). The diploid and heterozygous strains showed a SNP frequency distribution around 0.5, which confirms their diploid state. In the same way, GBFlor-C, which is triploid, showed a typical SNP frequency distribution around 0.33 and 0.66. As Temohaya-MI26 is completely homozygous, it was not possible to confirm its ploidy state with this method.

Fast adaptation to a stressful environment can be driven by large-scale genomic rearrangements. Among these, aneuploidies

**TABLE 1 |** Genes affected by non-synonymous changes exclusive of CECT10094 and GBFlor-C.

Systematic name	Gene name	Description	aa position	Ref aa	Alt aa
YNL155W	<i>CUZ1</i>	CDC48-associated UBL/Zn-finger protein	80	Y	F
YOR120W	<i>GCY1</i>	Galactose-inducible Crystallin-like Yeast prot	86	Q	E
YPR108W	<i>RPN7</i>	Regulatory Particle Non-ATPase	6	E	K
YPR141C	<i>KAR3</i>	KARyogamy	670	K	N
YPR175W	<i>DPB2</i>	DNA Polymerase B subunit 2	565	V	F
			584	E	Q
YPR185W	<i>ATG13</i>	AuTophagy-related	158	T	S





are getting much attention as a potential driver of adaptation of industrial relevance in *S. cerevisiae* (Gorter de Vries et al., 2017). We checked for the presence of aneuploidies in two ways: changes in read-depth between chromosomes and changes in heterozygous SNPs frequency compared to the overall genome frequency distribution (Figure 2). Interestingly, the highest ethanol-tolerant strains, CECT10094 and GBFlor-C were aneuploids. CECT10094 had an extra copy of chromosomes XII and III, and GBFlor-C also showed an extra copy of chromosome III. As chromosome III polysomy was shared between these two strains in different ploidy backgrounds, we further investigated if this could be of importance to explain their higher ethanol tolerance.

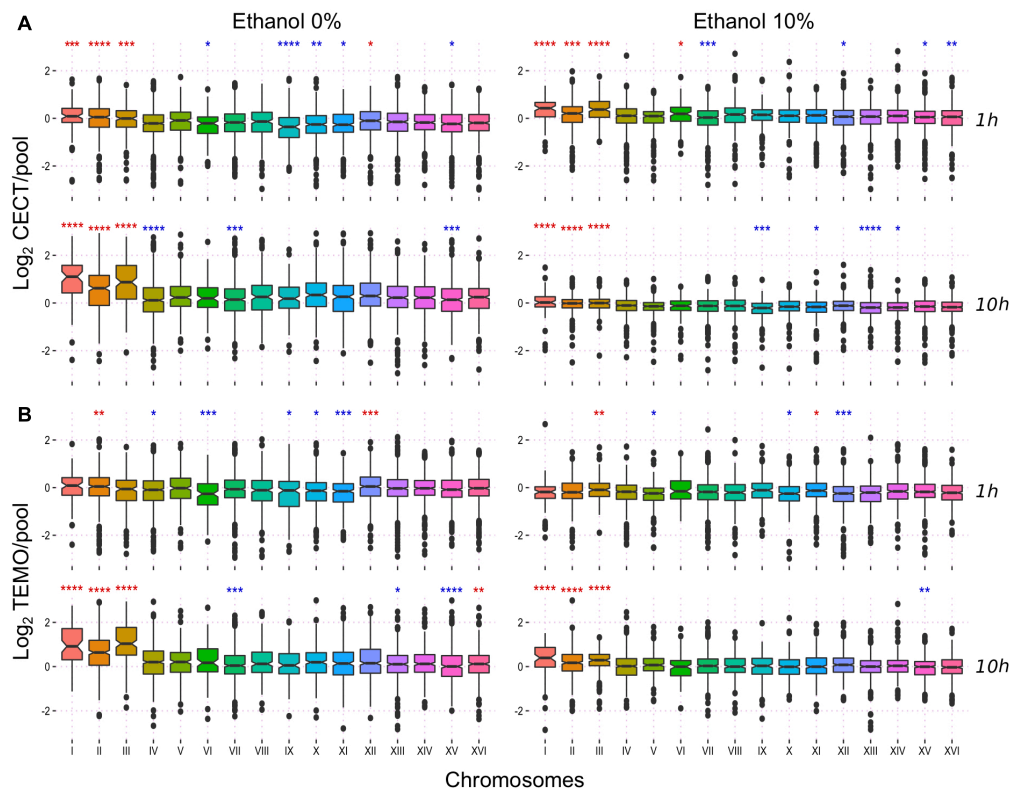
## Chromosome III Expression Increases With Ethanol Stress

A higher number of copies of a chromosome is related with a higher expression of the genes in this chromosome (Torres et al., 2007). We asked if in this case the higher expression of chromosome III could be related with ethanol tolerance. We used transcriptomic data from a previous study (Navarro-Tapia et al., 2016) to shed light on the importance of chromosomes expression on the ethanol tolerance. In brief, the strain with the lowest and highest ethanol tolerance of a set of strains from diverse isolation sources (Temohaya-MI26 and CECT10094, respectively) were selected and their RNA was extracted after 1 or 10 h of growth in two conditions: after a 10% ethanol shock or without stress. The genes were grouped by chromosomes to show the global contribution

of these in the expression profile of the different conditions (Figure 3). Contribution of each chromosome in the complete transcriptome of each strains were different but here we focused on chromosome III due to its aneuploidy in the highly ethanol-tolerant strains. Without ethanol stress, chromosome III showed a significantly higher expression at 1 h of growth compared to other chromosomes in CECT10094 but not in Temohaya-MI26. At 10 h of growth, chromosome III global expression is up-regulated in both strains and in both growth conditions. One hour after ethanol stress, however, chromosome III is significantly up-regulated in both Temohaya-MI26 and CECT10094. In Temohaya-MI26, chromosome III is the most significantly overexpressed chromosome in the genome after a short exposure to ethanol. Thus, the expression pattern observed here could be related with a higher expression in CECT10094 due to the aneuploidy. Furthermore, the change in the expression contribution of chromosome III in Temohaya-MI26 after ethanol shock is consistent with the presence of several genes in the chromosome contributing to the ethanol stress response, even in the low ethanol-tolerant strain.

## Chromosome III Aneuploidies Affect Growth on Ethanol in Different Backgrounds

Several studies showed that aneuploidies could be of importance in certain conditions (Gorter de Vries et al., 2017). In particular, chromosome III copy number variation was related with higher heat tolerance (Yona et al., 2012) and was duplicated in ethanol adaptation experiments (Voordeckers et al., 2015). In a recent



**FIGURE 3 |** Transcriptional response to ethanol at a chromosomal level. **(A)** expression fold change of CECT10094 compared to the pool at 1 and 10 h growth with 0 and 10% ethanol stress. **(B)** expression fold change for Temohaya-MI26. Statistics Wilcoxon test among all groups: ns:  $p > 0.05$ , \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ , \*\*\*\* $p < 0.0001$ . Significance symbols are colored in red if the chromosome is upregulated and in blue if it is downregulated.

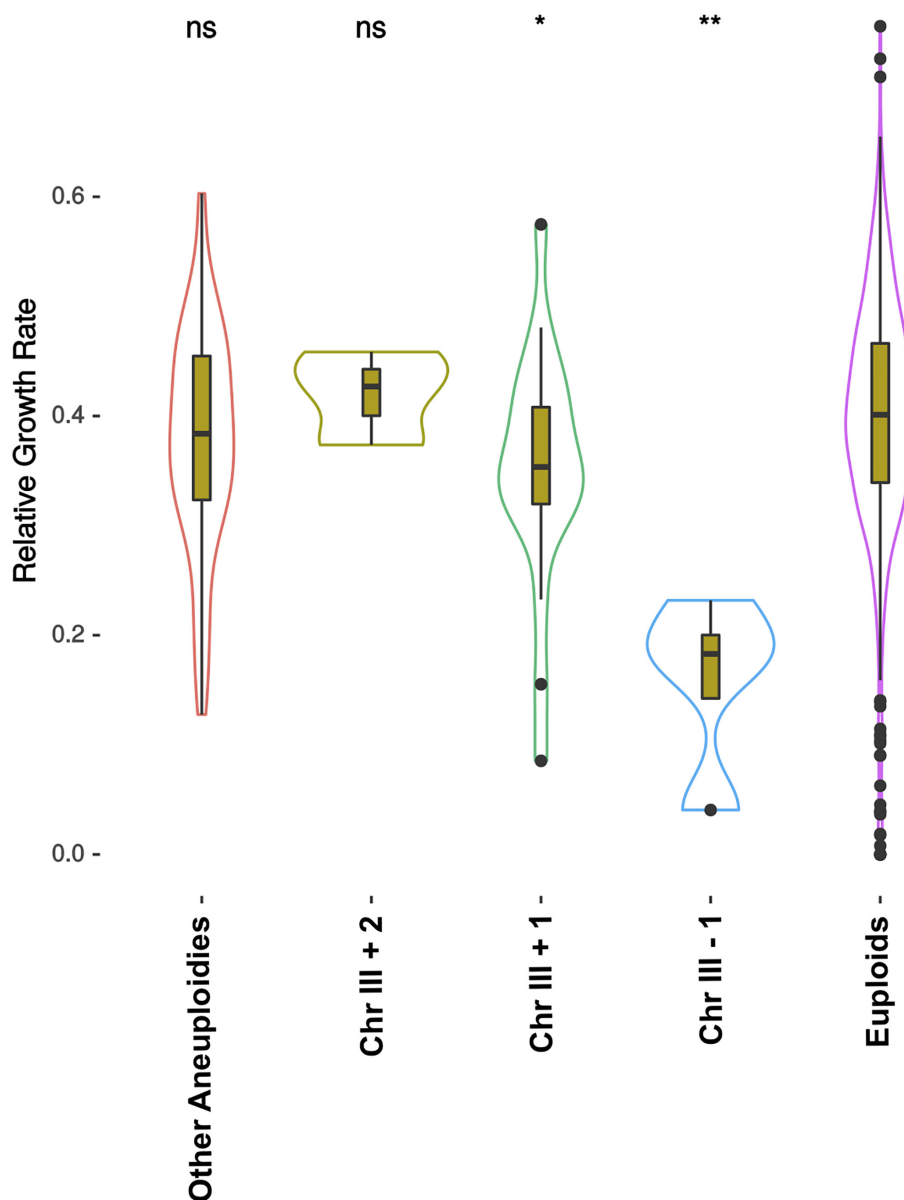
study (Peter et al., 2018), more than 1000 *S. cerevisiae* strains were sequenced and phenotyped in several conditions. Here we used the phenotype on 15% ethanol stress and the ploidy and aneuploidy information, and grouped the strains to search for differences in ethanol stress tolerance in a wider genetic background set (Figure 4 and Supplementary Figure S2). Most of the ploidies and chromosome copy numbers did not show many differences compared to diploid strains exhibiting a perfect euploidy. As groups are of different sizes and many factors are involved in ethanol tolerance, differences are hard to see. However, in diploids the number of chromosome III copies showed a specific trend (Figure 4). Strains lacking one of the chromosome copy were significantly worse than diploids growing on 15% ethanol respect to control condition. As the number of copies of chromosome III increases, higher is the relative growth rate exhibited. Strains with an extra copy were better than the euploids with two copies, and these better than monosomic strains, with one single chromosome III.

## Removing the Extra Copy of Chromosome III Strongly Affects Ethanol Tolerance

To confirm that the aneuploidy on chromosome III directly influenced the ethanol tolerance of the strains, we removed the

extra copy from the genome (see section Materials and Methods), returning strains to the euploid state. Unfortunately, we could not obtain any modified strain of CECT10094 and GBFlor-C with the experimental approach used. We therefore used a laboratory evolved strain (2-200-2) obtained by Voordeckers et al. (2015). These authors evolved six prototrophic, isogenic *S. cerevisiae* strains of different ploidy (1n, 2n, and 4n), all them generated from the haploid S288C-derivative FY5 strains, in chemostats with increasing ethanol concentrations (up to 12%) during 200 generations. The haploid and tetraploid lines showed rapid convergence toward a diploid state, and at the end of the experiment, all the evolved clones were highly tolerant to ethanol, and most of them shared the acquisition of an extra copy of chromosome III, although other specific aneuploidies were also present in some clones. The evolved clone 2-200-2 is a haploid-derived diploid with a chromosome III trisomy. As this evolved lab clone shared this genomic feature with our *flor* strains, it was used in an experiment to test if its ethanol tolerance was reverted after the removal of its extra chromosome III copy, by using the integration of a selection/counter-selection marker.

We first determined the copy number of chromosome III in the strain 2-200-2 and the modified strain 2-200-2-S4 by qPCR to confirm the removal of the extra copy. We used primers for three genes spread along the chromosome and compared chromosome



**FIGURE 4 |** Relative growth rate of 1011 strains on 15% ethanol. Relative growth rate is the ratio between growth of the strain on YPD at 30°C and growth on 15% ethanol. Strains are grouped by their ploidy and presence of different aneuploidies. In this figure only diploids are presented, all other ploidy levels are shown in **Supplementary Figure 2**. Chr III +1, Chr III +2, and Chr III -1 tagged contain one extra copy, two extra copies and one less copy of chromosome III, respectively, and can exhibit other aneuploidies. “Other Aneuploidies” group are strains that have different kind of aneuploidies but not on chromosome III, and “Euploids” have no aneuploidy. Significance is Wilcoxon paired with diploid euploid strains as reference (ns:  $p > 0.05$ , \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ , \*\*\*\* $p < 0.0001$ ). Original data were obtained from Peter et al. (2018).

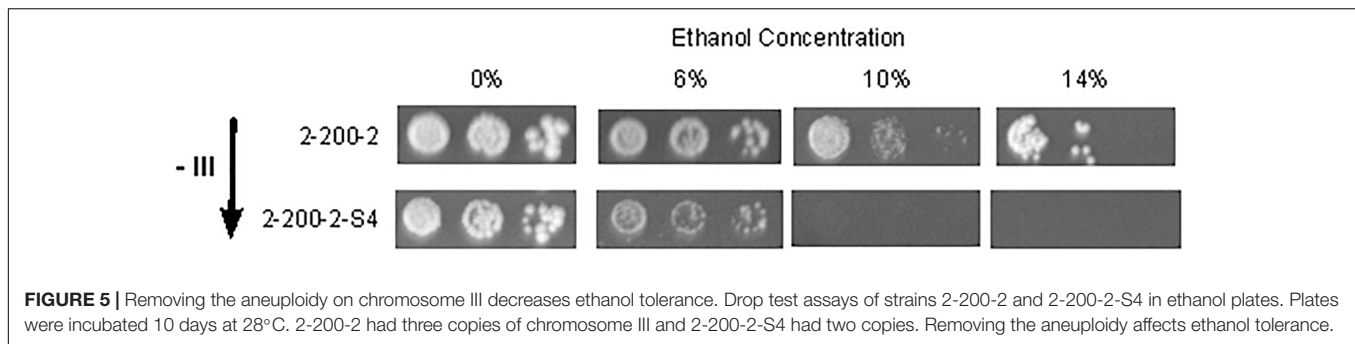
dosage using as reference genes *ACT1* and YFR057W from chromosome VI. The results showed that 2-200-2-S4 had lost an extra copy of chromosome III for each one of the tested genes resulting in a gene copy number close to 2 ( $1.99 \pm 0.31$  for *ARE1*,  $2.14 \pm 0.28$  for *YCL001W-A* and  $2.06 \pm 0.20$  for *POF1*).

After removing chromosome III extra copy, we tested growth of strains 2-200-2 and 2-200-2-S4 on GPY with different ethanol concentrations (**Figure 5**). The strain 2-200-2, which has three copies of chromosome III, was able to grow even on 14%

ethanol concentration. 2-200-2-S4, which had the extra copy of the chromosome III removed, was not able to grow on 10%-ethanol medium.

## DISCUSSION

Ethanol is one of the major stresses suffered by yeasts in industrial environments. Among the different species of its



genus, *S. cerevisiae* is the most tolerant to ethanol (Arroyo-López et al., 2010). Even if this characteristic is widely studied due to its importance in biotechnology and industry, it is still unknown what are the key factors that drive adaptation to high ethanol concentrations (Snoek et al., 2016). Here, we sequenced the genome of *S. cerevisiae* strains especially selected for their differential ethanol tolerance. A previous work showed that Temohaya-MI26 was low ethanol-tolerant, T73 and EC1118 had an intermediate tolerance and GBFlor-C and CECT10094 were high ethanol-tolerant (Arroyo-López et al., 2010).

The phylogenetic analysis performed showed that the wine strains could be divided in two subclades. The first one grouped typical wine strains and contained the T73 strain, and the second grouped *flor* strains (GBFlor-C and CECT10094) with EC1118. These results confirm that *flor* strains form a different subpopulation among wine strains, as previously described (Legras et al., 2014, 2016; Coi et al., 2017; Eldarov et al., 2018). Temohaya-MI26, in contrast, was not included in any of the groups considered. The ethanol tolerance was higher in *flor* strains but not in EC1118 which is in the same group and closely related to CECT10094 and GBFlor-C. This points that this phenotype is variable even within the same population.

Until recently, most of the sequenced *S. cerevisiae* strains were homoploid spore derivatives to improve assembly and analysis. These methods nevertheless shadow interesting parts of genome structure. Heterozygosity levels were related to differences in strains lifestyle (Magwene et al., 2011). In industrial environments, this species reproduces asexually and has higher heterozygosity levels than natural strains (Gallone et al., 2016; Peter et al., 2018). The strains studied also showed similar trend. Temohaya-MI26, which is not related to industrial strains, showed a low heterozygosity. This may mainly be due to the use of haploselfing in its environment. In contrast, wine related strains showed higher heterozygosity, with events of LOH, and were in the range of levels previously described for wine strains (Gallone et al., 2016; Peter et al., 2018).

We found that the highly ethanol-tolerant strains shared an aneuploidy on chromosome III in different ploidy backgrounds. A high fidelity of genome replication and segregation is vital for the survival of any organism as well as for the production of future generations. Errors in these steps during meiosis, and also during mitosis in unicellular organisms, can lead to a change in ploidy or chromosome numbers.

In fact, it has been suggested that the ethanol itself could induce chromosome malsegregation (Crebelli et al., 1989). These severe genome changes can be detrimental, causing a decrease in the fitness of the organism. However, during specific circumstances, such as periods of stress, in which gene dose increase can be beneficial, polyploidy or aneuploidy can provide a higher fitness (Todd et al., 2017). Aneuploidy is gaining attention for its relevance in industrial *S. cerevisiae* strains (Gorter de Vries et al., 2017) and for its possible implications in driving adaptation in general (Chen et al., 2012; Bennett et al., 2014). Consequences of aneuploidy are usually detrimental for strain growth (Mangado et al., 2018). However, it was described that specific chromosome copy-number variations could improve resistance to specific stresses. This way, chromosome III aneuploidy was related with improvement of heat tolerance (Yona et al., 2012). Other authors also found that chromosome III aneuploidies were generated as a response to ethanol stress (Gorter de Vries et al., 2017). Moreover, artificial segmental aneuploidies of chromosome III increased ethanol tolerance (Natesuntorn et al., 2015). Evolution on mild ethanol stress showed that different aneuploidies appeared, including chromosome III copy number increases (Adamczyk et al., 2016). Finally, a long-term evolution study showed that chromosome III aneuploidy was a common event (Voordeckers et al., 2015). Here, we found that relationship in non-laboratory strains, which may indicate that it is a long-term adaptation and its fixation seems to be important for ethanol tolerance. We also showed that the number of copies of the chromosome plays a role in this phenotype in different backgrounds (Figure 4), and that it is adaptive and affects directly to ethanol tolerance.

We dissected the expression profile by chromosomes of the high and low ethanol-tolerant strains. We found that the low ethanol-tolerant strains up-regulated chromosome III expression after ethanol stress and that the high ethanol-tolerant had its expression increased even in the absence of ethanol. Therefore, aneuploidy can be a way to change dosage of important genes present in chromosome III (Yona et al., 2012). This is consistent with our results, but further investigation is needed to find which genes in this chromosome could be involved in this process. Nevertheless, genes present in aneuploid chromosomes can change expression of other genes in other chromosomes, causing broad expression changes (Selmecki et al., 2008).



Aneuploidy itself affects the cell in different ways. Additional copies of a chromosome increases proteotoxic stress, which affect the protein folding processes in the cell (Torres et al., 2007). Interestingly, yeast were found to induce unfolded protein response under ethanol stress (Navarro-Tapia et al., 2016, 2017). Two out of seven nonsynonymous changes found in the high ethanol-tolerant strains affected genes related to protein homeostasis. We open here the possibility that ethanol and aneuploidy tolerance could involve similar processes, which may involve fixing variants affecting these processes and therefore aneuploidy itself could play a role on improving ethanol tolerance. As chromosome III is one of the smallest chromosome in *S. cerevisiae* genome, we cannot discard that the aneuploidy tolerance induced could be the cause of the observed phenotype.

## CONCLUSION

In conclusion, in this work we showed that ethanol tolerance was related to an aneuploidy on chromosome III in wine *S. cerevisiae*. Further work will be needed to elucidate the actual mechanism by which this phenomenon happens, but we confirmed that this is an adaptive trait that seems to be a widespread trend.

## DATA AVAILABILITY

All the genomic data in this article are available on NCBI under BioProject PRJNA493718.

## AUTHOR CONTRIBUTIONS

EB and CT conceived and designed the study. MM and LM performed all the genome sequence and phylogenetic analyses under CT and EB supervision. AA and RP-T did the chromosome III removal. ML-P carried out the ethanol tolerance assays by drop tests. MM, RP-T, and CT wrote the first versions of the article, and EB the final version.

## REFERENCES

- Adamczyk, J., Deregoska, A., Skoneczny, M., Skoneczna, A., Kwiatkowska, A., Potocki, L., et al. (2016). Adaptive response to chronic mild ethanol stress involves ROS, sirtuins and changes in chromosome dosage in wine yeasts. *Oncotarget* 7, 29958–29976. doi: 10.18632/oncotarget.8673
- Ansanay-Galeote, V., Blondin, B., Dequin, S., and Sablayrolles, J. M. (2001). Stress effect of ethanol on fermentation kinetics by stationary-phase cells of *Saccharomyces cerevisiae*. *Biotechnol. Lett.* 23, 677–681. doi: 10.1023/A:1010396232420
- Arroyo-López, F. N., Salvadó, Z., Tronchoni, J., Guillamón, J. M., Barrio, E., and Querol, A. (2010). Susceptibility and resistance to ethanol in *Saccharomyces* strains isolated from wild and fermentative environments. *Yeast* 27:1015. doi: 10.1002/yea.1809
- Auesukaree, C. (2017). Molecular mechanisms of the yeast adaptive response and tolerance to stresses encountered during ethanol fermentation. *J. Biosci. Bioeng.* 124, 133–142. doi: 10.1016/j.jbiosc.2017.03.009
- Barrick, J. E., Colburn, G., Deatherage, D. E., Traverse, C. C., Strand, M. D., Borges, J. J., et al. (2014). Identifying structural variation

## FUNDING

This work was funded by grant AGL2015-67504-C3-R from the Spanish Government and European Union ERDF-FEDER to EB. MM was supported by a Ph.D. student contract ACIF/2015/194 from the Regional Government of Valencia. ML-P acknowledges a Ph.D. student FPU contract FPU15/01775 from the Spanish Government. LM and RP-T are supported by the aforementioned grant AGL2015-67504-C3-R. CT acknowledges a “Juan de la Cierva” postdoctoral contract JCI-2012-14056 from the Spanish Government.

## ACKNOWLEDGMENTS

We are very grateful to Kevin Verstrepren and to Cristian Varela for providing strain 2-200-2, and plasmid pCORE5, respectively.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2019.00082/full#supplementary-material>

**FIGURE S1** | CIRCOS plot of heterozygous and homozygous SNPs. The wine and *flor* strains reads were mapped on T73 assembly for clarity purposes as described in Materials and Methods. Only SNP's on coding sequences are represented. In blue are represented heterozygous SNP's and in orange homozygous SNPs. Clear regions of LOH are observed on different chromosome regions.

**FIGURE S2** | Relative growth rate of 1011 strains on 15% ethanol. Relative growth rate on ethanol for all ploidies and aneuploidies (see **Figure 4**).

**TABLE S1** | Source of the *S. cerevisiae* genome sequences used in this study.

**TABLE S2** | Heterozygosity levels for each strain. The absolute number of heterozygotic and homozygotic SNPs in coding sequences are shown. The ratio is estimated as the number of heterozygotic positions divided by the total number of SNPs.

**TABLE S3** | Flow cytometry ploidy estimates.

- in haploid microbial genomes from short-read resequencing data using breseq. *BMC Genomics* 15:1039. doi: 10.1186/1471-2164-15-1039
- Bennett, R. J., Forche, A., and Berman, J. (2014). Rapid mechanisms for generating genome diversity: whole ploidy shifts, aneuploidy, and loss of heterozygosity. *Cold Spring Harb. Perspect. Med.* 4:a019604. doi: 10.1101/cshperspect.a019604
- Boetzer, M., Henkel, C. V., Jansen, H. J., Butler, D., and Pirovano, W. (2011). Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* 27, 578–579. doi: 10.1093/bioinformatics/btq683
- Boetzer, M., and Pirovano, W. (2012). Toward almost closed genomes with GapFiller. *Genome Biol.* 13:R56. doi: 10.1186/gb-2012-13-6-r56
- Brauner, A., Fridman, O., Gefen, O., and Balaban, N. Q. (2016). Distinguishing between resistance, tolerance and persistence to antibiotic treatment. *Nat. Rev. Microbiol.* 14:320. doi: 10.1038/nrmicro.2016.34
- Chen, G., Rubinstein, B., and Li, R. (2012). Whole chromosome aneuploidy: big mutations drive adaptation by phenotypic leap. *Bioessays* 34, 893–900. doi: 10.1002/bies.201200069

- Coi, A. L., Bigey, F., Mallet, S., Marsit, S., Zara, G., Gladieux, P., et al. (2017). Genomic signatures of adaptation to wine biological ageing conditions in biofilm-forming flor yeasts. *Mol. Ecol.* 26, 2150–2166. doi: 10.1111/mec.14053
- Crebelli, R., Conti, G., Conti, L., and Carere, A. (1989). A comparative study on ethanol and acetaldehyde as inducers of chromosome malsegregation in *Aspergillus nidulans*. *Mutat. Res.* 215, 187–195. doi: 10.1016/0027-5107(89)90183-90188
- Darling, A. E., Mau, B., and Perna, N. T. (2010). Progressive mauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS One* 5:e11147. doi: 10.1371/journal.pone.0011147
- Dayarian, A., Michael, T. P., and Sengupta, A. M. (2010). SOPRA: scaffolding algorithm for paired reads via statistical optimization. *BMC Bioinformatics* 11:345. doi: 10.1186/1471-2105-11-345
- Duan, S.-F., Han, P.-J., Wang, Q.-M., Liu, W.-Q., Shi, J.-Y., Li, K., et al. (2018). The origin and adaptive evolution of domesticated populations of yeast from Far East Asia. *Nat. Commun.* 9:2690. doi: 10.1038/s41467-018-05106-7
- Eldarov, M. A., Beletsky, A. V., Tanashchuk, T. N., Kishkovskaya, S. A., Ravin, N. V., and Mardanov, A. V. (2018). Whole-genome analysis of three yeast strains used for production of sherry-like wines revealed genetic traits specific to flor yeasts. *Front. Microbiol.* 9:965. doi: 10.3389/fmicb.2018.00965
- Esteve-Zarzoso, B., Peris-Toran, M. J., García-Máiquez, E., Uruburu, F., and Querol, A. (2001). Yeast population dynamics during the fermentation and biological aging of sherry wines. *Appl. Environ. Microbiol.* 67, 2056–2061. doi: 10.1128/AEM.67.5.2056-2061.2001
- Gallone, B., Steensels, J., Prahl, T., Soriaga, L., Sael, V., Herrera-Malaver, B., et al. (2016). Domestication and divergence of *Saccharomyces cerevisiae* beer yeasts. *Cell* 166, 1397.e16–1410.e16. doi: 10.1016/j.cell.2016.08.020
- Giudici, P., Solieri, L., Pulvirenti, A. M., and Cassanelli, S. (2005). Strategies and perspectives for genetic improvement of wine yeasts. *Appl. Microbiol. Biotechnol.* 66, 622–628. doi: 10.1007/s00253-004-1784-2
- Gorter de Vries, A. R., Pronk, J. T., and Daran, J.-M. G. (2017). Industrial relevance of chromosomal copy number variation in *Saccharomyces* yeasts. *Appl. Environ. Microbiol.* 83:e3206-16. doi: 10.1128/AEM.03206-16
- Haase, S. B., and Reed, S. I. (2002). Improved flow cytometric analysis of the budding yeast cell cycle. *Cell Cycle* 1, 132–136. doi: 10.4161/cc.1.2.114
- Hagman, A., and Piškur, J. (2015). A study on the fundamental mechanism and the evolutionary driving forces behind aerobic fermentation in yeast. *PLoS One* 10:e0116942. doi: 10.1371/journal.pone.0116942
- Hung, J.-H., and Weng, Z. (2017). Mapping short sequence reads to a reference genome. *Cold Spring Harb. Protoc.* 2017, 158–164. doi: 10.1101/pdb.prot093161
- Joshi, N. A., and Fass, J. N. (2011). *Sickle: A Sliding-Window, Adaptive, Quality-Based Trimming Tool for FastQ Files. Version 1.33*. Available at: <https://github.com/najoshi/sickle>.
- Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. doi: 10.1093/molbev/mst010
- Kurtz, S., Phillippy, A., Delcher, A. L., Smoot, M., Shumway, M., Antonescu, C., et al. (2004). Versatile and open software for comparing large genomes. *Genome Biol.* 5:R12. doi: 10.1186/gb-2004-5-2-r12
- Kutyna, D., Cordente, A. G., and Varela, C. (2014). “Genetic engineering of industrial *Saccharomyces cerevisiae* strains using a selection/counter-selection approach,” in *Yeast Metabolic Engineering: Methods and Protocols*, ed. V. Mapelli (New York, NY: Humana Press), 157–168.
- Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923
- Legras, J.-L., Erny, C., and Charpentier, C. (2014). Population structure and comparative genome hybridization of European flor yeast reveal a unique group of *Saccharomyces cerevisiae* strains with few gene duplications in their genome. *PLoS One* 9:e108089. doi: 10.1371/journal.pone.0108089
- Legras, J.-L., Galeote, V., Bigey, F., Camarasa, C., Marsit, S., Nidelet, T., et al. (2018). Adaptation of *S. cerevisiae* to fermented food environments reveals remarkable genome plasticity and the footprints of domestication. *Mol. Biol. Evol.* 35, 1712–1727. doi: 10.1093/molbev/msy066
- Legras, J.-L., Moreno-García, J., Zara, S., Zara, G., García-Martínez, T., Mauricio, J. C., et al. (2016). Flor yeast: new perspectives beyond wine aging. *Front. Microbiol.* 7:503. doi: 10.3389/fmicb.2016.00503
- Liti, G., Carter, D. M., Moses, A. M., Warringer, J., Parts, L., James, S. A., et al. (2009). Population genomics of domestic and wild yeasts. *Nature* 458, 337–341. doi: 10.1038/nature07743
- Magwene, P. M., Kayıkçı, Ö., Granek, J. A., Reininga, J. M., Scholl, Z., and Murray, D. (2011). Outcrossing, mitotic recombination, and life-history trade-offs shape genome evolution in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. U.S.A.* 108, 1987–1992. doi: 10.1073/pnas.1012544108
- Mangado, A., Morales, P., Gonzalez, R., and Tronchoni, J. (2018). Evolution of a yeast with industrial background under winemaking conditions leads to diploidization and chromosomal copy number variation. *Front. Microbiol.* 9:1816. doi: 10.3389/fmicb.2018.01816
- Natesuntorn, W., Iwami, K., Matsubara, Y., Sasano, Y., Sugiyama, M., Kaneko, Y., et al. (2015). Genome-wide construction of a series of designed segmental aneuploids in *Saccharomyces cerevisiae*. *Sci. Rep.* 5:12510. doi: 10.1038/srep12510
- Navarro-Tapia, E., Nana, R. K., Querol, A., and Pérez-Torrado, R. (2016). Ethanol cellular defense induce unfolded protein response in yeast. *Front. Microbiol.* 7:189. doi: 10.3389/fmicb.2016.00189
- Navarro-Tapia, E., Pérez-Torrado, R., and Querol, A. (2017). Ethanol effects involve non-canonical unfolded protein response activation in yeast cells. *Front. Microbiol.* 8:383. doi: 10.3389/fmicb.2017.00383
- Otto, T. D., Dillon, G. P., Degraeve, W. S., and Berriman, M. (2011). RATT: rapid annotation transfer tool. *Nucleic Acids Res.* 39:e57. doi: 10.1093/nar/gkq1268
- Peter, J., De Chiara, M., Friedrich, A., Yue, J.-X., Pflieger, D., Bergström, A., et al. (2018). Genome evolution across 1,011 *Saccharomyces cerevisiae* isolates. *Nature* 556, 339–344. doi: 10.1038/s41586-018-0030-35
- Piškur, J., Rozpedowska, E., Polakova, S., Merico, A., and Compagno, C. (2006). How did *Saccharomyces* evolve to become a good brewer? *Trends Genet.* 22, 183–186.
- Querol, A., Fernández-Espinar, M. T., del Olmo, M., and Barrio, E. (2003). Adaptive evolution of wine yeast. *Int. J. Food Microbiol.* 86, 3–10. doi: 10.1016/S0168-1605(03)00244-7
- Querol, A., Huerta, T., Barrio, E., and Ramón, D. (1992). Dry yeast-strain for use in fermentation of Alicante wines - selection and DNA patterns. *J. Food Sci.* 57, 183–185. doi: 10.1111/j.1365-2621.1992.tb05451.x
- Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M.-A., et al. (2000). Artemis: sequence visualization and annotation. *Bioinformatics* 16, 944–945. doi: 10.1093/bioinformatics/16.10.944
- Selmecki, A., Gerami-Nejad, M., Paulson, C., Forche, A., and Berman, J. (2008). An isochromosome confers drug resistance in vivo by amplification of two genes, *ERG11* and *TAC1*. *Mol. Microbiol.* 68, 624–641. doi: 10.1111/j.1365-2958.2008.06176.x
- Snoek, T., Verstrepen, K. J., and Voordeckers, K. (2016). How do yeast cells become tolerant to high ethanol concentrations? *Curr. Genet.* 62, 475–480. doi: 10.1007/s00294-015-0561-563
- Stamatakis, A. (2014). RAXML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313. doi: 10.1093/bioinformatics/btu033
- Stanke, M., and Morgenstern, B. (2005). AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Res.* 33, W465–W467. doi: 10.1093/nar/gki458
- Thomson, J. M., Gaucher, E. A., Burgan, M. F., De Kee, D. W., Li, T., Aris, J. P., et al. (2005). Resurrecting ancestral alcohol dehydrogenases from yeast. *Nat. Genet.* 37, 630–635. doi: 10.1038/ng1553
- Todd, R. T., Forche, A., and Selmecki, A. (2017). Ploidy variation in Fungi: polyploidy, aneuploidy, and genome evolution. *Microbiol. Spect.* 5:FUNK-0051-2016. doi: 10.1128/microbiolspec.FUNK-0051-2016
- Torres, E. M., Sokolsky, T., Tucker, C. M., Chan, L. Y., Boselli, M., Dunham, M. J., et al. (2007). Effects of aneuploidy on cellular physiology and cell division in haploid yeast. *Science* 317, 916–924. doi: 10.1126/science.1142210
- Voordeckers, K., Kominek, J., Das, A., Espinosa-Cantú, A., De Maeyer, D., Arslan, A., et al. (2015). Adaptation to high ethanol reveals complex evolutionary pathways. *PLoS Genet.* 11:e1005635. doi: 10.1371/journal.pgen.1005635

- Yona, A. H., Manor, Y. S., Herbst, R. H., Romano, G. H., Mitchell, A., Kupiec, M., et al. (2012). Chromosomal duplication is a transient evolutionary solution to stress. *Proc. Natl. Acad. Sci. U.S.A.* 109, 21010–21015. doi: 10.1073/pnas.1211150109
- Yu, G., Smith, D. K., Zhu, H., Guan, Y., and Lam, T. T.-Y. (2017). ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol. Evol.* 8, 28–36. doi: 10.1111/2041-210X.12628
- Zerbino, D. R., and Birney, E. (2008). Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.* 18, 821–829. doi: 10.1101/gr.074492.107

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Morard, Macías, Adam, Lairón-Peris, Pérez-Torrado, Toft and Barrio. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Characterization of a New *Saccharomyces cerevisiae* Isolated From Hibiscus Flower and Its Mutant With L-Leucine Accumulation for Awamori Brewing

## OPEN ACCESS

### Edited by:

Isabel Sá-Correia,  
University of Lisbon, Portugal

### Reviewed by:

Jean-luc Legras,  
Institut National de la Recherche  
Agronomique Centre Montpellier,  
France  
Jose Sampaio,  
Faculdade de Ciências e Tecnologia  
da Universidade Nova de Lisboa,  
Portugal

### \*Correspondence:

Masatoshi Tsukahara  
tsuka@biojet.jp  
Hiroshi Takagi  
hiro@bs.naist.jp

† These authors have contributed  
equally to this work

### Specialty section:

This article was submitted to  
Evolutionary and Genomic  
Microbiology,  
a section of the journal  
Frontiers in Genetics

**Received:** 10 December 2018

**Accepted:** 06 May 2019

**Published:** 28 May 2019

### Citation:

Abe T, Toyokawa Y, Sugimoto Y,  
Azuma H, Tsukahara K, Nasuno R,  
Watanabe D, Tsukahara M and  
Takagi H (2019) Characterization of a  
New *Saccharomyces cerevisiae*  
Isolated From Hibiscus Flower and Its  
Mutant With L-Leucine Accumulation  
for Awamori Brewing.  
Front. Genet. 10:490.  
doi: 10.3389/fgene.2019.00490

Takayuki Abe<sup>1†</sup>, Yoichi Toyokawa<sup>2†</sup>, Yukiko Sugimoto<sup>2</sup>, Haruna Azuma<sup>1</sup>,  
Keiko Tsukahara<sup>1</sup>, Ryo Nasuno<sup>2</sup>, Daisuke Watanabe<sup>2</sup>, Masatoshi Tsukahara<sup>1\*</sup> and  
Hiroshi Takagi<sup>2\*</sup>

<sup>1</sup> BioJet Co., Ltd., Okinawa, Japan, <sup>2</sup> Division of Biological Science, Graduate School of Science and Technology, Nara  
Institute of Science and Technology, Ikoma, Japan

Since flavors of alcoholic beverages produced in fermentation process are affected mainly by yeast metabolism, the isolation and breeding of yeasts have contributed to the alcoholic beverage industry. To produce awamori, a traditional spirit (distilled alcoholic beverage) with unique flavors made from steamed rice in Okinawa, Japan, it is necessary to optimize yeast strains for a diversity of tastes and flavors with established qualities. Two categories of flavors are characteristic of awamori; initial scented fruity flavors and sweet flavors that arise with aging. Here we isolated a novel strain of *Saccharomyces cerevisiae* from hibiscus flowers in Okinawa, HC02-5-2, that produces high levels of alcohol. The whole-genome information revealed that strain HC02-5-2 is contiguous to wine yeast strains in a phylogenetic tree. This strain also exhibited a high productivity of 4-vinyl guaiacol (4-VG), which is a precursor of vanillin known as a key flavor of aged awamori. Although conventional awamori yeast strain 101-18, which possesses the *FDC1* pseudogene does not produce 4-VG, strain HC02-5-2, which has the intact *PAD1* and *FDC1* genes, has an advantage for use in a novel kind of awamori. To increase the contents of initial scented fruity flavors, such as isoamyl alcohol and isoamyl acetate, we attempted to breed strain HC02-5-2 targeting the L-leucine synthetic pathway by conventional mutagenesis. In mutant strain T25 with L-leucine accumulation, we found a hetero allelic mutation in the *LEU4* gene encoding the Gly516Ser variant  $\alpha$ -isopropylmalate synthase (IPMS). IPMS activity of the Gly516Ser variant was less sensitive to feedback inhibition by L-leucine, leading to intracellular L-leucine accumulation. In a laboratory-scale test, awamori brewed with strain T25 showed higher concentrations of isoamyl alcohol and isoamyl acetate than that brewed with strain HC02-5-2. Such a combinatorial approach to yeast isolation, with whole-genome analysis and metabolism-focused breeding, has the potentials to vary the quality of alcoholic beverages.

**Keywords:** comparative genomics, hibiscus yeast, *Saccharomyces cerevisiae*, breeding, awamori, liquor flavors



## INTRODUCTION

Awamori, a distilled alcoholic beverage made from steamed rice, is brewed primarily in Okinawa, Japan. This traditional spirit, with more than 600 years of history, is representative of Okinawan culture and industry. Two microorganism species, the fungus *Aspergillus luchuensis* and the yeast *Saccharomyces cerevisiae* play essential roles in awamori brewing. *A. luchuensis*, as Kuro-koji mold, saccharifies steamed rice, and subsequently *S. cerevisiae* produces ethanol by multiple parallel fermentations. After 2 weeks of fermentation, the fermented mash (*moromi*), containing ethanol at a final concentration of approximately 20% (v/v), is obtained and applied to the distillation process to develop a rich and strong flavor. Awamori is composed of various aromatic compounds, which are distinguishable from other types of Japanese spirits (Nishiya, 1980). Two categories of flavors have been well studied in awamori: the initial volatile compounds commonly known as the “top notes” of flavors and the flavors that develop during the aging process (Koseki et al., 1996; Taira et al., 2012). These fruity and sweet flavors are favorable and demanded by consumers. For the awamori industry, it is important to enhance these two categories of flavors.

The initial volatile compounds, including fruity aroma esters and alcohols are recognized as important flavors in awamori (Taira et al., 2012). In particular, the biosynthetic pathway of isoamyl alcohol and its acetate, isoamyl acetate, is elucidated as follows. These metabolites are synthesized from  $\alpha$ -ketoisocaproate (KIC), a precursor of L-leucine, in *S. cerevisiae*. Isoamyl alcohol is synthesized from KIC catalyzed by  $\alpha$ -keto acid decarboxylase and alcohol dehydrogenase, and is finally converted into isoamyl acetate by alcohol acetyltransferase. In the L-leucine synthetic pathway,  $\alpha$ -isopropylmalate synthase (IPMS; the *LEU4* gene product) (EC 4.1.3.12) is the key enzyme, because its activity is regulated by feedback inhibition by L-leucine (Baichwal et al., 1983). Previously, we isolated a mutant of the diploid awamori yeast strain 101-18, which accumulated a higher amount of L-leucine selected from the L-leucine analog 5,5,5-trifluoro-DL-leucine (TFL)-resistant colonies. The mutant strain 101-T55 has mutations in the *LEU4* gene, which confers overproduction of isoamyl alcohol and isoamyl acetate (Takagi et al., 2015). Such yeast breeding technology targeted to specific flavors is applicable to the improvement of the quality of awamori.

Vanillin, which is regarded as a representative flavor of aged awamori, gives a sweet vanilla scent (Koseki et al., 1996). Vanillin is converted from 4-vinyl guaiacol (4-VG), which is produced mainly by koji mold assimilating ferulic acid during awamori fermentation (Koseki et al., 1998; Maeda et al., 2018). In *S. cerevisiae*, the *PAD1* and *FDC1* genes play essential roles in catalyzing the decarboxylation of ferulic acid (Mukai et al., 2010). However, awamori yeast strain 101-18 has the *FDC1* pseudogene with a nonsense mutation, leading to the loss of 4-VG-producing activity (Mukai et al., 2014). Therefore, the isolation or construction of a yeast strain with a functional ferulic acid decarboxylase is anticipated in order to brew novel awamori.

Our objective is to develop a new kind of awamori characterized by the top notes of flavors and altered flavors during aging. It is considered that yeast metabolism greatly influences the contents of isoamyl acetate and vanillin, known as the key flavor compounds in awamori. For awamori brewing, an awamori yeast strain of *S. cerevisiae*, Awamori 101, which is a diploid prototroph (supplied by National Research Institute of Brewing, Japan), has been dominantly used as a standard and conventional strain among awamori breweries. Strain 101-18 with high alcohol productivity has been isolated from Awamori 101 and is applicable for genetic analysis and awamori fermentation test (Takagi et al., 2015). Recently, several types of awamori have been manufactured and commercialized by awamori brewery, which brewed by use of yeast strains derived and isolated from nature in Okinawa, such as mango and brown sugar lump. These unique awamori possess the different characteristics in taste and flavor compared to those prepared by conventional yeast strain. However, there has been little study done concerning the genetic analysis of newly isolated yeast strains. Therefore, it is important to clarify the genetic differences between the novel yeast strain and the other conventional yeast strains. In this study, we aimed to isolate and characterize a new yeast strain that has the potential to promote the production of these compounds in awamori brewing. To achieve this, we applied combinatorial methods for the isolation and breeding of a new yeast strain. We also presented the whole-genome information on yeast strains for awamori to uncover the molecular basis of its fermentation properties.

## MATERIALS AND METHODS

### Culture Media

*S. cerevisiae* cells were cultured in a nutrient rich medium YPD (2% glucose, 1% yeast extract, 2% peptone). A synthetic defined medium SD+Am [2% glucose, 0.5% ammonium sulfate, 0.67% yeast nitrogen base without ammonium sulfate and amino acids (Difco Laboratories)] was used for analysis of intracellular amino acid contents in yeast. A SD medium containing 0.5% allantoin as a sole nitrogen source (SD+Alt) was used for selection of HC02-5-2 mutants resistant to TFL.

### Isolation of Yeast Strains for Awamori Brewing From Hibiscus Flowers

Wild hibiscus flowers (totally 23) were collected in Okinawa, Japan. The flowers were incubated in YPD medium containing 4% ethanol at 32°C. The resultant muddy media suspension was then incubated on YPD agar plates at 32°C. After 24–48 h incubation, 5 of well-grown single colonies were picked and subsequently applied to a laboratory-scale fermentation test to select yeast strains with high ethanol productivity. The genomic DNA was extracted from yeast cells that produced high concentrations of ethanol by a laboratory-scale fermentation test. PCR was performed to amplify a part of 26S rDNA region (560 bp). PCR was conducted with the following primers: NL1 (5'-GCA TAT CAA TAA GCG GAG GAA AAG-3') and NL4

(5'-GGT CCG TGT TTC AAG ACG G-3'). The amplified product was confirmed with DNA sequencing and queried the BLAST search program to identify yeast species.

## Awamori Fermentation Test and Distillation

The method for laboratory fermentation test was described previously (Takagi et al., 2015). In brief, 50 g of rice koji, 65 ml of water, and 0.1 ml of precultured yeast cells were mixed in a 200 ml Erlenmeyer flask. The fermented mash was incubated at 25°C for 14 days to prepare final fermented mash (*moromi*). After filtration of the supernatant of *moromi*, ethanol concentrations were analyzed using portable alcohol detector (AL-3; Riken Keiki). To analyze the production of flavor, the distillation of *moromi* was performed by atmospheric distillation using a water distiller (MH943SBS; Megahome) and stopped when the ethanol concentration reached 10%. The distillate was collected and subjected to liquid chromatography and gas chromatography according to the below methods.

## Liquid and Gas Chromatography Analysis

Quantification of 4-VG was performed using Shimadzu HPLC system. Distilled awamori liquids were passed with a 0.45  $\mu$ m-pore filter. The filtrated samples were loaded on the C18 column and eluted at a flow rate of 0.5 ml/min with an acetic buffer containing methanol. The elution was started with solvent A (50 mM acetic acid, pH4) and gradually increasing solvent B (100% methanol) up to 90%. On chromatogram at 254 nm of absorbance, the peak of 4-VG was detected at the same retention time as the standard solution. The peak area was calculated and the concentrations were determined by comparing to the standard solution.

The volatile compounds were analyzed using the electric nose of Heracles II system (Alpha MOS). The system composed of the solid phase absorption of vapor from samples, the dual gas chromatography columns and the FID detection. The distilled liquids from the laboratory scale fermentation were diluted with distilled water and prepared in 15% alcohol. Each 10 ml sample in glass vials was set on the equipment. The analysis was performed according to the manufacturer's installation instruction. The volatile compounds were quantified by comparing the peak areas with the standards.

## Genome Sequencing and Phylogenetic Analysis

The extracted genome DNA from wild-type strain HC02-5-2 isolated from hibiscus flower, its mutant strain T25, and Japanese sake strain Kyokai no.7 (K7) underwent quantification by Qubit (Thermo). Next generation sequencing library was constructed for each genome using the Nextera DNA Library Preparation Kit (Illumina) according to the manufacturer's instructions. The genome libraries were sequenced using MiSeq (Illumina) with MiSeq Reagent Kit v2 or v3 (Illumina). Sequencing data processing of strains HC02-5-2, T25, and K7, as well as sequencing data from the Sequence Read Archive (SRA) was

performed with CLC Genomics Workbench v 10.1.1 (QIAGEN). This included trimming, mapping, and variants calling against the reference genome of *S. cerevisiae* S288c (GCA\_000146045). Reads bases not matching in the alignment were scored as variants. The coverage table files and the variants table files were exported from Genomics Workbench and retained for further analysis. These files were converted to a fasta file of synthetic sequences with custom scripts<sup>1</sup>. These scripts generate the sequences of homozygous SNPs from the data of coverage and variants. The fasta file was applied to the phylogenetic analysis with the Neighbor-joining method using MEGA X: Molecular Evolutionary Genetics Analysis software (Kumar et al., 2018). Parameters: Statistical method; Neighbor-joining, Substitution model; Maximum composite likelihood, Substitutions to include; d: transitions + transversions, Rates among sites; uniform rates, pattern among lineages same (homogeneous). SRA accession numbers are shown in **Supplementary Table 1**. The parent strain HC02-5-2 has 10,569 homozygous and 388 heterozygous variants to the reference genome of *S. cerevisiae* S288c. The mutant strain T25 has 10,763 homozygous and 483 heterozygous variants to the S288c reference genome. The comparison of T25 to its parental strain HC02-5-2 revealed 287 homozygous and 231 heterozygote specific variant sites.

## Isolation of TFL-Resistant Yeast Mutant

To induce mutations, HC02-5-2 cells were treated with 6% ethyl methanesulfonate (EMS) (Rose and Broach, 1991). The mutagenized cells were spread onto SD+Alt agar plates containing 40  $\mu$ g/ml of TFL and incubated at 30°C for 3 days. Resultant colonies were cultured in SD liquid medium and subsequently subjected to an amino acid analyzer for selecting the L-leucine-accumulating mutants.

## Gene Cloning and Plasmid Construction

The centromere-based low-copy-number plasmid pYC130 containing the G418 resistance gene (KanMX4) (supplied by National Research Institute of Brewing, Hiroshima, Japan) (Calera et al., 2000) and the 2  $\mu$ -based high-copy-number plasmid pAD4 containing the *ADH1* promoter and terminator (supplied by J. Nikawa, Kyushu Institute of Technology, Fukuoka, Japan) (Nikawa et al., 2006) were used to subclone and express the *LEU4* gene, respectively. *Escherichia coli* strain DH5 $\alpha$  [ $F^{-}\lambda^{-}\Phi80lacZ\Delta M15 \Delta(lacZYA argF)U169 deoR recA1 endA1 hsdR17(r_k^{+}m_k^{-}) supE44 thi-1 gyrA96$ ] was used to construct plasmids.

Full-length *LEU4* gene was amplified from the genomic DNA of HC02-5-2 cells by using KOD FX Neo polymerase (Toyobo), with addition of the *Hind*III and *Sac*I recognition site at 5' and 3' ends of the coding region of *LEU4*, respectively. PCR products were cloned into *Hind*III – *Sac*I site in pAD4 (pAD4-*LEU4*) and then *PADH1*-*LEU4*-*TADH1* fusion was amplified from pAD4-*LEU4*, with addition of the *Kpn*I and *Mlu*I recognition site at 5' and 3' ends, respectively, at the region of *PADH1*-*LEU4*-*TADH1* fusion. Amplified products were subcloned into pYC130 to construct expression plasmid

<sup>1</sup>[https://github.com/BiojetCoLtd/GW\\_to\\_phylogeny](https://github.com/BiojetCoLtd/GW_to_phylogeny)

(pYC130-LEU4) in yeast cells. Plasmids for expressing the Leu4 variants [pYC130-LEU4(G516S), pYC130-LEU4(S452F/A551V) and pYC130-LEU4(G516S/S542F/A551V)] were prepared using Quick-Change II Mutagenesis Kit (Agilent Technologies). Site-specific mutagenesis for constructing plasmids were performed with the following primers: *LEU4* (G516S) q-change Fw (5'-GAA GGT ACA GGT AAT AGT CCA ATC TCT-3'), *LEU4* (G516S) q-change Rv (5'-AGA AGA GAT TGG ACT ATT ACC TGT ACC TTC-3'), *LEU4* (S452F/A551V) q-change Fw (5'-TCT CGT AGC AAA CTA CAC AGA GCA TTT TCT AGG TTC TGG TTC TTC TAC GCA AGT TGC TTC TTA CAT CCA TC-3'), *LEU4* (S452F/A551V) q-change Rv (5'-GAT GGA TGT AAG AAG CAA CTT GCG TAG AAG AAC CAG AAC CTA GAA AAT GCT CTG TGT AGT TTG CTA CG-3'). Expression plasmids were transformed into HC02-5-2 cells using the lithium acetate method (Rose and Broach, 1991) and transformants were selected on YPD agar plates containing 200 µg/ml of G418. The DNA sequences of the amplified products with PCR and plasmids newly constructed in this study were confirmed by sequence analyses based on the Sanger method by using ABI PRISM 3130 Genetic Analyzer (Applied Biosystems).

## **α-Isopropylmalate Synthase (IPMS) Activity Assay**

Yeast cells were cultured on YPD liquid medium at 30°C for 2 days. Culture medium was centrifuged for collecting yeast cells and the pellet was suspended in 250 mM Tris-HCl (pH 8.5) containing 1 mM phenylmethylsulfonyl fluoride (PMSF). Cell suspension was treated with a Multi-Beads Shocker (Yasui Kikai) at 4°C, for preparing the whole-cell extracts by disrupting with glass beads, and resultant supernatant was used for crude enzyme solution. IPMS activity was measured as described previously (Takagi et al., 2015). One unit of enzyme activity is defined as the amount of coenzyme A (CoA) liberated from acetyl-CoA, which produced by an enzymatic transacylation reaction from 2-ketoisovalerate to 2-isopropylmalate, at 37°C for 60 min. Protein concentration was measured using a Bio-Rad Protein Assay Dye reagent concentrate, which employs the Bradford method. The protein concentrations of each sample were estimated from the index of absorbance at 595 nm on the basis of a standard curve of bovine serum albumin.

## **Quantification of Intracellular Amino Acids Contents**

Yeast cells were cultured in 5 ml of SD+Am medium containing 200 µg/ml of G418 (if necessary) at 30°C for 2 days ( $OD_{600} = 10.0$ ). Collected cells were suspended in 500 µl of distilled water and boiled at 100°C for 20 min to extract intracellular amino acids. After centrifugation at  $13,000 \times g$  for 5 min, supernatant was filtered with 0.2 µm syringe filter (mdiTm). Filtrated samples were subjected to an amino acid analyzer (AminoTacTM JLC-500/V, JEOL) or a LC/MS amino acid system (UF-Amino Station, Shimadzu) for quantifying amino acids contents in yeast cells. Experimental procedures for analyzing amino acid content by LC/MS were conducted as

reported previously (Shimbo et al., 2009). The content of each amino acid was expressed as a percentage of dry cell weight.

## **Homology Modeling of IPMS**

To analyze the effect of each amino acid substitution in IPMS, we constructed the wild-type and variant IPMS structures by homology modeling using SWISS-MODEL<sup>2</sup>. The template structure used for modeling was the structure of IPMS (LeuA) bound to L-leucine from *Mycobacterium tuberculosis* (PDB ID code: 3FIG). The amino acid sequence identity of the *M. tuberculosis* LeuA with IPMS from strain HC02-5-2 was 45.37%.

## **RESULTS**

### **Isolation and Characterization of Yeast Strain From Hibiscus Flowers for Awamori Brewing**

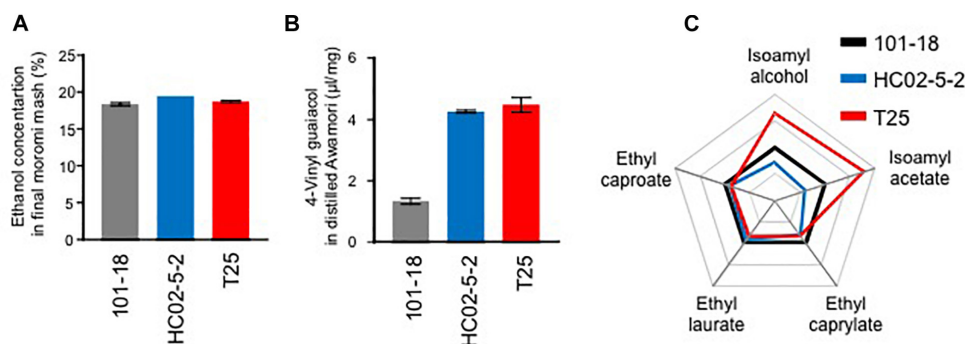
We aimed to obtain a new yeast strain applicable to awamori brewing. Microorganisms were screened from wild hibiscus flowers in Okinawa for ethanol production. Among the isolated yeast colonies, strain HC02-5-2, with high-ethanol productivity, was further applied to genetic analysis in order to identify species. The BLAST search revealed that the 26S rDNA sequence of strain HC02-5-2 was identical to that of the yeast *S. cerevisiae*.

To examine the applicability of hibiscus yeast HC02-5-2 for awamori brewing, we performed a laboratory-scale awamori fermentation test. First, the concentration of ethanol in moromi fermented with HC02-5-2 reached 18.75%, whereas strain 101-18, a conventional awamori yeast, reached 17.71% (Figure 1A). This indicates that strain HC02-5-2 produces sufficient ethanol during fermentation.

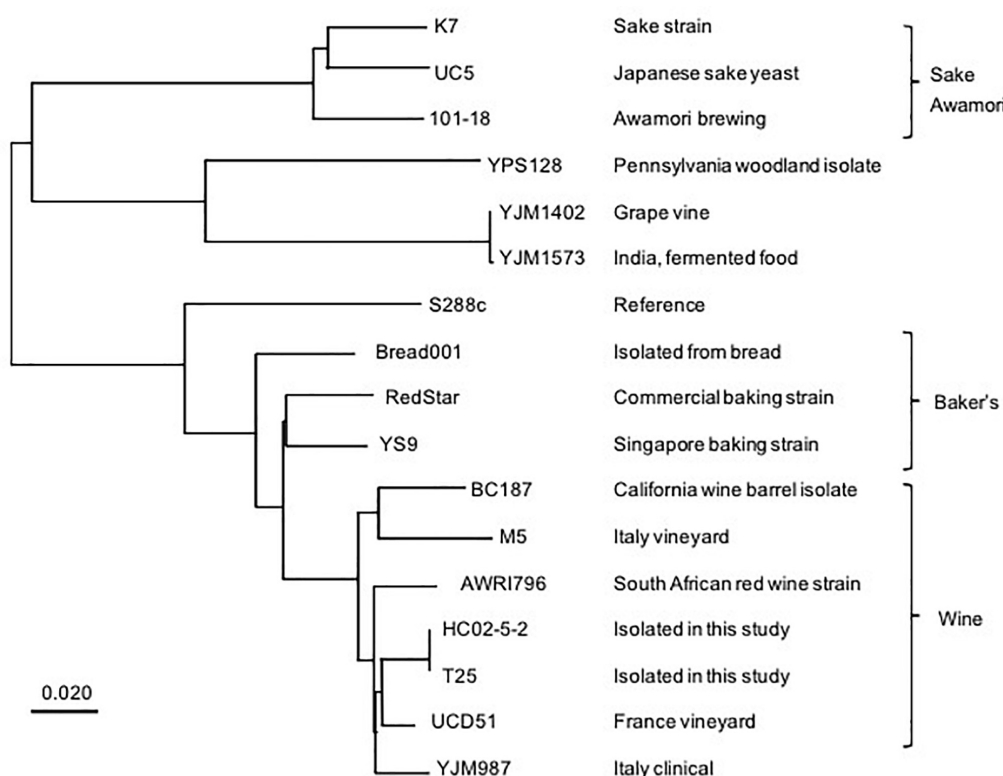
### **Comparative Analysis of Yeast Whole Genomes**

We then conducted next-generation sequencing for the genome of strain HC02-5-2. As a result, 99.01% reads were mapped to the S288c reference genome and mean sequencing depth exhibited x113, confirming that strain HC02-5-2 belongs to *S. cerevisiae*. To examine the relationship between strain HC02-5-2 and other yeast strains used for fermentation, a phylogenetic analysis using whole genome information was performed with comparing single-nucleotide variants (SNVs). Interestingly, the depicted phylogenetic tree showed that strains HC02-5-2 and 101-18 were assigned to different clades of the tree. Strain HC02-5-2 is in a clade that includes yeast strains for wine brewing, whereas awamori strain 101-18 is in a clade that includes yeast strains for sake and shochu brewing (Figure 2). The evolutionary tree was also supported by phylogenetic analysis focusing on specific genes, as shown in previous reports (Fay and Benavides, 2005; Futagami et al., 2017). These results suggest that strain HC02-5-2 does not share ancestry with sake or shochu yeast strains.

<sup>2</sup><http://swissmodel.expasy.org/>



**FIGURE 1** | A laboratory-scale fermentation test of strains 101-18, HC02-5-2, and T25. **(A)** Ethanol concentrations in final fermented mash (*moromi*). **(B)** 4-Vinyl guaiacol concentrations (4-VG) in distilled *moromi*. The data are the mean, maximum, and minimum value of results from two independent experiments. **(C)** A radar chart of the initial volatile compounds. The each scale was standardized by the amounts derived from strain 101-18.



**FIGURE 2** | A phylogenetic tree based on whole genome analysis of various *S. cerevisiae* strains. A single nucleotide variants-based phylogenetic analysis was performed using the Neighbor-joining method. The BioSample database (<https://www.ncbi.nlm.nih.gov/biosample/>) information was used as a reference for the description of yeast strains. Scale bar represents 0.020 nucleotide changes per site.

## Evaluation of the Flavor Compounds in Awamori Brewed by Hibiscus Yeast

The contents of odorants in awamori produced in a laboratory-scale test were measured. First, 4-VG was quantified by liquid chromatography. Its concentration observed in strain HC02-5-2 (6.47 μg/ml) was approximately 3 times higher than that observed in strain 101-18 (1.99 μg/ml) (**Figure 1B**). Based on a previous report (Mukai et al., 2010), the variants in the

*PAD1* and *FDC1* genes essential for the decarboxylation of phenylacrylic acids were compared. It was confirmed that the intact and protein-coding *PAD1* and *FDC1* genes were present in the HC02-5-2 genome, whereas a nonsense mutant was found in the *FDC1* gene in strain 101-18. Next, the initial volatile compounds in awamori were quantified by gas chromatography. The concentration ratios of isoamyl alcohol and isoamyl acetate observed in strain HC02-5-2 were

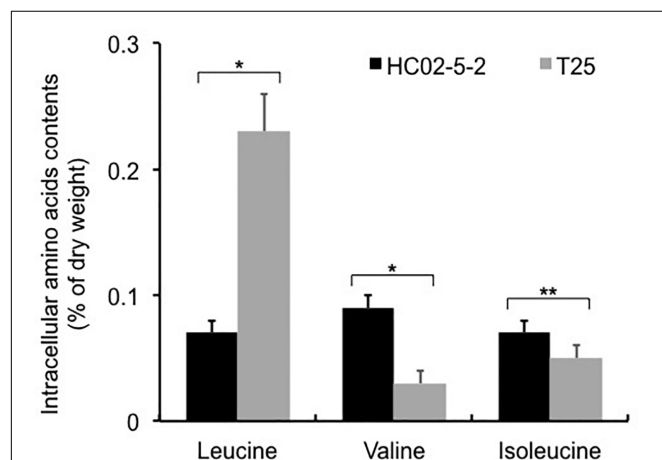


lower than those observed in strain 101-18 (Figure 1C). Although the initial volatile compounds in awamori brewed with strain HC02-5-2 were not prominent, the greater 4-VG production obtained with strain HC02-5-2 is desirable for awamori brewing.

## Isolation of Hibiscus Yeast Mutants With L-Leucine Accumulation From TFL-Resistant Mutants

We previously isolated TFL-resistant mutant strain 18-T55 with L-leucine accumulation from awamori yeast strain 101-18 (Takagi et al., 2015). By brewing with strain 18-T55, which overproduces isoamyl alcohol and isoamyl acetate, a new kind of awamori has been successfully commercialized. Therefore, to give distinctive characteristics to hibiscus yeast HC02-5-2, we attempted to isolate the TFL-resistant mutants that accumulate intracellular L-leucine. Strain HC02-5-2 was incubated in the presence of 6% EMS for 1 h and EMS-treated cells, which showed 30% of survival rate of untreated cells, were directly plated on SD agar medium containing 40  $\mu$ g/ml of TFL. As a result, many TFL-resistant colonies were obtained, and among them one mutant strain, T25, exhibited larger amounts of L-leucine in the cells.

Next, we determined the intracellular L-leucine quantities in both parent strain HC02-5-2 and mutant strain T25. As we expected, the intracellular concentration of L-leucine in strain T25 ( $0.225 \pm 0.033\%$  of dry cell weight) was significantly higher than that in parent strain HC02-5-2 ( $0.066 \pm 0.006\%$  of dry cell weight) (Figure 3). Interestingly, decreased concentrations of both L-valine and L-isoleucine were observed in strain T25 (Figure 3), suggesting that one or more mutations in strain T25 affect the biosynthetic pathway of branched-chain amino acids.



**FIGURE 3 |** Intracellular amino acids contents in strains HC02-5-2 and T25. Yeast cells were cultured in 5 ml of SD+Am medium at 30°C for 2 days ( $OD_{600} = 10.0$ ). Black and gray bars indicate strain HC02-5-2 and T25, respectively. The values are the means and standard deviations of results from three independent experiments. Statistically significant differences were determined by Student's *t*-test (\* $p < 0.01$  and \*\* $p < 0.05$ ).

## IPMS Activity and the LEU4 Locus of a Hibiscus Yeast Mutant With L-Leucine Accumulation

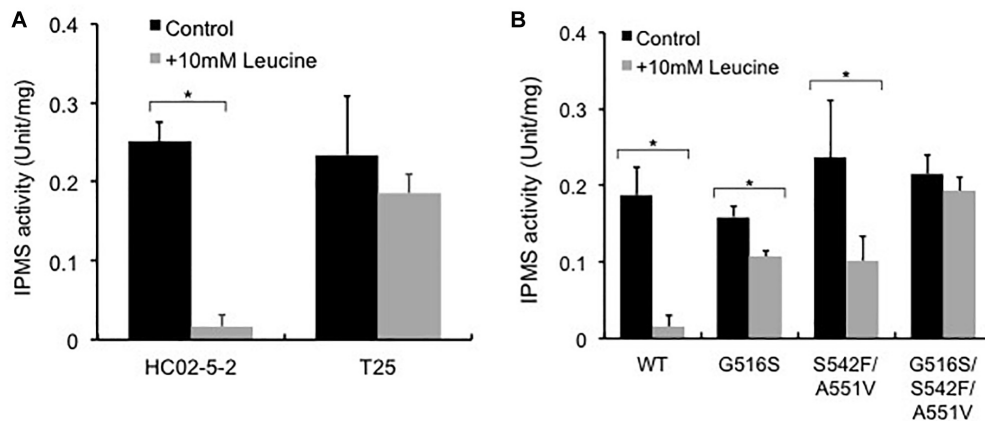
Previous reports indicated that reduced sensitivity to L-leucine feedback inhibition in the IPMS variants causes oversynthesis of L-leucine in the cell (Oba et al., 2005; Takagi et al., 2015). To elucidate the molecular mechanisms underlying L-leucine accumulation in strain T25, we examined IPMS activity in strains T25 and HC02-5-2. The IPMS activity in the presence or absence of 10 mM L-leucine was assayed using the crude cell extracts from both strains. As we expected, the IPMS activity from strain HC02-5-2 was remarkably inhibited by 10 mM L-leucine, indicating that the IPMS activity of this strain is regulated by L-leucine feedback inhibition. In contrast, the level of IPMS activity from strain T25 was approximately 80% even in the presence of 10 mM L-leucine, indicating that IPMS in mutant strain T25 is less sensitive to feedback inhibition by L-leucine than the parent strain HC02-5-2 (Figure 4A).

Next, we determined the DNA sequences of the *LEU4* gene coding IPMS in strains HC02-5-2 and T25. Gene amplification by PCR and the subsequent sequencing result showed that strain T25 has a mixture of nucleotides A and G at position 1,546, whereas strain HC02-5-2 has a G at the same position. The mutation of G to A leads to the amino acid replacement of Gly to Ser at position 516, suggesting that strain T25 has a heteroallelic mutant of the *LEU4* gene.

## Effects of the LEU4 Mutations on IPMS Activity and L-Leucine Biosynthesis

To analyze the *LEU4* mutation identified in strain T25, the expression plasmids for the *LEU4* mutants were introduced to strain HC02-5-2. First, we assayed IPMS activity in the transformants overexpressing the *LEU4*, *LEU4*<sup>G516S</sup>, *LEU4*<sup>S542F/A551V</sup>, and *LEU4*<sup>G516S/S542F/A551V</sup> genes. When the wild-type *LEU4* gene was overexpressed, IPMS activity was markedly inhibited in the presence of 10 mM L-leucine. As previously reported (Takagi et al., 2015), IPMS activity in cells overexpressing the *LEU4*<sup>S542F/A551V</sup> gene was less sensitive than the wild-type *LEU4* gene to L-leucine. We also found that overexpression of the *LEU4*<sup>G516S</sup> and *LEU4*<sup>G516S/S542F/A551V</sup> genes increased IPMS activity in the presence of L-leucine (Figure 4B). These results indicated that the IPMS activity in strain T25 was mimicked by *LEU4*<sup>G516S</sup> overexpression.

To check whether the *LEU4*<sup>G516S</sup> gene is sufficient to confer TFL-resistance and L-leucine accumulation in strain HC02-5-2, we examined the growth of transformants on SD agar plates containing TFL and determined the intracellular L-leucine levels. Yeast cells overexpressing the wild-type *LEU4* gene could not grow in the presence of TFL (Figure 5A). On the other hand, overexpression of the *LEU4*<sup>G516S</sup>, *LEU4*<sup>S542F/A551V</sup>, and *LEU4*<sup>G516S/S542F/A551V</sup> genes resulted in resistance to TFL. In these transformants, the intracellular L-leucine concentration was higher than that in wild-type strain HC02-5-2 (Figure 5B). These results indicate that the Gly516Ser variant of IPMS reduces sensitivity to L-leucine, leading to L-leucine accumulation in the hibiscus yeast strain.



**FIGURE 4 |** Effect of L-leucine on IPMS activity in (A) strains HC02-5-2 and T25, and (B) strain HC02-5-2 expressing the wild-type *LEU4* (WT), *LEU4*<sup>G516S</sup> (G516S), *LEU4*<sup>S542F/A551V</sup> (S542F/A551V), and *LEU4*<sup>G516S/S542F/A551V</sup> (G516S/S542F/A551V). Yeast cell extract was used as enzyme solution. One unit of IPMS activity was defined as the amount of CoA generated from acetyl-CoA by the enzyme reaction at 37°C for 60 min. Black and gray bars indicate IPMS activities in the absence or presence of 10 mM L-leucine, respectively. The values are the means and standard deviations of results from three independent experiments. Statistically significant differences were determined by Student's *t*-test (\**p* < 0.01).

## Awamori Fermentation Test for T25 and Analysis of Ethanol and Flavor Compounds

Finally, we evaluated the potential of strain T25 for awamori brewing. First, to determine ethanol productivity, a laboratory-scale fermentation test was carried out. The concentration of ethanol in final moromi fermented with strain T25 reached 18.05%, which is equivalent to that with strains HC02-5-2 and 101-18 (Figure 1A). Next, we analyzed distilled awamori by gas chromatography. As we expected, in proportion to the cellular L-leucine level, strain T25 produced 2.3-fold more isoamyl alcohol and isoamyl acetate than strain HC02-5-2 (Figure 6). On the other hand, the concentrations of ethyl caprylate, ethyl laurate, and ethyl caproate were almost the same among the 3 strains (HC02-5-2, T25, and 101-18) (Figure 1C), suggesting that breeding selectively promotes the production of odorant compounds in awamori. Thus, mutant strain T25 derived from hibiscus yeast is expected to be applicable to awamori brewing.

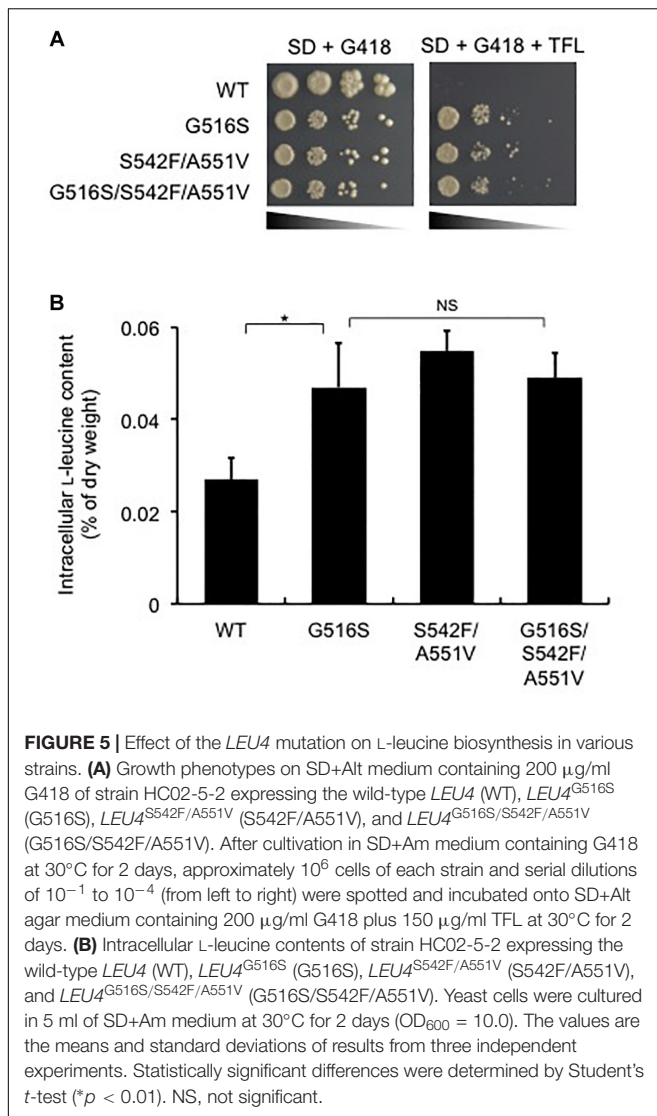
## DISCUSSION

By virtue of its high productivity of ethanol and favorable flavors, strain 101-18, an awamori yeast strain of *S. cerevisiae*, has been used commercially to brew awamori by most awamori manufacturers (Shinzato et al., 1989). Since demands to expand the diversity of awamori qualities have increased, the development of novel yeast strains that vary in taste and flavor may contribute to the awamori industry. In this study, we found that novel yeast strain HC02-5-2, isolated from hibiscus flowers, produced enough ethanol for awamori brewing. In fact, “Hibiscus Awamori”, a new kind of awamori brewed with strain HC02-5-2, has been sold on the Japanese market. Subsequent breeding of strain HC02-5-2 succeeded in obtaining the desirable mutant

T25, which produces more isoamyl alcohol and isoamyl acetate than HC02-5-2, which imbue awamori with fruity flavors. In a pilot-scale fermentation, greater amounts of isoamyl alcohol and isoamyl acetate were produced in awamori fermented by strain T25 than in those produced by strain 101-18 (data not shown). Our data indicate that strains HC02-5-2 and T25 are suitable for awamori brewing.

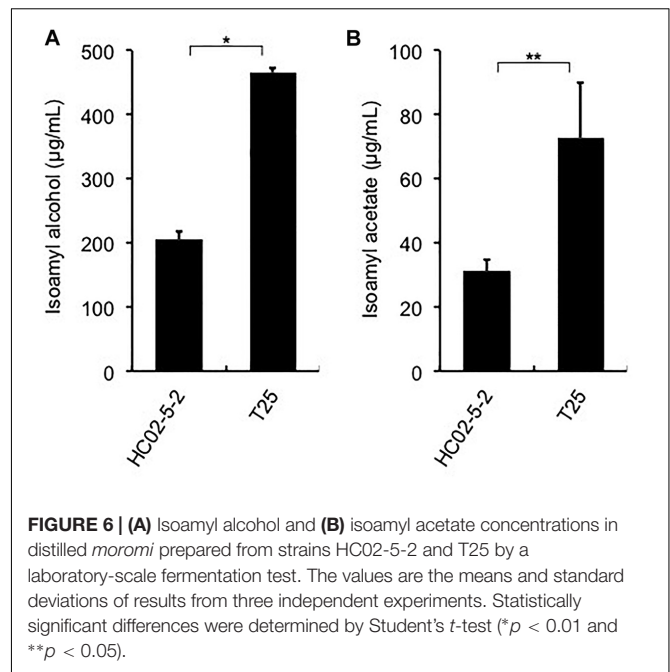
Strain HC02-5-2 produced more 4-VG production than strain 101-18 (Figure 1B). For several alcoholic beverages, such as beer, 4-VG is recognized as phenolic off flavor (POF). A comparative study revealed that the yeast strains with strong domestication like beer strains exhibit the POF negative phenotype (Gallone et al., 2016). In awamori, 4-VG is known as a precursor compound of vanillin, which confers sweet vanilla flavor and is characteristic of aged awamori (Maeda et al., 2018). During storage, awamori flavors are commonly believed to change, and storage methods for beneficial aging have been established (Kanauchi, 2012). One of the molecular bases for awamori aging is explained by the conversion of 4-VG into vanillin (Koseki et al., 1996). Therefore, strain HC02-5-2 is suggested to have a distinct feature for awamori brewing, whereas strain 101-18 is a POF-negative strain with a nonsense mutation in the *FDC1* gene.

The initial scented fruity flavors are preferable to awamori and sake consumers. Several studies have increased fruity aromatic compounds in Japanese sake by introducing spontaneous one or more mutations into industrial yeast strains (Arikawa et al., 2000; Takahashi et al., 2017). We recently reported that L-leucine analog-resistant mutants overproduced isoamyl alcohol and isoamyl acetate (Takagi et al., 2015). Since strain HC02-5-2 showed no striking features in initial volatile compounds, we expected to breed this yeast strain for a diversity of tastes and flavors. As a result, the obtained mutant strain T25 with L-leucine accumulation produced more isoamyl alcohol and isoamyl acetate than its parent strain HC02-5-2. It is known that



an increase in L-leucine leads to high-levels of isoamyl acetate production (Ashida et al., 1987; Quilter et al., 2003; Takagi et al., 2015). Moreover, strain T25 produced more isoamyl acetate than strain HC02-5-2 did. Isoamyl acetate is converted mainly from isoamyl alcohol catalyzed by alcohol acetyltransferase (the *ATF1* gene product) in *S. cerevisiae* (Inoue et al., 1997). In the present study, we did not observe any sequence difference in the *ATF1* gene between strains HC02-5-2 and T25. Further investigation is needed to understand the mechanism by which strain T25 overproduces isoamyl acetate.

Interestingly, strain T25 produced approximately 3 times more L-leucine than the parent strain, whereas intracellular levels of L-valine and L-isoleucine in strain T25 were decreased to about 30 and 70% of those observed in the parent strain, respectively. Previous studies have shown that the balance of intracellular amino acids is tightly controlled in *S. cerevisiae* (Watson, 1976; Messenguy et al., 1980). Moreover, L-valine and L-isoleucine, which are categorized as branched-chain amino acids (BCAAs)



containing L-leucine are bioynthesized by commonly sharing the same enzymes (acetolactate synthase, acetohydroxiacid reductoisomerase, and dihydroxiacid dehydratase) in the first four steps in mitochondria (Kohlhaw, 2003). It is possible that an increase in L-leucine accounts for a decrease in other BCAAs in yeast cells. We found that L-leucine accumulation in strain T25 is caused by expression of the Gly516Ser variant of IPMS (Figures 3, 5B). IPMS is the key enzyme that regulates L-leucine biosynthesis via the mechanism of negative feedback inhibition by L-leucine in *S. cerevisiae*. Homology modeling analysis suggests that Gly516 is located on  $\alpha$  14 helix comprising the L-leucine binding site, which is the allosteric regulation domain with L-leucine in the bacterial IPMS (Koon et al., 2004). Furthermore, the amino acid change of Gly to Ser at position 516 in IPMS was supposed to directly interfere with L-leucine binding due to the steric hindrance at the binding cleft, leading to the desensitization of the L-leucine feedback inhibition of IPMS (Supplementary Figure 1).

## CONCLUSION

In conclusion, newly isolated strain HC02-5-2 from hibiscus flowers was identified as a *S. cerevisiae* strain related to the wine lineage of this species. Strains HC02-5-2 and its mutant T25 possess favorable characteristics for developing both the initial scented fruity flavors and the sweet flavors associated with aging. Our data supported the practical use of these isolated yeast strains. Moreover, we can now explore the certain molecular basis of fermentation properties of these strains. Since spirits contain many fragrant ingredients and their balance determines the quality, it is important to control multiple flavor compounds. This combinatorial approach to yeast isolation from nature and

its breeding is applicable to the variation of the quality of alcoholic beverages in the fermentation industry.

TA, YT, and HT wrote the manuscript. All authors reviewed and approved the final version of manuscript.

## AUTHOR CONTRIBUTIONS

DW, MT, and HT conceived the study and designed the experiments. TA, YT, YS, HA, and KT performed the experiments. TA, YT, RN, DW, MT, and HT analyzed the data.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2019.00490/full#supplementary-material>

## REFERENCES

- Arikawa, Y., Yamada, M., Shimosaka, M., Okazaki, M., and Fukuzawa, M. (2000). Isolation of sake yeast mutants producing a high level of ethyl caproate and/or isoamyl acetate. *J. Biosci. Bioeng.* 90, 675–677. doi: 10.1016/S1389-1723(00)90016-1
- Ashida, S., Ichikawa, E., Suginami, K., and Imayasu, S. (1987). Isolation and application of mutants producing sufficient isoamyl acetate, a sake flavor component. *Agric. Biol. Chem.* 51, 2061–2065. doi: 10.1271/bbb1961.51.2061
- Baichwal, V. R., Cunningham, T. S., Gatzek, P. R., and Kohlhaw, G. B. (1983). Leucine biosynthesis in yeast. *Curr. Genet.* 7, 369–377. doi: 10.1007/BF00445877
- Calera, J. A., Herman, D., and Calderone, R. (2000). The pYC plasmids, a series of cassette-based yeast plasmid vectors providing means of counter-selection. *Yeast* 16, 1035–1043. doi: 10.1002/1097-0061(200008)16:11<1035::aid-yea606>3.0.co;2-p
- Fay, J. C., and Benavides, J. A. (2005). Evidence for domesticated and wild populations of *Saccharomyces cerevisiae*. *PLoS Genet.* 1:e5. doi: 10.1371/journal.pgen.0010005
- Futagami, T., Kadooka, C., Ando, Y., Okutsu, K., Yoshizaki, Y., Setoguchi, S., et al. (2017). Multi-gene phylogenetic analysis reveals that shochu-fermenting *Saccharomyces cerevisiae* strains form a distinct sub-clade of the Japanese sake cluster. *Yeast* 34, 407–415. doi: 10.1002/yea.3243
- Gallone, B., Steensels, J., Prah, T., Soriaga, L., Saels, V., Herrera-Malaver, B., et al. (2016). Domestication and divergence of *Saccharomyces cerevisiae* beer yeasts. *Cell* 166, 1397.e–1410.e. doi: 10.1016/j.cell.2016.08.020
- Inoue, Y., Trevanichi, S., Fukuda, K., Izawa, S., Wakai, Y., and Kimura, A. (1997). Roles of esterase and alcohol acetyltransferase on production of isoamyl acetate in *Hansenula mrakii*. *J. Agric. Food Chem.* 45, 644–649. doi: 10.1021/jf960648o
- Kanauchi, M. (2012). “Characteristics and role of feruloyl esterase from *Aspergillus awamori* in Japanese spirits, ‘awamori’ production,” in *Scientific, Health and Social Aspects of the Food Industry*, ed. B. Valdez (London: InTech), 145–162. doi: 10.5772/31190
- Kohlhaw, G. B. (2003). Leucine biosynthesis in fungi: entering metabolism through the back door. *Microbiol. Mol. Biol. Rev.* 67, 1–15. doi: 10.1128/MMBR.67.1.1-15.2003
- Koon, N., Squire, C. J., and Baker, E. N. (2004). Crystal structure of LeuA from *Mycobacterium tuberculosis*, a key enzyme in leucine biosynthesis. *Proc. Natl. Acad. Sci. U.S.A.* 101, 8295–8300. doi: 10.1073/pnas.0400820101
- Koseki, T., Furuse, S., Iwano, K., and Matsuzawa, H. (1998). Purification and characterization of a feruloyl esterase from *Aspergillus awamori*. *Biosci. Biotechnol. Biochem.* 62, 2032–2034. doi: 10.1271/bbb.62.2032
- Koseki, T., Ito, Y., Furuse, S., Ito, K., and Iwano, K. (1996). Conversion of ferulic acid into 4-vinylguaiacol, vanillin and vanillic acid in model solutions of shochu. *J. Ferment. Bioeng.* 82, 46–50. doi: 10.1016/0922-338X(96)89453-0
- Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* 35, 1547–1549. doi: 10.1093/molbev/msy096
- Maeda, M., Tokashiki, M., Tokashiki, M., Uechi, K., Ito, S., and Taira, T. (2018). Characterization and induction of phenolic acid decarboxylase from *Aspergillus luchuensis*. *J. Biosci. Bioeng.* 126, 162–168. doi: 10.1016/j.jbiosc.2018.02.009
- Messenguy, F., Colin, D., and ten Have, J. P. (1980). Regulation of compartmentation of amino acid pools in *Saccharomyces cerevisiae* and its effects on metabolic control. *Eur. J. Biochem.* 108, 439–447. doi: 10.1111/j.1432-1033.1980.tb04740.x
- Mukai, N., Masaki, K., Fujii, T., and Iefuji, H. (2014). Single nucleotide polymorphisms of *PAD1* and *FDC1* show a positive relationship with ferulic acid decarboxylation ability among industrial yeasts used in alcoholic beverage production. *J. Biosci. Bioeng.* 118, 50–55. doi: 10.1016/j.jbiosc.2013.12.017
- Mukai, N., Masaki, K., Fujii, T., Kawamukai, M., and Iefuji, H. (2010). *PAD1* and *FDC1* are essential for the decarboxylation of phenylacrylic acids in *Saccharomyces cerevisiae*. *J. Biosci. Bioeng.* 109, 564–569. doi: 10.1016/j.jbiosc.2009.11.011
- Nikawa, J.-I., Yata, M., Motomura, M., Miyoshi, N., Ueda, T., and Hisada, D. (2006). Ternary complex formation of Ino2p-Ino4p transcription factors and Apl2p adaptin  $\beta$  subunit in yeast. *Biosci. Biotechnol. Biochem.* 70, 2604–2612. doi: 10.1271/bbb.60146
- Nishiyama, T. (1980). Honkaku shochu no koumi. *J. Soc. Brewing Jpn.* 75, 944–952.
- Oba, T., Nomiyama, S., Hirakawa, H., Tashiro, K., and Kuhara, S. (2005). Asp578 in LEU4p is one of the key residues for leucine feedback inhibition release in sake yeast. *Biosci. Biotechnol. Biochem.* 69, 1270–1273. doi: 10.1271/bbb.69.1270
- Quilter, M. G., Hurley, J. C., Lynch, F. J., and Murphy, M. G. (2003). The production of isoamyl acetate from amy alcohol by *Saccharomyces cerevisiae*. *J. Inst. Brewing* 109, 34–40. doi: 10.1002/j.2050-0416.2003.tb00591.x
- Rose, M. D., and Broach, J. R. (1991). Cloning genes by complementation in yeast. *Methods Enzymol.* 194, 195–230. doi: 10.1016/0076-6879(91)94017-7
- Shimbo, K., Oonuki, T., Yahashi, A., Hirayama, K., and Miyano, H. (2009). Precolumn derivatization reagents for high-speed analysis of amines and amino acids in biological fluid using liquid chromatography/electrospray ionization tandem mass spectrometry. *Rapid Commun. Mass Spectrom.* 23, 1483–1492. doi: 10.1002/rcm.4026
- Shinzato, S., Miyagi, T., Takaesu, C., and Maruyama, S. (1989). Characteristics of a non-foaming awamori yeast isolated from awamori No.1 yeast. *J. Brew. Soc. Jpn.* 84, 121–123. doi: 10.6013/jbrewsocjapan1988.84.121
- Taira, J., Tsuchiya, A., and Furudate, H. (2012). Initial volatile aroma profiles of young and aged awamori shochu determined by GC/MS/pulsed FPD. *Food Sci. Technol. Res.* 18, 177–181. doi: 10.3136/fstr.18.177
- Takagi, H., Hashida, K., Watanabe, D., Nasuno, R., Ohashi, M., Iha, T., et al. (2015). Isolation and characterization of awamori yeast mutants with l-leucine accumulation that overproduce isoamyl alcohol. *J. Biosci. Bioeng.* 119, 140–147. doi: 10.1016/j.jbiosc.2014.06.020
- Takahashi, T., Ohara, Y., and Sueno, K. (2017). Breeding of a sake yeast mutant with enhanced ethyl caproate productivity in sake brewing using rice milled at a high polishing ratio. *J. Biosci. Bioeng.* 123, 707–713. doi: 10.1016/j.jbiosc.2017.01.014
- Watson, T. G. (1976). Amino-acid pool composition of *Saccharomyces cerevisiae* as a function of growth rate and amino-acid nitrogen source. *J. Gen. Microbiol.* 96, 263–268. doi: 10.1099/00221287-96-2-263

**Conflict of Interest Statement:** TA, HA, KT, and MT were employed by BioJet Co.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Abe, Toyokawa, Sugimoto, Azuma, Tsukahara, Nasuno, Watanabe, Tsukahara and Takagi. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.





# Controlled Reduction of Genomic Heterozygosity in an Industrial Yeast Strain Reveals Wide Cryptic Phenotypic Variation

Nadia M. V. Sampaio<sup>1,2†</sup>, Ruth A. Watson<sup>1</sup> and Juan Lucas Argueso<sup>1,2\*</sup>

## OPEN ACCESS

### Edited by:

Isabel Sá-Correia,  
iBB-Institute for Bioengineering and  
Biosciences (IST), Portugal

### Reviewed by:

Philippe Marullo,  
BIOLAFFORT, France  
Jean-Baptiste Leducq,  
Université du Québec à Montréal,  
Canada  
Jean-luc Legras,  
Institut National de la Recherche  
Agronomique Centre Montpellier,  
France

### \*Correspondence:

Juan Lucas Argueso  
lucas.argueso@colostate.edu

### †Present address:

Nadia M.V. Sampaio,  
Department of Biomedical  
Engineering, Boston University,  
Boston, MA, United States

### Specialty section:

This article was submitted to  
Evolutionary and Genomic  
Microbiology,  
a section of the journal  
Frontiers in Genetics

**Received:** 07 January 2019

**Accepted:** 24 July 2019

**Published:** 11 September 2019

### Citation:

Sampaio NMV, Watson RA and  
Argueso JL (2019) Controlled  
Reduction of Genomic Heterozygosity  
in an Industrial Yeast Strain Reveals  
Wide Cryptic Phenotypic Variation.  
Front. Genet. 10:782.  
doi: 10.3389/fgene.2019.00782

<sup>1</sup> Department of Environmental and Radiological Health Sciences, Colorado State University, Fort Collins, CO, United States,  
<sup>2</sup> Cell and Molecular Biology Graduate Program, Colorado State University, Fort Collins, CO, United States

Abundant genomic heterozygosity can be found in wild strains of the budding yeast *Saccharomyces cerevisiae* isolated from industrial and clinical environments. The extent to which heterozygosity influences the phenotypes of these isolates is not fully understood. One such case is the PE-2/JAY270 strain, a natural hybrid widely adopted by sugarcane bioethanol distilleries for its ability to thrive under harsh biotic and abiotic stresses during industrial scale fermentation, however, it is not known whether or how the heterozygous configuration of the JAY270 genome contributes to its many desirable traits. In this study, we took a step toward exploring this question by conducting an initial functional characterization of JAY270's heteroalleles. We manipulated the abundance and distribution of heterozygous alleles through inbreeding and targeted uniparental disomy (UPD). Unique combinations of homozygous alleles in each inbred strain revealed wide phenotypic variation for at least two important industrial traits: Heat stress tolerance and competitive growth. Quantitative trait loci analyses allowed the identification of broad genomic regions where genetic polymorphisms potentially impacted these traits, and there was no overlap between the loci associated with each. In addition, we adapted an approach to induce bidirectional UPD of three targeted pairs of chromosomes (IV, XIV, and XV), while heterozygosity was maintained elsewhere in the genome. In most cases UPD led to detectable phenotypic alterations, often in opposite directions between the two homozygous haplotypes in each UPD pair. Our results showed that both widespread and regional homozygosity could uncover cryptic phenotypic variation supported by the heteroalleles residing in the JAY270 genome. Interestingly, we characterized multiple examples of inbred and UPD strains that displayed heat tolerance or competitive growth phenotypes that were superior to their heterozygous parent. However, we propose that homozygosity for those regions may be associated with a decrease in overall fitness in the complex and dynamic distillery environment, and that may have contributed to slowing down the erosion of heterozygosity from the JAY270 genome. This study also laid a foundation for approaches that can be expanded to the identification of specific alleles of interest for industrial applications in this and other hybrid yeast strains.

**Keywords:** Loss-of-Heterozygosity (LOH), *Saccharomyces cerevisiae*, industrial yeast, bioethanol, uniparental disomy

## INTRODUCTION

In the budding yeast *Saccharomyces cerevisiae*, abundant heterozygosity appears to be prevalent in strains isolated from clinical and industrial settings (Borneman et al., 2011; Magwene et al., 2011; Cromie et al., 2013; Borneman et al., 2016; Peter et al., 2018). One of the first heterozygous wild strains to have its genome characterized was PE-2/JAY270 (referred to here simply as JAY270) (Argueso et al., 2009). This strain was originally isolated as an aggressive wild contaminant of sugarcane-based batch-fed fermentations (Basso et al., 2008). In addition to robust competitive growth, this strain also displays excellent fermentation yield and stress tolerance traits, thus it was selected for commercial propagation, and has since been widely adopted by bioethanol distilleries as a primary inoculum (Basso et al., 2008; Della-Bianca et al., 2013).

The industrial environment where JAY270 thrives represents an interesting model for studying the dynamics of microbial populations. During each batch of fermentation, cells are exposed to significant and variable biotic and abiotic stresses, including high osmotic pressure that transitions to ethanol toxicity, oxidative and heat stresses, and steady introduction of wild bacterial and fungal contaminants (Amorim et al., 2011). In addition, a peculiar feature of this system is that the microbial population is recycled twice daily from one batch to the next for up to eight consecutive months during the sugarcane harvest season. The combination of these factors creates a highly competitive environment, in which the most adapted yeast strains persist and may evolve over time. JAY270's defining characteristic is its extraordinary ability to out-compete external contaminants in this environment, dominating the microbial population in the distillery and thus ensuring stable and predictable operational conditions (Basso et al., 2008).

The genetic characterization of JAY270 suggests this strain was formed as a natural hybrid that resulted from the mating of two diverged parent haploid strains (Argueso et al., 2009; Rodrigues-Prause et al., 2018). Analogous examples of such mosaic strains have been described recently, including yeasts used in the production of distilled alcoholic beverages from sugarcane juice (Barbosa et al., 2018; Legras et al., 2018; Peter et al., 2018). JAY270 is heterothallic (i.e., its meiotic spores are unable to switch mating type to self-mate and generate fully homozygous diploids), and it has a complex diploid genomic architecture, marked by abundant structural and single nucleotide polymorphisms between most pairs of homologous chromosomes (Argueso et al., 2009). This heterozygous genomic architecture is also a feature of other bioethanol strains (e.g., CAT-1, BG-1) that, like JAY270, were isolated as robust contaminants at sugarcane distilleries (Babrzadeh et al., 2012; Carvalho-Netto et al., 2013; Della-Bianca et al., 2013; Coutoune et al., 2017).

We recently mapped the distribution of heterozygous loci in JAY270 (Figure S1; Rodrigues-Prause et al., 2018) and found that heterozygosity is not uniformly distributed across its genome. Instead, only ~60% of the genome corresponds to regions with a high density of heterozygous loci, interspersed by long homozygous regions. Thus, by the time this strain was isolated, ~40% of the heterozygosity originally present in the

ancestral hybrid diploid had already eroded away through cycles of mitotic and/or meiotic recombination. Presumably, the heteroalleles formerly present at those regions were likely dispensable for JAY270's distinctive performance in the sugarcane fermentation environment. An intriguing question that follows is whether some of the heteroalleles that remain in the genome contribute to the desirable industrial traits that JAY270 displays today.

In this study, we took a step toward exploring this question by conducting an initial functional characterization of the heteroalleles present in the JAY270 genome. We employed two different approaches to reduce genomic heterozygosity, and then systematically assessed the phenotypic consequences of loss of heterozygosity (LOH). In the primary approach, we used controlled inbreeding to generate a collection of experimental strains, each harboring a unique combination of homozygous alleles distributed genome-wide. We compared the phenotypes of those inbred strains to their fully heterozygous parent (JAY270) under various culture conditions and identified candidate genomic regions where genetic polymorphisms impacted two important industrial traits: Heat stress tolerance and competitive co-culture growth kinetics. In a second, more conservative approach, we constructed strain sets in which bidirectional LOH was restricted to one chromosome pair at a time (uniparental disomy; UPD), while preserving heterozygosity elsewhere. We found that UPD also affected the two traits above, and did so in a way that was specific to the chromosomes and haplotypes that were made homozygous in each strain. Taken together, our results showed that a wide phenotypic variation can be uncovered by shuffling the combinations of heteroalleles present in JAY270. We interpreted these results in light of a model in which the current heterozygous genomic configuration of this strain may correspond to an optimal set of alleles which collectively allow it to be highly versatile, and thus well adapted to long term propagation in the industrial sugarcane fermentation environment.

## RESULTS

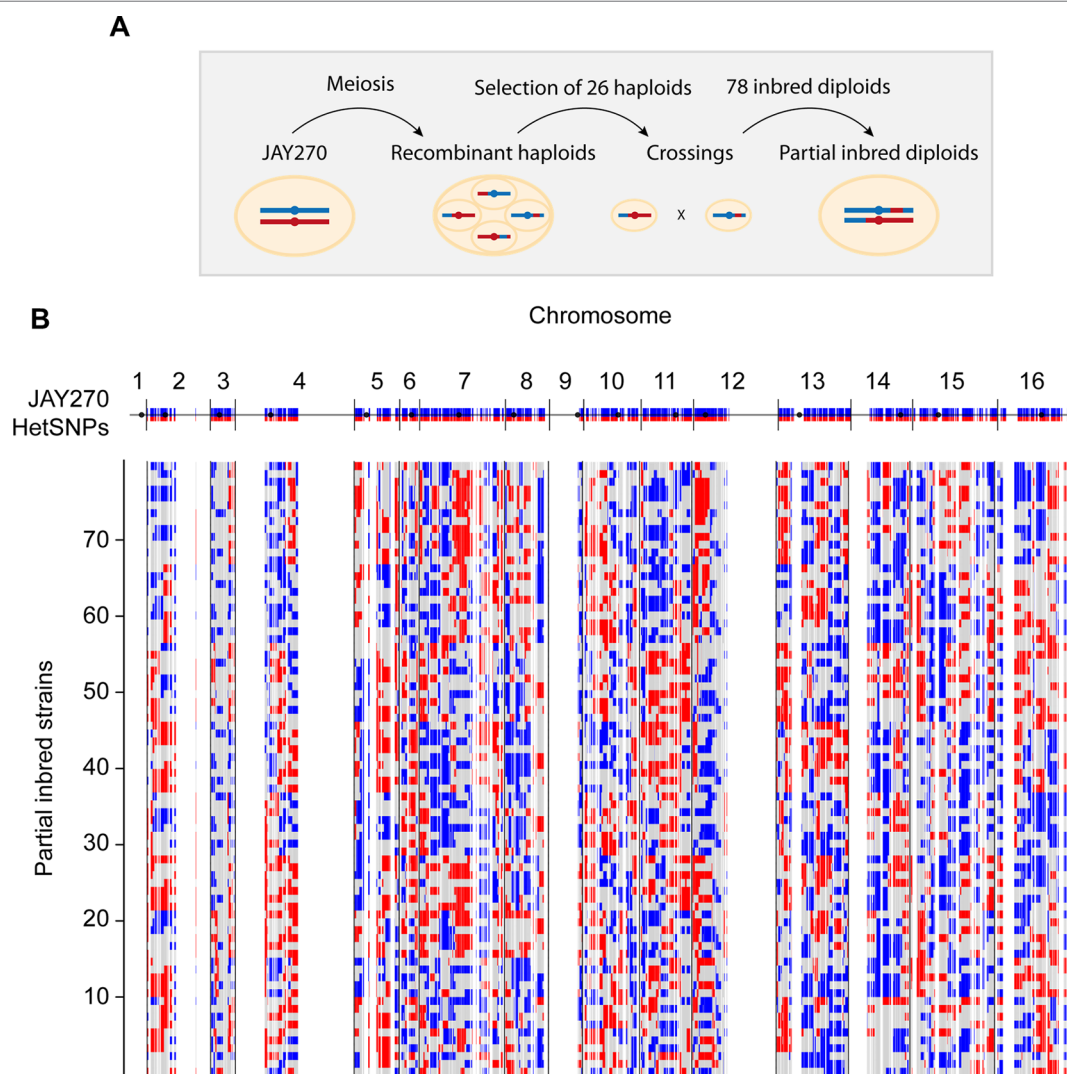
### Controlled Reduction of Heterozygosity in the JAY270 Genome Through Inbreeding

In order to characterize the phenotypic contributions of the heteroalleles present in the JAY270 genome, we explored how changes in the abundance and distribution of heterozygous sites would affect the traits of the strain. Recently, we reported a draft phased map of ~12,000 heterozygous single nucleotide polymorphisms (HetSNPs) unevenly distributed across JAY270's genome (Figure S1 and Rodrigues-Prause et al., 2018). In order to keep track of the two specific allele variants present at each HetSNP, we arbitrarily named the two phased haplotypes for each chromosome pair as M or P, making an analogy to haplotypes of maternal or paternal origin in a classic F1 cross (M and P alleles are represented in red and blue in all figures, respectively).

Our primary strategy to create JAY270 derivatives containing reduced heterozygosity was based on inbreeding. Our group had previously isolated and whole-genome sequenced 52 haploid spore clones originated from thirteen sets of JAY270 four-spored

tetrads (Rodrigues-Prause et al., 2018). It has been estimated that each meiotic cell division in *S. cerevisiae* produces about 90 crossovers distributed across the genome (Mancera et al., 2008; Chakraborty et al., 2018). These events result in the formation of recombinant chromatids that are sorted into haploid spores, each containing approximately half maternal and half paternal alleles (Figure 1A). In order to maximize the genotypic variation of the haploids used in our crossings, we selected for mating only one *MATa* and one *MATa* spore from each of the thirteen sequenced tetrad sets. This ensured that all inbred diploids were formed by joining recombinant haplotypes generated from independent

meiotic crossover events. An additional criterion for selection of the parent spores was based on their genotype at the *ACE2* locus. We recently showed that JAY270 is heterozygous for a frameshift mutation at *ACE2* (*ace2-A7*) and diploid derivatives homozygous for the mutant allele display a cell-cell aggregation phenotype that could confound the phenotypic analysis of inbred diploids (Rodrigues-Prause et al., 2018). Thus, 13 *MATa* *ACE2* and 13 *MATa* *ace2-A7* spores were crossed in inter-tetrad pairwise combinations (Table S1 and Figure S2A), resulting in a collection of 78 inbred diploid strains directly derived from JAY270. This inbred collection enabled us to examine the effects



**FIGURE 1 |** Schematic representation of the construction of inbred diploids. **(A)** A parental diploid JAY270 cell is shown with a pair of homologous chromosomes colored in blue (paternal haplotype or P) or red (maternal haplotype or M). Cells were induced to undergo meiosis, resulting in the formation of tetrads containing four recombinant haploid spores which had their whole genomes sequenced. One *MATa* *ACE2* spore and one *MATa* *ace2-A7* spore was selected from each of 13 tetrads and used for pairwise crossings (Figure S2A). This setup resulted in a collection of 78 inbred diploid strains derived from inter-tetrad matings. Each resulting inbred diploid was expected to remain heterozygous M/P at ~50% of the genome, and to become homozygous M/M (~25%) or P/P (~25%) for the other half. **(B)** Genome-wide maps of the genotypes of the 78 inbred diploids. The top blue and red line corresponds to the linear representation of the distribution of 12,023 HetSNPs in the JAY270 parent genome (Rodrigues-Prause et al., 2018; Figure S1). Regions without blue/red lines represent genomic regions of homozygosity in JAY270. Each line below shows the maps and the distributions of genotypes in the inbred strains: Heterozygous M/P (gray); homozygous P/P (blue); homozygous M/M (red). The cumulative distribution and the overall quantification of each genotype class are shown in Figure S2B–C.

of homozygosity at most regions of the JAY270 genome. The only exceptions were loci genetically linked to *MAT* and *ACE2*, respectively on chromosomes III and XII (Chr3 and Chr12), which tended to remain heterozygous in the inbred clones due to the parental haploid selection criteria described above.

Because only one generation of inbreeding was carried out, the genome of each inbred strain in the collection was predicted to be approximately a quarter homozygous for maternal alleles, a quarter homozygous for paternal alleles, and half heterozygous. Importantly, since each haploid parent inherited a unique combination of maternal and paternal alleles, no two inbred diploids were heterozygous for the same half HetSNPs. Based on the whole-genome sequence information of all 26 parental haploids, we derived precise genotype maps for each inbred diploid. These maps show all loci that remained heterozygous (M/P), and the loci that became homozygous for either allele (M/M or P/P), and illustrate the genetic variation present in our collection (**Figure 1B** and **File S1**). We analyzed the genotype maps to determine the overall level of hetero- and homozygosity in each of the inbred diploids and overall in the strain set (**Figure S2B** and **S2C**). The average inbred was heterozygous for 51% of the JAY270 HetSNPs, within a range of ~40 to ~62% for the least and most heterozygous inbreds. The average of M/M and P/P homozygosity was well balanced (~26 and ~23%, respectively) and consistent with the levels expected for a single generation of inbreeding.

## Characterization of Phenotypic Variation in the Inbred Diploid Collection

We next explored how homozygosity in each inbred diploid affected different traits in comparison to their fully heterozygous parent (JAY270). A Petri plate spotting assay format was used as an initial screen for growth phenotypes under a variety of individual stress conditions (detailed information in **Table S3**), some of which are known to be present in the sugarcane fermentation industrial environment (Della-Bianca et al., 2013). No significant changes in cell viability or growth characteristics were observed when cells were plated and grown in the presence of 7 or 11% *v/v* ethanol, 30 mM furfural (a byproduct of lignocellulose biomass fermentation), 0.75 mM of menadione (an inducer of oxidative stress), or 100 and 150 J/m<sup>2</sup> ultraviolet light exposure and 0.01% methyl methanesulfonate (DNA damaging agents). The abilities to metabolize galactose and the non-fermentable carbon sources ethanol and glycerol were also apparently uniform across all inbred strains. Mild phenotypic variation was observed when cells were grown on raffinose as the sole carbon source, or in the presence of 100 mM of hydroxyurea, an inducer of DNA replication stress (data not shown).

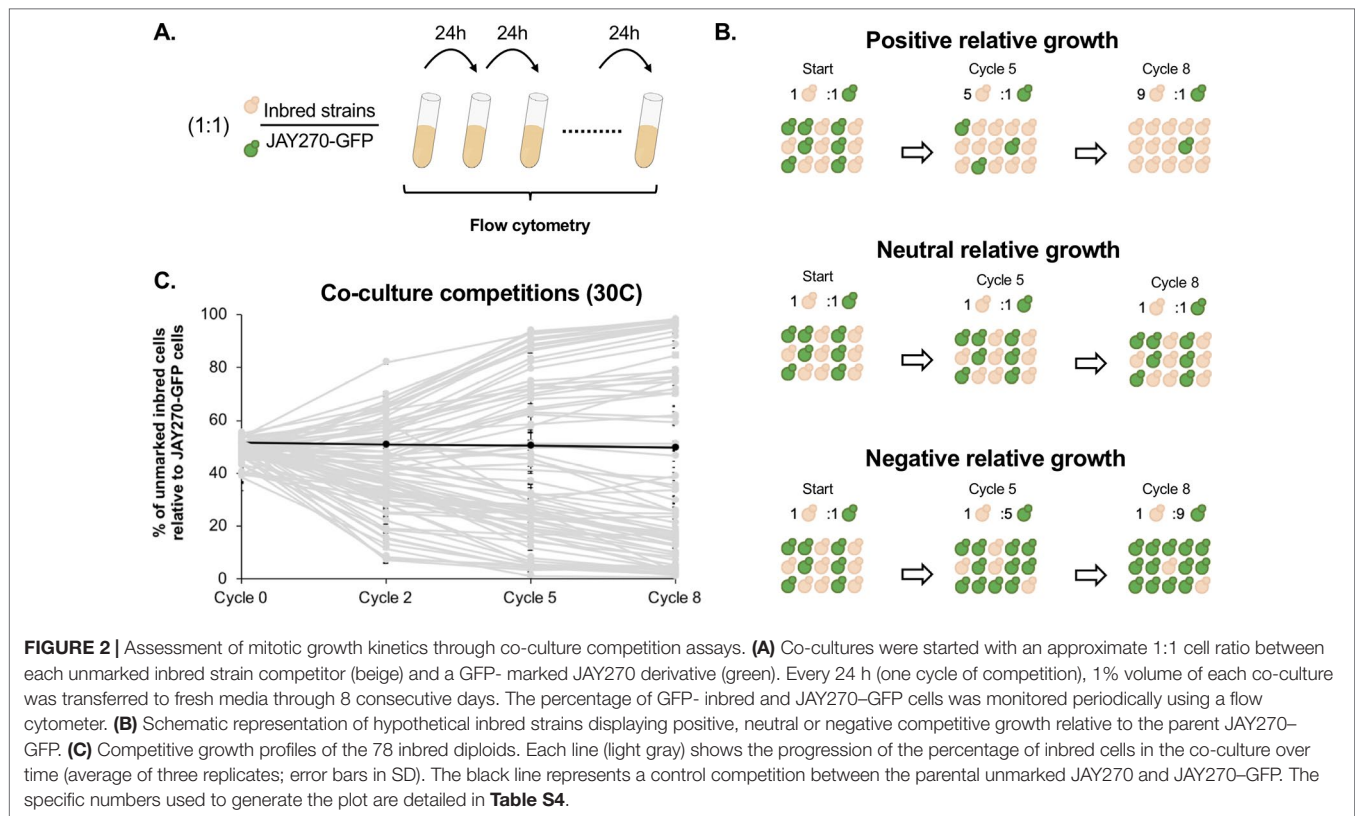
Finally, a pronounced variation in tolerance to heat stress (growth at 39°C) was observed among strains in the inbred collection (**Figure S3**). The wide range in the distribution of this phenotype during the screening phase made it suitable for a subsequent detailed phenotypic characterization and quantitative trait loci (QTL) analysis. We categorized the inbred strains into six phenotypic groups using a qualitative colony-size scoring system (**Figure S4**). JAY270 displayed an intermediate heat tolerance phenotype (score 3), characterized by good viability, but substantial variation in colony diameter, ranging from small

to medium sized colonies. Roughly 40% of the inbred strains (32 of 78) displayed a similar phenotype. The remaining inbred strains displayed heat tolerance patterns that were either lower or higher than JAY270. At the extremes were strains that either showed no growth at all or formed very small colonies when incubated at 39°C (scores 0 and 1, respectively), while others that formed uniformly large colonies and were classified as the most heat tolerant strains (score 5). The median heat stress tolerance scores for each inbred and the phenotypic distribution in the full strain set are shown in **Figure 3A**.

In addition to the phenotypes examined through the plating assays above, we also investigated a more subtle variation in mitotic growth kinetics. JAY270 is known to grow very robustly, and this trait is likely a key factor contributing to its ability to outcompete wild yeast contaminants in the sugarcane fermentation process. Thus, we sought to explore the variation in the growth kinetics phenotype among the inbred strains through a cumulative co-culture competition assay (**Figure 2A**). Each inbred strain was co-cultured with a GFP-marked JAY270 derivative (JAY270-GFP) under optimal *S. cerevisiae* growth conditions (YPD liquid rich medium at 30°C under rotation). The co-cultures were started with an approximately equal inoculum of the two competitors (~2.5 × 10<sup>6</sup> cells each) and were incubated for 24 h, past nutrient depletion and population saturation (~15 h). At the end of each daily growth cycle, 1% volume of each co-culture was transferred to fresh liquid medium to allow continued growth. The percentage of GFP-negative (inbreds) and GFP-positive (fully heterozygous JAY270) present in the co-cultures was measured periodically with a high-throughput flow cytometer (**Figure S5A–C**), and used as a parameter to quantify the growth kinetics of each of the inbred diploids relative to JAY270. Inbred strains with intact growth kinetics should maintain steady ~50% over time, whereas deviations up or down would indicate a phenotypic change (**Figure 2B**).

Besides the genotype of the inbreds, another factor that may cause the relative abundance of the GFP– and GFP+ competitors to deviate is the emergence of beneficial *de novo* mutations within the co-cultures. However, this effect should be delayed until the newly formed mutants become numerous enough to be detected. In order to determine the period of time during which the GFP– to GFP+ ratio can be confidently attributed solely to the initial genotype of the inbreds, we performed control co-culture competitions of each of four independently generated GFP– marked JAY270 clones versus the unmarked JAY270 parent strain. We carried out a total of twelve co-cultures (four GFP– marked clones, three replicates each) with daily 1% volume transfer cycles to fresh media for 22 consecutive days, and the percentage of GFP– was measured at 7-day intervals (**Figure S5D**). The GFP ratios in all 12 independent co-cultures remained steady at ~50%:~50% by the end of the first week (cycle 8). By the end of the second and third weeks, some of the ratios had diverged up or down, presumably through emergence of beneficial mutations in the GFP– or GFP+ strains. Therefore, we limited our experimental competitions of the inbred diploids versus JAY270–GFP to a maximum of 8 daily transfer cycles, in order to insulate the measured GFP ratios from the effect of *de novo* mutations. These control experiments also showed that integration of the GFP cassette into the JAY270 genome did not by itself have an





effect on growth kinetics. Additional JAY270–GFP versus JAY270 control co-culture competitions were included every time a new experimental evaluation of the inbred collection was performed (39 replicates), and in no cases a significant deviation in the GFP ratio was observed before or at transfer cycle 8.

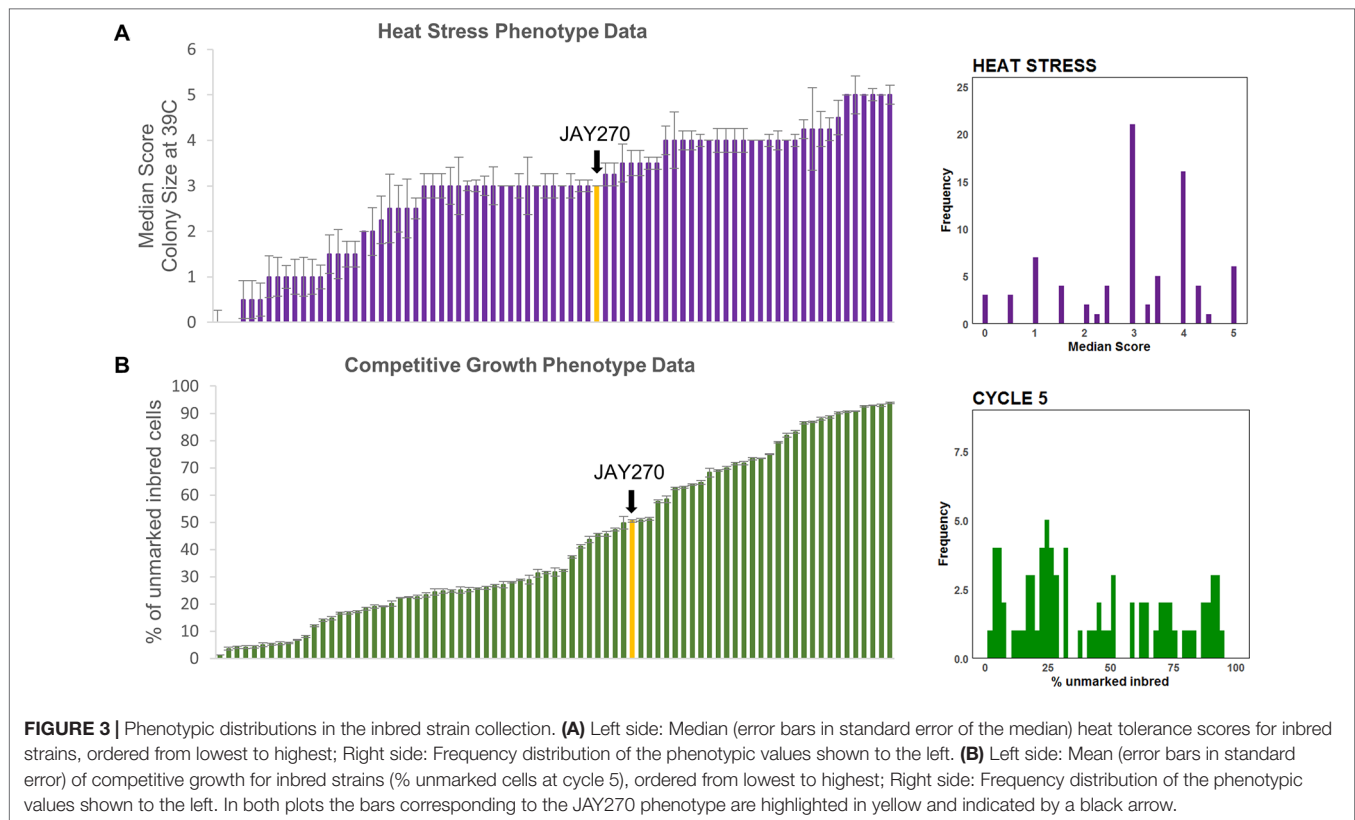
The competitive growth profiles of the inbreds were characterized collectively by a “fan out” shape, showing that a wide range of phenotypic variation existed in the strain set (**Figure 2C**). Many of the inbred diploids displayed growth kinetics that were substantially different from JAY270, not only slower but also faster. Of this group at the extremes of the competitiveness range, 18 displayed a strong reduction in growth kinetics and comprised less than 10% of the total cell population by the last cycle of co-culture; while 13 inbreds showed a substantial improvement in growth, outcompeting JAY270 to reach more than 90% of the cells in the co-cultures. Importantly, all but one of the inbreds, regardless of the neutral, positive or negative relative growth kinetics profiles, followed a steady unidirectional trajectory from the early cycles until the end of co-culture. This result is consistent with their phenotype being a function of their initial genotype, and not due to the random appearance of *de novo* mutations during the experiments. In addition, there was very little variation between the independent replicates of each inbred co-culture, further disfavoring a potential influence of *de novo* mutations over the observed phenotypes.

It is important to note that all 78 inbreds, even those with the poorest performance in the co-culture competition, grew apparently normally and indistinguishably from JAY270 on solid/

agar rich medium at 30°C. This shows that the cumulative liquid co-culture competition assay was able to reliably and consistently uncover extremely subtle relative differences in growth kinetics. We estimate that the most extreme competition phenotypes among the inbreds, reaching <10% or >90% of the total co-culture cell population by transfer cycle 8, should have a rate of cell division only ~3% longer or shorter than JAY270, respectively. Thus, the co-culture competition assay offered an opportunity to reliably measure minor phenotypic changes that resulted from the different genotype combinations represented in the inbred collection. Even though we collected data for cycles 0, 2, 5, and 8, we used data only from cycle 5 for the downstream QTL analysis as it offered an optimal quantification of relative growth kinetics. The cumulative nature of this assay meant that by cycle 8 some of the extreme GFP ratios had already started to reach a plateau, which could lead to an underestimation of their full phenotypic differential. This cumulative trend can be visualized in **Figure S6** as the progressive shift in the phenotypic distribution away from center (cycles 0 and 2) and toward the low and high extremes over time (cycle 8). The specific mean percentage of each inbred in the co-culture at cycle 5 and the phenotypic distribution in the full strain set are shown in **Figure 3B**.

## Identification of Genomic Regions Associated With Phenotypic Variation

We next performed a QTL analysis to identify possible relationships between the specific genotypes at JAY270 HetSNPs and the

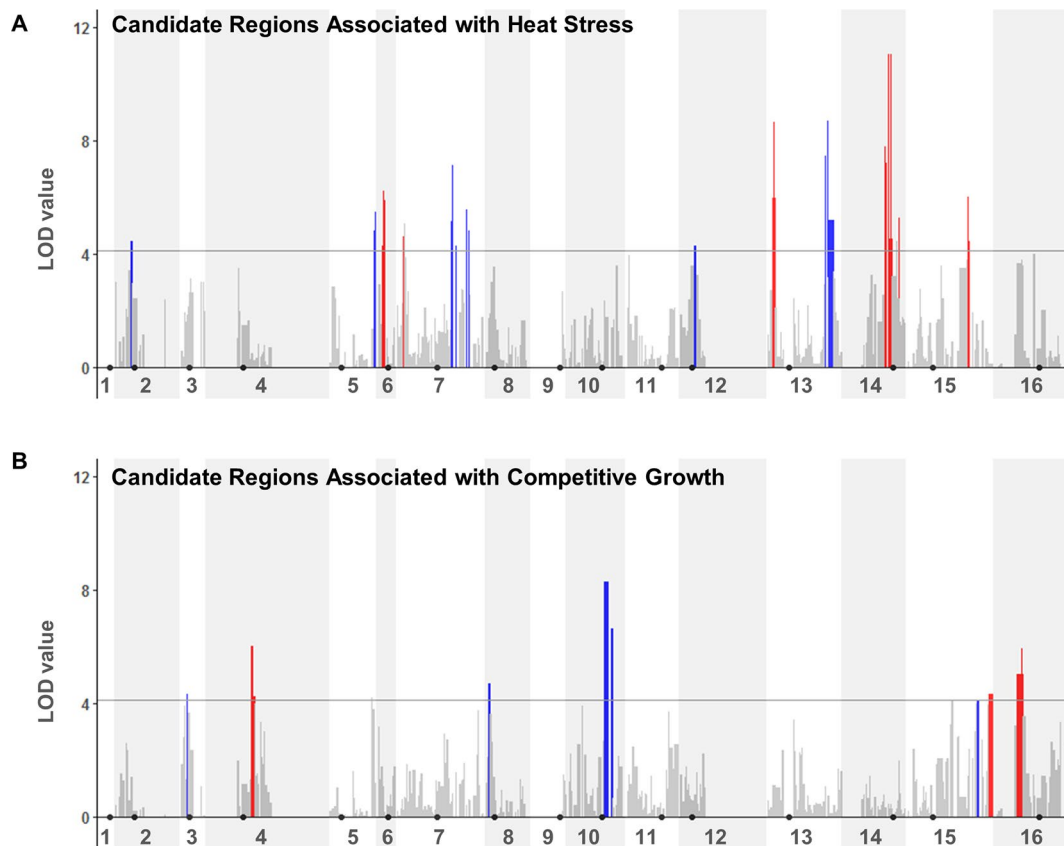


phenotypic variation in heat stress tolerance and competitive growth among inbred strains. Because all the inbred diploids in the collection were necessarily heterozygous at Chr3 *MAT* (*MATa/MATα*) and at Chr12 *ACE2* (*ACE2/ace2-A7*), we excluded markers genetically linked to those loci from the analyses (within ~50 and ~75 Kb up and downstream of each, respectively). This resulted in a final list of 11,742 HetSNPs that were included in the QTL analyses. We used the genotype maps of all inbreds (**File S1**) to determine the frequencies of homozygous M/M and P/P, and heterozygous M/P genotypes at each HetSNP marker. Then, for each marker we calculated the mean phenotype value measured among strains with M/M, P/P and M/P genotypes. Using a one-dimensional scan of the genome, log<sub>10</sub> likelihood ratio (LOD) scores were determined for each marker for each trait. The statistical significance thresholds for the identification of candidate loci associated to each trait were established by randomized phenotype by genotype permutation tests (five independent runs of 10,000 iterations for each trait) at the  $p < 0.05$  significance level. The significance threshold values determined independently from the heat tolerance and competitive growth permutation tests were the same: LOD > 4.11. The genome-wide LOD scores for each trait are plotted in **Figure 4**, and regions that rose above the 4.11 thresholds were considered to be statistically significant. A two-dimensional scan of the genome was also performed, but no significant pairwise epistatic interactions were detected (data not shown).

Although our inbred population size was relatively small, this analysis was sufficient to reveal multiple genomic segments that

may make important contributions to the traits of interest. In total, thirteen regions from eight chromosomes showed association to heat tolerance, and seven regions from six chromosomes to competitive growth (**Table S6**). For each trait, these regions corresponded to combined total sizes of ~332 Kb with ~189 annotated genes in them, ~25% of which were heterozygous for non-synonymous substitutions. This narrower list included 47 and 48 candidate genes within the genomic regions that were significantly associated to heat tolerance and competitive growth, respectively. The genes in both lists belonged to diverse functional annotation groups (i.e., no specific Gene Ontology terms were significantly enriched at  $p < 0.01$ ).

We evaluated which quantitative inheritance model better fit the observations from each region (**Table S7**). Most regions (16 of 20) were consistent with an additive variance model in which the heterozygote has an intermediate phenotype. We also found four regions with likely dominance, but no cases of overdominance. Finally, we estimated the percent variance explained (PVE; **Table S7**) for the HetSNP with the highest LOD value within each region using a single-QTL model analysis. In order to facilitate a comparison of the relative contributions between regions to each trait, we also calculated relative PVE values normalized to the locus with the highest PVE. As an additional approach, we fit a multi-QTL model (**Table S8**) and determined that three of the regions identified for heat tolerance synergistically explained 76% of the variance and three of the regions for competitive growth worked together to explain 57% of the variance.



**FIGURE 4 |** QTL analyses to identify loci potentially underlying heat stress tolerance **(A)** and relative growth kinetics **(B)**. LOD scores for each HetSNP (y-axis) were plotted over linear genomic coordinates of the HetSNPs for all 16 chromosomes (x-axis). The horizontal gray line corresponds to the LOD = 4.11 threshold of significant association. The color code in the plots indicates the allele contributing the higher phenotype value (red: maternal; blue: paternal). Black circles indicate the positions of centromeres. Plots were generated in R (version 3.4.0) using the ggplot2 package (version 1.42-8) and a custom script.

The identification and characterization of specific major genes and alleles that contribute to these traits in JAY270 was beyond the scope of this particular study. However, we noted that none of the significant association regions overlapped between the two traits. In addition, there was no overlap between the inbred strains ranked in the upper or lower tiers of heat tolerance and competitive growth (**Figure S7**). This suggested that the two traits are controlled independently of each other, so different combinations of alleles present at different sets of JAY270 genomic regions contributed in their own way to the phenotypic variation observed for each trait.

### Controlled Reduction of Heterozygosity in the JAY270 Genome Through Targeted Uniparental Disomy

In the inbred collection approach described above, each strain had lost roughly half of the heterozygosity present in JAY270, thus a large fraction of the genome was affected. We next took an independent and more conservative approach in which fewer heterozygous loci were manipulated at a time. To do so, we adapted a procedure to induce targeted uniparental disomy

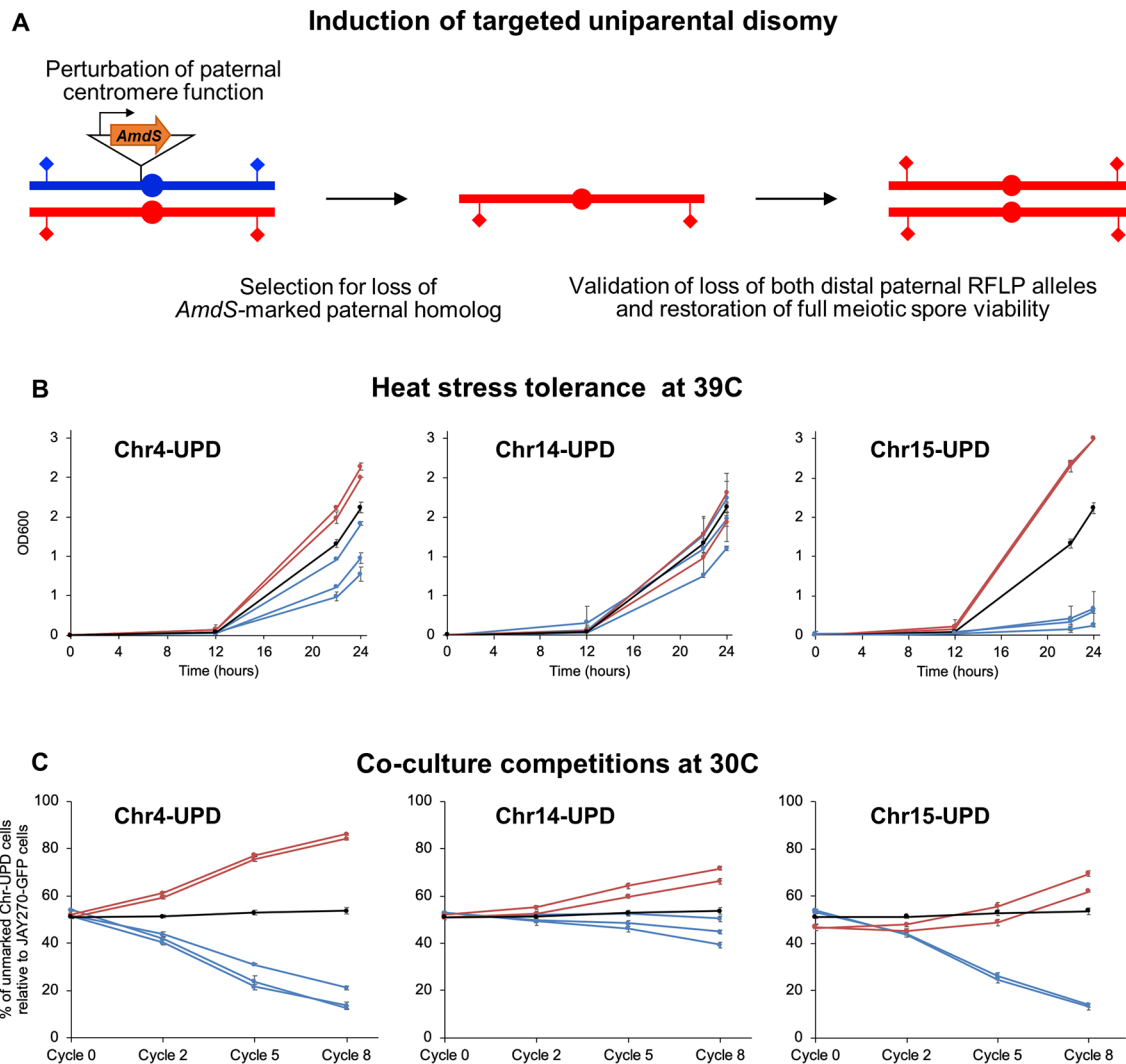
(UPD) (i.e., homozygosis for an individual whole chromosome), while preserving heterozygosity in the other chromosome pairs. Our strategy took advantage of previous demonstrations that driving transcription through centromeric regions leads to perturbation of the function of centromeres, and can be used to induce targeted chromosome loss, resulting in  $2n - 1$  monosomic diploid cells (Hill and Bloom, 1987). This strategy was successfully applied to map mutations to individual chromosomes in a *ura3/ura3* auxotrophic diploid laboratory strain background (W303; (Reid et al., 2008)), by inducing transcription of a *pGAL1-URA3* cassette integrated at centromeric regions, and then applying counter selection for 5-FOA resistance to recover clones that had lost the targeted chromosomes.

Here, we adapted this approach for use in prototrophic diploid strains by integrating a hemizygous copy of the heterologous forward and counter selectable marker *AmdS* (Solis-Escalante et al., 2013) immediately upstream of specific JAY270 centromeric regions. We modified the *AmdS* cassette by removing the transcriptional terminator sequence, thus enabling constitutive transcription to continue past the ORF and extend through the centromeric sequence. Insertions of *AmdS* cassettes adjacent to centromeres of each the M and P homologs

of targeted chromosomes were obtained and stably maintained through forward selection for growth in media containing acetamide as the sole nitrogen source (**Figure 5A**). Then, counter selection for loss of the cassette (fluoroacetamide resistance) was used to isolate candidate clones carrying chromosome loss. The final phase, and a key part of the strategy, relied on the observation that monosomic diploid *S. cerevisiae* cells tend to rapidly and spontaneously endoduplicate the remaining homolog,

which results in reestablishment of the normal chromosomal complement through UPD (Reid et al., 2008). Another possible mechanism is that UPD may be formed in a single step through meiosis I-like co-segregation of sister chromatids in mitotic cells (Andersen and Petes, 2012).

We conducted a proof-of-concept experiment focused on the generation of strain pairs carrying bidirectional UPD for three chromosomes (Chr4, Chr14, and Chr15), chosen on the basis



**FIGURE 5 |** Construction and phenotypes of UPD strain pairs. **(A)** A cassette containing the counter-selectable marker *AmdS* under the transcriptional control of the *TEF1* constitutive promoter and lacking a terminator sequence was integrated immediately upstream of the centromeric regions of each M or P homolog of Chr4, Chr14, and Chr15 (insertion of *AmdS* at a P homolog is shown in this case). Transcription of *AmdS* perturbs centromere function and induces targeted chromosome mis-segregation during mitosis. Cells that lost the *AmdS* marker were selected for in media containing fluoroacetamide. Spontaneous endoduplication of the remaining homolog results in strains containing UPD, in this case represented as the maternal (red) homolog. Loss of each homolog was validated by RFLP-PCR genotyping analysis at distal markers on both chromosome arms (diamonds), and confirmation of endoduplication was obtained by tetrad dissection and spore viability analysis (4 viable spores per tetrad indicate disomy). Multiple independently-generated UPD strains were isolated and used in phenotypic tests. **(B)** Growth profiles under heat stress of UPD strain pairs. Each line shows the growth curves of an individual UPD strain under high temperature conditions (39°C) in liquid culture; error bars in SD. The OD600 at 0, 12, 22, and 24 h are shown on the y-axis. Red and blue lines represent strains containing two copies of the maternal (M/M) or paternal (P/P) homologs, respectively, of each chromosome analyzed (Chr4, Chr14, and Chr15). The black line represents the JAY270 control. **(C)** Competitive growth profiles of UPD strain pairs. Each line shows the relative growth profile of an individual UPD strain in co-culture with JAY270-GFP, grown at 30°C; error bars in SD. The percentage of UPD cells relative to JAY270-GFP cells at the 0, 2, 5 and 8<sup>th</sup> cycle of co-culture are shown in the y-axis. Color scheme is the same as in B. The black line corresponds to a control competition between the parental unmarked JAY270 and JAY270-GFP.



of their overall chromosome size, and number and distribution of HetSNPs (**Figure S1**). It has been proposed that loss of long *S. cerevisiae* chromosomes may impose a heavier phenotypic burden than loss of a small chromosome (Reid et al., 2008), thus the likelihood to recover UPD through endoduplication should be higher. Chr4 was an attractive candidate for this analysis because it is a large chromosome but has a relatively low number of HetSNPs, which are all clustered in a central region. We also chose to study Chr15, because it is also a large chromosome, but in contrast to Chr4 it has a large number of HetSNPs (> 1,400; ~12% of the genome's total) scattered throughout its whole length. Finally, a third interesting case study was Chr14, which is a mid-size chromosome containing a relatively large number of HetSNPs (~700) and a long homozygous segment.

We integrated the terminator-less *AmdS* cassette immediately adjacent to the centromeres of each homolog of these three chromosomes. We then screened multiple independently-generated chromosome loss clones by PCR-RFLP genotyping followed by tetrad analysis to identify those that had undergone UPD to become homozygous for each of the three respective targeted chromosomes (**Figure S8**). We tested by tetrad analysis a subset of the fluoroacetamide resistant clones that had LOH at both the left and right centromere-distal HetSNP markers. All of the clones tested through tetrad analysis were found to be disomic (*i.e.*, four viable meiotic spores per tetrad; data not shown), indicating that monosomy of Chr4, Chr14, and Chr15 was short-lived, and UPD was readily acquired. As an additional control, we determined the genotypes of all UPD strains at three heterozygous loci that were not targeted for UPD (**Figure S9**). All strains remained heterozygous at those loci, except for the specific chromosome homolog targeted for UPD.

## Phenotypic Consequences of Chromosome-Scale LOH

Our goal was to test whether localized reduction of heterozygosity in these three chromosomes would be sufficient to cause detectable variations in the heat tolerance and competitive growth phenotypes. It is important to note that, if UPD induction indeed led to phenotypic changes, we would not necessarily expect those to correlate with the presence of significant peaks identified through the QTL analyses derived from phenotypes of inbred strains. The association of the HetSNPs present in those recombinant chromosomes likely reflects complex interactions with the homo- and heterozygous genotypes of loci present in other regions of the genome. In contrast, any phenotypes detectable in the UPD strains would likely be dependent on the collective and coordinated homozygosity of all alleles present in each M or P haplotypes for the respective chromosome homologs, within a background of heterozygosity everywhere else.

Earlier in the study we used a colony size qualitative scoring system to describe the variation in the heat tolerance phenotype among the strains in the inbred collection. In order to improve the characterization of more subtle phenotypic differences among the UPD strains, we monitored optical density (OD600) in pure cultures in liquid media under rotation at 39°C. We validated this approach by generating 39°C pure culture liquid

growth curves for JAY270, and for two heat tolerant and two heat sensitive inbreds, and compared them to the results of parallel solid media qualitative colony scoring assays of these same control strains. The heat tolerance profiles for each of the strains were quite consistent between the two assays (**Figure S10**). The liquid growth assay is more laborious and thus less suitable for the analysis of large strain sets (*i.e.*, the whole inbred collection). However, it provided more informative data because its broader dynamic range allowed us to better monitor refined gradations of the heat stress tolerance phenotype.

Using this enhanced method (**Figure 5B**), we found that UPD for the two Chr15 haplotypes influenced heat tolerance significantly and in opposite directions. Chr15-UPD M/M strains were more heat tolerant than JAY270, while Chr15-UPD P/P strains were quite sensitive. A similar, but less pronounced pattern was observed for Chr4-UPD M/M and Chr4-UPD P/P, which showed, respectively, slightly higher and lower tolerance to heat stress compared to JAY270. Finally, Chr14-UPD in either direction did not appear to cause substantial difference in heat tolerance.

Co-culture growth competition assays also revealed pronounced phenotypic shifts in the UPD strains (**Figure 5C**). Homozygosis for the two Chr4 haplotypes influenced the growth kinetics significantly and in opposite directions, in a symmetric fashion. Chr4-UPD M/M strains outgrew the fully heterozygous parent JAY270, whereas Chr4-UPD P/P strains displayed the opposite phenotype. Homozygosis for the two Chr15 haplotypes resulted in a less symmetrical change in growth kinetics, but still followed a similar trend in which each haplotype displayed opposite competition profiles. Chr15-UPD M/M strains displayed a subtle but steady growth advantage, while Chr15-UPD P/P strains were outcompeted by the parent strain JAY270 at a faster pace. Finally, the changes in growth competition profiles of the Chr14-UPD strains were more subtle, but also showed a divergent trend between haplotypes.

Notably, all independently-generated strains within each of the six UPD sets displayed very similar phenotypes for both heat tolerance and competitive growth. This indicated that any non-targeted alterations that may have arisen in their genomes during construction, if at all present, were not sufficient to influence these phenotypes. Taken together, our results showed that even though each of the three UPD pairs retained ~88–96% of the overall HetSNPs of JAY270, their relatively small and localized erosions of heterozygosity were sufficient to create significant and often symmetric alterations in the two phenotypes examined.

## DISCUSSION

The work presented above showed that the heterozygous genome of JAY270 harbors a diversity of alleles that can support a wide phenotypic variation for competitive growth and heat stress tolerance. Our QTL analyses using inbred diploids pointed to broad regions scattered throughout the genome that were associated with these two traits. The genomic regions identified in each case were distinct between them (**Figure 4** and **Table S6**), and the groups of inbred diploids ranked at the top and bottom

of each phenotypic range were non-overlapping (**Figure S7**). Interestingly, we found that the inbred clones displayed superior just as often as inferior performances compared to the heterozygous JAY270 parent strain for the two narrowly defined traits analyzed.

The phenotypic analysis of strains engineered for carrying bidirectional UPD of targeted chromosomes also provided important clues about the extent to which heterozygosity influences the overall phenotypes of JAY270. The cumulative effects of homozygosity in entire chromosomes resulted in detectable changes in both heat stress tolerance and competitive growth (**Figure 5**). Homozygosity for each haplotype within a chromosome often led to opposite phenotypic outcomes, characterized mostly by a symmetric response relative to the heterozygous parent. This pattern was especially noticeable in the Chr4-UPD strains in growth kinetics and the Chr15-UPD strains in heat tolerance. In these two cases, the M/M UPD derivatives were superior to JAY270 while the P/P derivatives were inferior. However, it is entirely possible, and we believe likely, that other phenotypes not specifically tested in this study could have diverged in opposite directions such that P/P UPDs would be superior to JAY270 and M/M UPDs.

The cryptic phenotypic variation uncovered in the inbred and UPD strain sets engineered for this study may also be accessible naturally to the JAY270 lineage during industrial fermentations. This may be achieved quickly and globally through meiotic recombination followed by inbreeding, and/or gradually and locally through mitotic recombination (Magwene, 2014; Dutta et al., 2017). Assuming that the ancestor of the lineage that gave rise to JAY270 was a hybrid diploid formed by mating of two diverged haploids, it would have been heterozygous for loci distributed across 100% of its genome. One cycle of meiosis in that ancestor followed by mating between sibling spores from the same tetrad would produce a diploid bearing only ~66% of the heterozygosity originally present in the ancestor (Johnson et al., 2004), or ~50% heterozygosity if the mating occurred between spores from two different tetrads (**Figure S2**). A single such cycle of meiosis is sufficient to explain most or all of the distribution of heterozygosity observed in the JAY270 genome (~60%; **Figure S1**). Two or more meiotic cycles over the life history of this lineage would result in substantially less heterozygous diploids (~44% or less). The second and more gradual path to the erosion of the heterozygosity originally present in the hybrid ancestor would be multiple rounds of allelic inter-homolog mitotic recombination over successive generations of vegetative growth. This would lead to the progressive accumulation of many tracts of homozygosity distributed genome-wide. Either mechanism, or a combination of them, may have contributed to shaping the JAY270 genome. However, we favor a predominant role for mitotic recombination, given the long record of propagation of the PE-2 industrial strain (Basso et al., 2008), from which JAY270 was purified (Argueso et al., 2009).

In many species inbreeding is known to correlate with inferior phenotypes and decreased fitness (Charlesworth and Willis, 2009). In contrast, in this study we described multiple cases in which homozygosity resulted in superior performance. However, it is essential to note that neither of the individual and narrow phenotypic assays we used reproduced, nor approached, the

complex and dynamic sugarcane fermentation environment that likely shaped the present genomic configuration of JAY270. If it had been possible for us to exactly reproduce the biotic and abiotic challenges found in sugarcane bioethanol distilleries, then we predict that most inbred and UPD clones would perform poorly relative to JAY270, and few or none would be superior. The challenges posed by such varied and simultaneous stress conditions might be better met by the heterozygous genomic configuration that enables JAY270 to be a well-rounded generalist, the feature that makes it so attractive to bioethanol producers.

We interpret our results within the context of a model in which the erosion of heterozygosity in the JAY270 genome through meiotic and/or mitotic recombination, while frequent and potentially beneficial in specific circumstances, may be generally disfavored and curtailed by natural selection. For example, homozygosity at a specific chromosomal region might lead to faster growth, but it may also decrease tolerance to elevated temperatures or other unrelated stress sensitivities. Once cells carrying such new LOH tracts arise, selective pressures to which the cells are subjected to in the distillery environment would determine their fate of expansion or disappearance from the yeast population. The adaptive potential of LOH has been nicely characterized in inter- and intra-species yeast hybrids grown in chemostats over several generations under different and specific growth conditions (Smukowski Heil et al., 2017). Mitotic recombination leading to LOH was shown to be a major driver of adaptation in those hybrids. However, when clones carrying an LOH event that conferred superior fitness in a specific growth condition were tested in an alternate condition, their fitness was often reduced. This is consistent with our findings that a genome-wide or even regional reduction in heterozygosity through inbreeding or UPD can have positive effects for some specific traits, but may not necessarily support overall fitness.

We speculate that the net result of such LOH events in JAY270 would often be disadvantageous in its natural environment where optimal performance is constantly demanded for all phenotypes. This opposing interaction between LOH steadily introduced by recombination versus selection for optimal adaptation to a complex and dynamic environment could explain the persistence of heterozygous regions in the JAY270 genome. Given that LOH events occur at a substantial rate in yeast genomes and that JAY270 had been clonally propagated at industrial scale for years prior to isolation (and perhaps longer in natural environments), we reason that most of its genomic heterozygosity could have already been eroded away. However, the fact that a substantial portion (~60%) of the JAY270 genome still retains heterozygosity suggests that an opposing force (*i.e.* natural selection) might have acted to disfavor cells carrying LOH spanning loci and heteroalleles that contribute to the strain's overall fitness.

Here, we leveraged inbreeding and induction of targeted UPD to characterize the phenotypic consequences of controlled reductions in the levels of genomic heterozygosity in a natural hybrid yeast strain. Our results using both approaches revealed a wide phenotypic variability provided by the HetSNPs distributed through the JAY270 genome, and allowed the initial identification of broad genomic regions associated with two important industrial traits. This study

laid the foundation for future experimental work aimed at refining these maps to identify specific heteroalleles. A higher resolution of genetic mapping can likely be achieved by substantially increasing the size of the inbred collection, by generating UPD strain pairs for the remaining thirteen chromosomes, and also by targeting mitotic recombination and LOH to specific chromosomal regions of interest (Sadhu et al., 2016). Coupling these expanded strain sets with automated high throughput phenotyping for a wider variety of growth conditions might allow identification of individual heteroalleles and loci associated with traits of interest for industrial applications. In cases where superior performance through homozygosity can be narrowed down to specific heteroalleles, it may be possible to use targeted allele engineering approaches for strain improvement, minimizing or eliminating the negative effects of losing heterozygosity in the neighboring loci. In addition, it would be interesting to characterize these strain sets under conditions designed to more closely recapitulate the bioethanol fermentation environment (Raghavendran et al., 2017), as well as in actual industrial scale sugarcane distilleries. Those experiments would be extremely valuable to determine whether any of the inbred or UPD strains are capable of outperforming JAY270 when challenged by multiple simultaneous stressors. Finally, our work demonstrated the use of the heterologous *AmdS* counter-selectable marker for induction of targeted chromosome loss, which allowed us to bypass the need to introduce auxotrophic markers that often influence yeast phenotypic analyses (Swinnen et al., 2015), while also broadening the UPD approach applicability to wild and industrial strains.

## MATERIAL AND METHODS

### Growth Media

Yeast cells were grown in YPD (20 g/L glucose, 20 g/L peptone, 10 g/L yeast extract, 20 g/L bacteriological agar for solid media), unless otherwise noted. Transformants carrying the *GFP-kanMX* cassette were selected in YPD plates supplemented with 400 mg/L of geneticin. Selection of *AmdS* positive (*AmdS*+) clones was performed in acetamide media (20 g/L glucose, 6.6 g/L potassium sulfate, 1.7 g/L YNB without amino acids, 0.6 g/L acetamide, 20 g/L bacteriological agar). Fluoroacetamide media was used for *AmdS* counter-selection (20 g/L glucose, 5 g/L ammonium sulfate, 1.7 g/L YNB without amino acids, 1.4 g/L complete dropout mix, 2.3 g/L fluoroacetamide, 20 g/L bacteriological agar). Spot assays for phenotypic screening of the inbred collection was performed in different types of media, including: 2% YPGE (20 g/L peptone, 10 g/L yeast extract, 30 mL/L glycerol; 30 mL/L 100% ethanol, 20 g/L bacteriological agar), 2% YP Galactose (20 g/L galactose, 20 g/L peptone, 10 g/L yeast extract, 20 g/L bacteriological agar), 2% YP Raffinose (20 g/L raffinose, 20 g/L peptone, 10 g/L yeast extract, 20 g/L bacteriological agar).

### Yeast Genetic Backgrounds and Microbiology Procedures

All *S. cerevisiae* strains used in this study descended from the JAY270 background (Table S1) (Argueso et al., 2009). Standard

procedures for yeast culture, transformation, crossing and sporulation were followed (Ausubel et al., 2003).

### Construction of a Collection of Inbred Diploids Derived From JAY270

A detailed description of the strategy adopted to construct the inbred strain collection is described in Results section. Genome sequencing data associated with this study is available in the Sequence Read Archive (SRA) database under study number SRP082524 and described previously (Rodrigues-Prause et al., 2018).

### Construction of a GFP-Tagged JAY270 Derivative

A *GFP-KanMX* cassette with homology to a non-coding region located 365 bp upstream to the centromere 5 (*CEN5*, genomic coordinate = 151,522) was built. A *GFP* cassette was amplified from pFA6a-TEF2P-GFP-ADH1-NATMX4, kindly provided by Dr. Maitreya Dunham's laboratory, using the primers JAO1385 and JAO1386. The *KanMX4* cassette was amplified from pFA6-KanMX4 using primers JAO1387 and JAO466 (Wach et al., 1994). Both cassettes were fused by double-joining PCR and transformed into JAY270. Four transformants were selected, purified and tested in 22-cycles of co-culture with the wild type JAY270 strain, one of which (JAY2208) was used for the co-culture competitions against the inbred and UPD strains. See Table S2 for the primers used in this construction.

### Construction of UPD Strains

Destabilization of centromere function was achieved by the insertion of the counter-selectable gene *AmdS* (Solis-Escalante et al., 2013) at the consensus centromeric region of each chromosome analyzed. Cassettes targeting different integration sites (~100 bp or ~5 bp upstream to the targeted centromere), as well as including or excluding a terminator sequence, were tested. Clones showing UPD were more frequently recovered when cassettes that excluded the terminator region of *AmdS* were integrated immediately upstream to the consensus centromere sequence (data not shown). All cassettes were amplified from pCfB2399 (Stovicek et al., 2015) (gift from Irina Borodina; addgene plasmid # 67550) and targeted *CEN4*, *CEN14* and *CEN15*. Transformants were selected and purified in acetamide media. The integration of the cassettes was confirmed by PCR that amplified the left and right junctions between the *AmdS* cassette and the centromeric region. PCR products were designed to span at least one centromere-proximal HetSNP that was used to determine in which homolog the integration occurred using Sanger sequencing. At least two independent clones containing the insertion in each homolog were selected. Cells were grown on YPD plates for 24 h to allow for loss of the *AmdS*-marked homolog, diluted in 200 µl of water and plated in the counter-selection fluoroacetamide media. At least four types of cells should be able to grow in this selective media: (1) cells with inactivation of the *AmdS* gene through point mutation; (2) cells that acquired loss-of-heterozygosity (LOH) tracts spanning *AmdS* as a result of a recombination event, such as gene conversion or mitotic crossover leading to LOH, (3) cells that lost the whole homolog containing *AmdS* and persisted as monosomics; and (4), our



targeted UDP class, cells that lost the whole homolog containing *AmdS*, undergoing a transient monosomic state followed by endoduplication of the remaining homolog. Two sequential tests were performed to screen for targeted UDP clones. First, candidates were genotyped using PCR restriction fragment length polymorphism (PCR-RFLP; **Table S5**) analysis at two centromere-distal positions (HetSNPs near the left and right ends of the chromosome arms, and near the centromeric region). Clones genotyped as homozygous at all three markers could be either monosomic or homozygous disomic for the chromosome of interest. To distinguish between these cell types, candidates were sporulated and tetrads were dissected. Monosomic clones should generate tetrads with two viable and two inviable spores, whereas UDP clones should generate tetrads with four viable spores. Candidates that were homozygous for all three PCR-RFLP markers and produced tetrads with four viable spores were selected for phenotypic tests. See **Table S2** for the primers used in the construction and validation of these UDP strains.

## Phenotypic Assessment of the Inbred Collection

### Phenotypic Screenings Through Plate Spotting Assay

Three cultures of JAY270 and of each inbred strain were grown to saturation at 30°C in 96-well plates containing 200  $\mu$ L of YPD. Cultures were diluted by immersing a 96-pin replicator in the resuspended saturated cultures and subsequently in a 96-well plate containing 100  $\mu$ L of distilled water. Diluted cells were pinned in different types of plates and allowed to grow under different conditions as detailed in **Table S3**.

### Heat Stress Tolerance Assay Using Colony Size Scores

Cells were thawed from  $-80^{\circ}\text{C}$  freezer stocks and incubated at 30°C in YPD plates for 24 h. Cells were inoculated into 5 ml liquid YPD, and incubated for 24 h at 30°C in a rotating drum. Saturated cultures were diluted 10,000-fold and 40  $\mu$ L were plated onto four YPD plates, two of which were incubated for 48 h at 30°C and two were incubated for 96 h at 39°C. Growth at high temperature between strains was assessed through a colony size scoring system (**Figure S4**). Assays were repeated independently at least three times for the whole collection of inbred strains. Median scores between repetitions were calculated for each strain, and standard error of the median was determined by multiplying the standard error of the mean by 1.253 (or square root of  $\pi/2$ ).

### Heat Stress Tolerance Assay in Liquid Cultures

Cells were thawed from  $-80^{\circ}\text{C}$  freezer stocks and incubated at 30°C on YPD plates for 48 h. Cells were inoculated into 5 ml liquid YPD and incubated for 24 h in a cell culture-rotating drum at 30°C. Saturated cultures were diluted 10-fold and 10  $\mu$ L were used to inoculate 5 ml of liquid YPD. Cultures were incubated for 24 h in a rotating drum in a warm room at 39°C. OD600 was determined using a spectrophotometer at 0, 12, 22, and 24 h of incubation. To assure cultures incubated at 39°C started with a comparable number of cells, the saturated cultures used as

inoculum were also plated on YPD (50  $\mu$ L, 10,000-fold dilution) and colony-forming units (CFU/mL) were determined after 48 h incubation at 30°C (data not shown). This assay was repeated independently two times for the collection of UDP strains with three biological replicates per strain each time.

## Flow Cytometry-Based Competitive Growth Fitness Assay

Yeast cells were thawed from  $-80^{\circ}\text{C}$  freezer stocks and incubated at 30°C on YPD plates for 24 h. Cells were inoculated into 5 ml liquid YPD, and grown until saturation for 24 h at 30°C in a rotating drum. Equal volumes of each inbred culture and the JAY270-GFP marked culture were mixed and used to inoculate three assay tubes containing 5 ml of fresh liquid YPD, establishing “Cycle 0” of the competition assay. An aliquot of each mixture was also run through the flow cytometer to determine the starting (pre-culture) percentages of inbred (GFP $-$ ) and JAY270-GFP marked cells. Co-cultures were incubated at 30°C in a rotating drum, and every 24 h (one cycle of competition) 1% of the co-culture volume (50  $\mu$ L) was transferred to 5 mL of fresh YPD medium. Each co-culture competition was performed in triplicate. The percentages of inbred and JAY270 cells was assessed at the beginning of cycle 0 and at the end of cycles 2, 5 and 8 using a Cyan ADP7 color flow cytometer coupled to a HyperCyt Rapid Sampler system for 96-well plate-based assays. Ninety-six well plates for flow cytometry readings were prepared by diluting 10  $\mu$ L of each culture in 190  $\mu$ L of 1% PBS buffer. A PBS-only well was placed after each triplicate and triplicates of a control competition between unmarked JAY270 and JAY270-GFP were included every time a new experiment was initiated. Flow cytometry parameters were optimized by applying a series of gatings that excluded from the analysis cell debris (**Figure S5A**) and cell agglomerates (**Figure S5B**), resulting in a final cell count that was gated into FITC $-$  and FITC $+$  populations based on their fluorescence signals (**Figure S5C**).

## Genome-Wide Association Analyses in the Inbred Collection

### Genotype Calling of Inbred Diploids

A previously described phased map of 12,023 heterozygous SNPs (HetSNPs) in JAY270's genome (**File S1**) (Rodrigues-Prause et al., 2018) was used for calling the genotype of the recombinant haploid strains that originated the collection of inbreds. The phased JAY270 HetSNP haplotypes were arbitrarily designated as maternal (M) or paternal (P).

CLC genomics workbench software was used for mapping sequencing reads from each parental recombinant haploid onto the S288c reference and detecting SNPs across their whole genome. The nucleotides present at each of the 12,023 loci in the JAY270 HetSNP list were determined for each haploid. When no SNPs were detected at those positions, the reference nucleotide genotype was called, while the alternative nucleotide was called when the alternative SNP was detected at a frequency higher than 0.95. After the genotypes were determined they were converted to the respective haplotype designations as M or P.



In order to deduce the diploid genotype of each partial inbred strain, we examined the genotype of their respective parents for each of the 12,023 HetSNPs. Heterozygous M/P loci were called whenever the haploid parents presented distinct nucleotides at a specific position. Whenever the parents presented the same nucleotide at a specific position, the locus was designated either homozygous M/M or homozygous P/P. A genotype map of all haploid parents and inbred strains is provided in **File S1**.

### Statistical Analysis of Genotype/Phenotype Association

Analysis was done using R version 3.4.0, the R/qlt package version 1.42-8, SAS version 9.4, and the SAS GLM procedure (Broman et al., 2003; SAS, 2013; R\_Core\_Team, 2017). Phenotype data from two independent assays (high temperature stress tolerance score and percent inbred cells in co-culture competition at cycle 5) and genotypes at 11,742 HetSNPs across the genome from 78 partial inbred strains were used as inputs for the QTL analysis (Results section, **File S1**, and **Table S4**). Each phenotype was analyzed separately using a one-dimensional scan of the genome using Haley-Knott regression, and log10 likelihood ratio (LOD) scores were determined for each marker position. Five independent 10,000 iteration permutation tests were run to determine the null distribution of our data and the genome-wide LOD threshold value (median of five permutations). Using the 95<sup>th</sup> percentile of the distribution of maximum LOD scores generated from the permutation tests, this resulted in genome-wide LOD thresholds of LOD > 4.11 for heat stress tolerance and LOD > 4.11 for co-culture competition. To determine the allele (M or P) contributing the higher phenotypic value at each locus, we simply noted which genotype showed the highest mean phenotype value. We used two methods to determine the inheritance model of each QTL (additive, dominant/recessive, or overdominance), focusing on the significant regions in the genome and the marker with the highest LOD score for each region. First, each significant region of the genome was visually inspected using effect plots to show the mean phenotype values for each genotype at each locus of interest. Second, Tukey pairwise comparisons were used to determine which genotypes had significantly different mean phenotype values from each other at an alpha level of 0.05 at each locus of interest. Plots showing magnitude of LOD value vs. whole genome position were generated using the ggplot2 package version 2.2.1 using a custom script (Wickham, 2009). Three statistically significant regions were excluded from further analysis and subsequently not color-coded in **Figure 4** due to having no ORFs in the region (Competitive growth: Chr5, Heat Stress: Chr7, Chr14).

A two-dimensional scan of the genome was also performed, but no evidence was found for interactions between QTL (data not shown).

## REFERENCES

Amorim, H. V., Lopes, M. L., de Castro Oliveira, J. V., Buckeridge, M. S., and Goldman, G. H. (2011). Scientific challenges of bioethanol production in Brazil. *Appl. Microbiol. Biotechnol.* 91 (5), 1267–1275. doi: 10.1007/s00253-011-3437-6

Loci identified through the one-dimensional scans and the genome-wide LOD threshold using Haley-Knott regression (R/qlt package) were interrogated further to determine the percent variance explained (PVE) by each QTL using the SAS GLM (general linear models) procedure. Single-factor ANOVA was used to assess the association between genotype frequencies at individual loci and each phenotype (high temperature stress tolerance and co-culture competitive growth), using a comparison-wise error rate of 0.05. Least squares means of genotypic classes were calculated with the LSMEANS option of the GLM procedure. PVE was determined by multiplying the R<sup>2</sup> value in the ANOVA output by 100, and p-values were obtained. This procedure can result in relatively high PVE values for each locus that totaled to over 100%, thus overestimating the absolute single locus PVE. Thus, we also calculated relative contribution of the loci to serve as a comparison parameter. In a second approach to estimate PVEs, multi-locus models were fit for each phenotype and tested with the GLM procedure. The model presented in **Table S8** explained the highest proportion of the phenotypic variance while, at the same time, individual loci were significant at the 0.01 significance level.

## AUTHOR CONTRIBUTIONS

NS and JA conceived the study. NS, RW, and JA designed and performed the experiments, analyzed data, generated figures, and wrote the manuscript.

## FUNDING

NS received a pre-doctoral fellowship from Brazil's CAPES (0316/13-0). Research reported here was supported by an NIH grant to JA (R35GM119788).

## ACKNOWLEDGMENTS

We thank Patrick Byrne, Joshua Granek, John McKay, and Lydia Heasley for valuable insights and comments on the manuscript. We also thank Chris Allen, Marcelo Bassalo and Steven Watson for technical assistance.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2019.00782/full#supplementary-material>

Andersen, S. L., and Petes, T. D. (2012). Reciprocal uniparental disomy in yeast. *Proc. Natl. Acad. Sci. U. S. A.* 109 (25), 9947–9952. doi: 10.1073/pnas.1207736109

Argueso, J. L., Carazzolle, M. F., Mieczkowski, P. A., Duarte, F. M., Netto, O. V., Missawa, S. K., et al. (2009). Genome structure of a *saccharomyces cerevisiae*

- strain widely used in bioethanol production. *Genome Res.* 19 (12), 2258–2270. doi: 10.1101/gr.091777.109
- Ausubel, F. M., Brent, R., Kingston, R., Moore, D., Seidman, J. G., Smith, J. A., et al. (2003). “Yeast,” in *Current Protocols in Molecular Biology*. Eds. R. B., F. M. Ausubel, R. Kingston, D. Moore, J. G. Seidman, J. Smith, and K. Struhl. (New York: John Wiley & Sons Inc.).
- Babrzadeh, F., Jalili, R., Wang, C., Shokralla, S., Pierce, S., Robinson-Mosher, A., et al. (2012). Whole-genome sequencing of the efficient industrial fuel-ethanol fermentative *Saccharomyces cerevisiae* strain CAT-1. *Mol. Genet. Genomics* 287 (6), 485–494. doi: 10.1007/s00438-012-0695-7
- Barbosa, R., Pontes, A., Santos, R. O., Montandon, G. G., de Ponzzes-Gomes, C. M., Morais, P. B., et al. (2018). Multiple rounds of artificial selection promote microbe secondary domestication—the case of cachaca yeasts. *Genome Biol. Evol.* 10 (8), 1939–1955. doi: 10.1093/gbe/evy132
- Basso, L. C., de Amorim, H. V., de Oliveira, A. J., and Lopes, M. L. (2008). Yeast selection for fuel ethanol production in Brazil. *FEMS Yeast Res.* 8 (7), 1155–1163. doi: 10.1111/j.1567-1364.2008.00428.x
- Borneman, A. R., Desany, B. A., Riches, D., Affourtit, J. P., Forgan, A. H., Pretorius, I. S., et al. (2011). Whole-genome comparison reveals novel genetic elements that characterize the genome of industrial strains of *Saccharomyces cerevisiae*. *PLoS Genet.* 7 (2), e1001287. doi: 10.1371/journal.pgen.1001287
- Borneman, A. R., Forgan, A. H., Kolouchova, R., Fraser, J. A., and Schmidt, S. A. (2016). Whole genome comparison reveals high levels of inbreeding and strain redundancy across the spectrum of commercial wine strains of *saccharomyces cerevisiae*. *G3 (Bethesda)* 6 (4), 957–971. doi: 10.1534/g3.115.025692
- Broman, K. W., Wu, H., Sen, S., and Churchill, G. A. (2003). R/qtl: QTL mapping in experimental crosses. *Bioinformatics* 19 (7), 889–890. doi: 10.1093/bioinformatics/btg112
- Carvalho-Netto, O. V., Carazzolle, M. F., Rodrigues, A., Braganca, W. O., Costa, G. G., Argueso, J. L., et al. (2013). A simple and effective set of PCR-based molecular markers for the monitoring of the *Saccharomyces cerevisiae* cell population during bioethanol fermentation. *J. Biotechnol.* 168 (4), 701–709. doi: 10.1016/j.jbiotec.2013.08.025
- Chakraborty, P., Pankajam, A. V., Dutta, A., and Nishant, K. T. (2018). Genome wide analysis of meiotic recombination in yeast: for a few SNPs more. *IUBMB Life* 70 (8), 743–752. doi: 10.1002/iub.1877
- Charlesworth, D., and Willis, J. H. (2009). The genetics of inbreeding depression. *Nat. Rev. Genet.* 10 (11), 783–796. doi: 10.1038/nrg2664
- Coutoune, N., Mulato, A. T., Riano-Pachon, D. M., and Oliveira, J. V. (2017). Draft genome sequence of *Saccharomyces cerevisiae* Barra Grande (BG-1), a Brazilian industrial bioethanol-producing strain. *Genome Announc.* 5 (13), e00111-17. doi: 10.1128/genomeA.00111-17
- Cromie, G. A., Hyma, K. E., Ludlow, C. L., Garmendia-Torres, C., Gilbert, T. L., May, P., et al. (2013). Genomic sequence diversity and population structure of *Saccharomyces cerevisiae* assessed by RAD-seq. *G3 (Bethesda)* 3 (12), 2163–2171. doi: 10.1534/g3.113.007492
- Della-Bianca, B. E., Basso, T. O., Stambuk, B. U., Basso, L. C., and Gombert, A. K. (2013). What do we know about the yeast strains from the Brazilian fuel ethanol industry? *Appl. Microbiol. Biotechnol.* 97 (3), 979–991. doi: 10.1007/s00253-012-4631-x
- Dutta, A., Lin, G., Pankajam, A. V., Chakraborty, P., Bhat, N., Steinmetz, L. M., et al. (2017). Genome dynamics of hybrid *Saccharomyces cerevisiae* during vegetative and meiotic divisions. *G3 (Bethesda)* 7 (11), 3669–3679. doi: 10.1534/g3.117.1135
- Hill, A., and Bloom, K. (1987). Genetic manipulation of centromere function. *Mol. Cell Biol.* 7 (7), 2397–2405. doi: 10.1128/MCB.7.7.2397
- Johnson, L. J., Koufopanou, V., Goddard, M. R., Hetherington, R., Schafer, S. M., and Burt, A. (2004). Population genetics of the wild yeast *Saccharomyces paradoxus*. *Genetics* 166 (1), 43–52. doi: 10.1534/genetics.166.1.43
- Legras, J. L., Galeote, V., Bigey, F., Camarasa, C., Marsit, S., Nidelet, T., et al. (2018). Adaptation of *S. cerevisiae* to fermented food environments reveals remarkable genome plasticity and the footprints of domestication. *Mol. Biol. Evol.* 35 (7), 1712–1727. doi: 10.1093/molbev/msy066
- Magwene, P. M. (2014). Revisiting mortimer's genome renewal hypothesis: heterozygosity, homothallism, and the potential for adaptation in yeast. *Adv. Exp. Med. Biol.* 781, 37–48. doi: 10.1007/978-94-007-7347-9\_3
- Magwene, P. M., Kayikci, O., Granek, J. A., Reininga, J. M., Scholl, Z., and Murray, D. (2011). Outcrossing, mitotic recombination, and life-history trade-offs shape genome evolution in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. U. S. A.* 108 (5), 1987–1992. doi: 10.1073/pnas.1012544108
- Mancera, E., Bourgon, R., Brozzi, A., Huber, W., and Steinmetz, L. M. (2008). High-resolution mapping of meiotic crossovers and non-crossovers in yeast. *Nature* 454 (7203), 479–485. doi: 10.1038/nature07135
- Peter, J., De Chiara, M., Friedrich, A., Yue, J. X., Pflieger, D., Bergstrom, A., et al. (2018). Genome evolution across 1,011 *Saccharomyces cerevisiae* isolates. *Nature* 556 (7701), 339–344. doi: 10.1038/s41586-018-0030-5
- Raghavendran, V., Basso, T.P., da Silva, J.B., Basso, L.C., and Gombert, A.K. (2017). A simple scaled down system to mimic the industrial production of first generation fuel ethanol in Brazil. *Antonie Van Leeuwenhoek* 110 (7), 971–983. doi: 10.1007/s10482-017-0868-9
- R\_Core\_Team (2017). “R: A language and environment for statistical computing,” in *R. Found. Stat. Comput.*, 3.4.0th ed. (Vienna, Austria: R Foundation for Statistical Computing).
- Reid, R. J., Sunjevaric, I., Voth, W. P., Ciccone, S., Du, W., Olsen, A. E., et al. (2008). Chromosome-scale genetic mapping using a set of 16 conditionally stable *Saccharomyces cerevisiae* chromosomes. *Genetics* 180 (4), 1799–1808. doi: 10.1534/genetics.108.087999
- Rodrigues-Prause, A., Sampaio, N. M. V., Gurol, T. M., Aguirre, G. M., Sedam, H. N. C., Chapman, M. J., et al. (2018). A case study of genomic instability in an industrial strain of *Saccharomyces cerevisiae*. *G3 (Bethesda)* 8 (11), 3703–3713. doi: 10.1534/g3.118.200446
- Sadhu, M. J., Bloom, J. S., Day, L., and Kruglyak, L. (2016). CRISPR-directed mitotic recombination enables genetic mapping without crosses. *Science* 352 (6289), 1113–1116. doi: 10.1126/science.aaf5124
- SAS (2013). SAS 9.4. 9.4th ed. Cary, NC: SAS Institute Inc.
- Smukowski Heil, C. S., DeSevo, C. G., Pai, D. A., Tucker, C. M., Hoang, M. L., and Dunham, M. J. (2017). Loss of heterozygosity drives adaptation in hybrid yeast. *Mol. Biol. Evol.* 34 (7), 1596–1612. doi: 10.1093/molbev/msx098
- Solis-Escalante, D., Kuijpers, N. G., Bongaerts, N., Bolat, I., Bosman, L., Pronk, J. T., et al. (2013). amdSYM, a new dominant recyclable marker cassette for *Saccharomyces cerevisiae*. *FEMS Yeast Res.* 13 (1), 126–139. doi: 10.1111/1567-1364.12024
- Stovicek, V., Borja, G.M., Forster, J., and Borodina, I. (2015). EasyClone 2.0: expanded toolkit of integrative vectors for stable gene expression in industrial *Saccharomyces cerevisiae* strains. *J. Ind. Microbiol. Biotechnol.* 42 (11), 1519–1531. doi: 10.1007/s10295-015-1684-8
- Swinnen, S., Goovaerts, A., Schaeferlaens, K., Dumortier, F., Verdyck, P., Souverein, K., et al. (2015). Auxotrophic mutations reduce tolerance of *Saccharomyces cerevisiae* to very high levels of ethanol stress. *Eukaryot. Cell* 14 (9), 884–897. doi: 10.1128/EC.00053-15
- Wach, A., Brachat, A., Pohlmann, R., and Philippsen, P. (1994). New heterologous modules for classical or PCR-based gene disruptions in *Saccharomyces cerevisiae*. *Yeast* 10 (13), 1793–1808.
- Wickham, H. (2009). *ggplot2: elegant graphics for data analysis*. 2nd ed. (New York: Springer-Verlag). doi: 10.1007/978-0-387-98141-3

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Sampaio, Watson and Argueso. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Quantitative Trait Nucleotides Impacting the Technological Performances of Industrial *Saccharomyces cerevisiae* Strains

Emilien Peltier<sup>1,2</sup>, Anne Friedrich<sup>3</sup>, Joseph Schacherer<sup>3</sup> and Philippe Marullo<sup>1,2\*</sup>

<sup>1</sup> Department Sciences du vivant et de la sante, Université de Bordeaux, UR CEnologie EA 4577, Bordeaux, France, <sup>2</sup> Biolauffort, Bordeaux, France, <sup>3</sup> Department Micro-organismes, Génomes, Environnement, Université de Strasbourg, CNRS, GMGM UMR 7156, Strasbourg, France

## OPEN ACCESS

### Edited by:

Isabel Sá-Correia,  
University of Lisbon,  
Portugal

### Reviewed by:

Johan M. Thevelein,  
KU Leuven, Belgium  
Himanshu Sinha,  
Indian Institute of  
Technology Madras,  
India

### \*Correspondence:

Philippe Marullo  
philippe.marullo@u-bordeaux.fr

### Specialty section:

This article was submitted to  
Evolutionary and Genomic  
Microbiology,  
a section of the journal  
Frontiers in Genetics

**Received:** 15 April 2019

**Accepted:** 01 July 2019

**Published:** 23 July 2019

### Citation:

Peltier E, Friedrich A, Schacherer J  
and Marullo P (2019) Quantitative  
Trait Nucleotides Impacting the  
Technological Performances of  
Industrial *Saccharomyces*  
*cerevisiae* Strains.  
*Front. Genet.* 10:683.  
doi: 10.3389/fgene.2019.00683

The budding yeast *Saccharomyces cerevisiae* is certainly the prime industrial microorganism and is related to many biotechnological applications including food fermentations, biofuel production, green chemistry, and drug production. A noteworthy characteristic of this species is the existence of subgroups well adapted to specific processes with some individuals showing optimal technological traits. In the last 20 years, many studies have established a link between quantitative traits and single-nucleotide polymorphisms found in hundreds of genes. These natural variations constitute a pool of QTNs (quantitative trait nucleotides) that modulate yeast traits of economic interest for industry. By selecting a subset of genes functionally validated, a total of 284 QTNs were inventoried. Their distribution across pan and core genome and their frequency within the 1,011 *Saccharomyces cerevisiae* genomes were analyzed. We found that 150 of the 284 QTNs have a frequency lower than 5%, meaning that these variants would be undetectable by genome-wide association studies (GWAS). This analysis also suggests that most of the functional variants are private to a subpopulation, possibly due to their adaptive role to specific industrial environment. In this review, we provide a literature survey of their phenotypic impact and discuss the opportunities and the limits of their use for industrial strain selection.

**Keywords:** biotechnology, fermentation, QTL, QTN, QTG, yeast, variant, aroma

## INTRODUCTION

Between individuals of the same species, a broad palette of genetic variants is found, including large chromosomal rearrangements (deletions, duplications, inversions, translocations, and introgressions) and punctual mutations (Griffiths et al., 2000). This latter type includes small insertions/deletions (InDels) as well as single-nucleotide polymorphisms (SNPs) that are by far

**Abbreviations:** 3'UTR, 3' untranslated transcribed region; ALE, adaptive laboratory evolution; CNVs, copy number variants; eQTL, expression QTL; gTME, global transcription machinery engineering; GM, genetically modified; GWAS, genome-wide association studies; GxE, genetic per environment; HMF, 5-hydroxy-methyl-furfural; InDels, insertions/deletions; MAF, minor allele frequency; MAS, marker-assisted selection; nsSNP, non-synonymous SNP; sSNP, synonymous SNP; QTL, quantitative trait loci; QTG, quantitative trait genes; QTN, quantitative trait nucleotides; RHA, reciprocal hemizygosity assay; SNP, single-nucleotide polymorphisms.

the most frequent polymorphic event found at the intraspecific level in fungi (Doniger et al., 2008), human (Sachidanandam et al., 2001), and plants (Ching et al., 2002). Depending on the organism and the genomic position, the SNP/InDel frequency ranges from  $1 \times 10^{-2}$  to  $1 \times 10^{-3}$  per base and constitutes a vast pool of genetic variants (Stucki et al., 2012; Almeida et al., 2014; Scozzari et al., 2014; Peter et al., 2018).

With the relative ease of obtaining genome-wide SNP-data, their impact on complex trait can be tracked by either genome-wide association studies (GWAS) or quantitative trait loci mapping (QTL mapping) in medicine (Beck et al., 2014) or agronomy (Brachi et al., 2011; Sharma et al., 2015). When they are statistically linked to a phenotype, these SNPs become QTNs (quantitative trait nucleotides) and could be listed in large databases for research communities (Grant et al., 2010; Youens-Clark et al., 2011). In contrast to multicellular eukaryotes, large SNP-database regrouping several studies for fungi and yeasts are not really developed. For *Saccharomyces cerevisiae* species, the first attempts to set up SNP databases have been made 10 years ago (Doniger et al., 2008; Schacherer et al., 2009) but without establishing a link between genotypic data and phenotypes.

With the emergence of high-throughput sequencing in the last 10 years, the number of available complete genomes rose impressively providing a quite complete landscape of genetic polymorphism for this species (Peter et al., 2018). A particular focus was done on strains belonging to food fermentation including wine (Borneman et al., 2016), beer (Gallone et al., 2016), distillery (Barbosa et al., 2018), and cheese/beer/distillery (Legras et al., 2018). As previously demonstrated, the *S. cerevisiae* population appears to be clearly structured according to the geography, the environmental niche, and the relation to human environment (Peter and Schacherer, 2016; Marsit et al., 2017; Peter et al., 2018). Among the food-related strains, the beer and bakery strains are polyphyletic and characterized by a high ploidy level (Gallone et al., 2016), whereas wine or sake strains are mostly diploid and derived from the genetic drift of a limited number of founders (Ohnuki et al., 2017). Since each industrial application is characterized by distinct populations, the strains of each group have been faced with specific selective pressures. These conditions have likely promoted the emergence of adaptive alleles conferring a phenotypic advantage to each particular industrial process. The identification of those adaptive mutations in the wide pool of natural variations that discriminate the different subpopulations remains a challenging task. It has been recently shown by GWAS that CNVs (copy number variants) and gross chromosomal reorganization exert a sound effect on phenotypic variation (Peter et al., 2018). However, the yeast strains of each subpopulation also exhibit a wide set of SNPs shaping their technological properties. The growing number of QTNs found in the last decade suggests that numerous functional variants will be found in the future.

In this review, we established an extensive catalog of *S. cerevisiae* QTNs experimentally validated that impact traits of biotechnological interest. First, we analyzed their allelic frequencies and their dispersion within a large population. Second, we reported their physiological effect. Third, we discussed

how these QTNs can be used for significantly improving the technological properties of industrial yeast strains.

## GENES AND POLYMORPHISMS IMPACTING QUANTITATIVE TRAITS OF INDUSTRIAL INTEREST

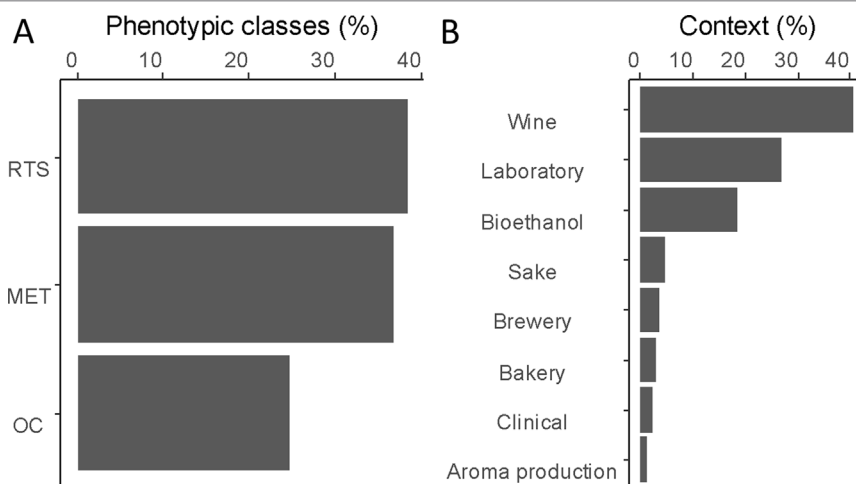
To establish an exhaustive catalog of functional variants, we focus our literature survey on QTNs impacting yeast traits relevant for biotechnological applications. Traits were sorted in three main phenotypic classes: traits linked to metabolism (e.g., nitrogen, carbon, vitamin, and fermentation activity), traits linked to stress resistance (e.g., acidic and basic, temperature, osmotic, and ethanol), and traits impacting the organoleptic properties of the products (Figure 1A). Most of QTNs were identified by linkage analysis, demonstrating the efficiency of this strategy in yeast (Liti and Louis, 2012; Fay, 2013). Other functional variants were identified by mutagenesis, comparative genomics, and adaptive laboratory evolution (ALE) approaches (Gresham and Hong, 2014). More recently, genome-wide associations were performed on an extensive set of fully sequenced natural isolates, providing a large list of SNPs statistically associated with the measured phenotypic diversity (Peter et al., 2018; Sardi et al., 2018).

To provide a functional analysis, we included only genes that have been experimentally validated by reciprocal hemizyosity assay (RHA) (Steinmetz et al., 2002) or allele swapping (Storici and Resnick, 2006). In this context, a total of 147 QTGs (quantitative trait genes) were reported and described across a set of 85 articles (Table S1). In total, 71% of these genes were identified in an industrial context since they concern media and/or strains related to aroma production (1%), bioethanol (18%), or traditional fermented goods including wine (41%), sake (5%), bakery (3%), or beer (3%). The remaining 30% were identified by using laboratory or clinical strains cultivated in non-industrial conditions (Figure 1B). However, these genes were included since they possibly affect industrial properties [e.g., growth fitness, stress resistance, and flocculation]. We tested the overall distribution of such genes across yeast genome, and no hotspot was found (hypergeometric distribution, with sliding windows of 100 kb and a 10-kb step, 1,000 permutation tests, False Discovery Rate (FDR) = 5%)(Figure 2).

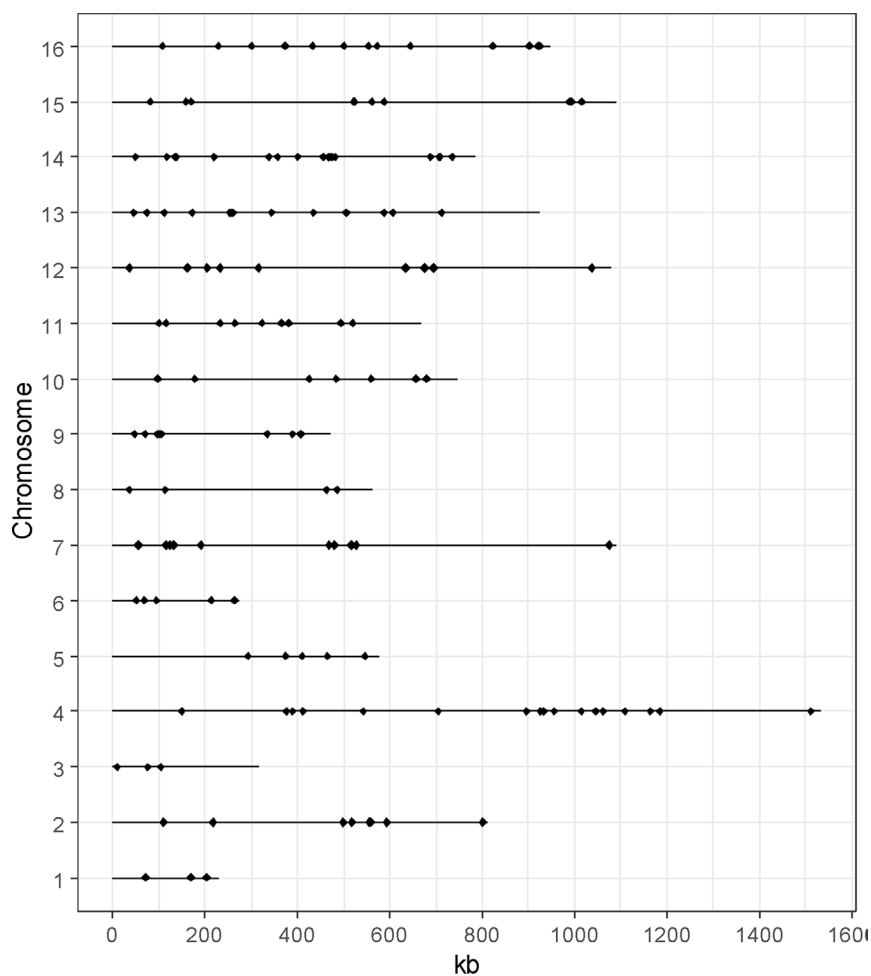
Among these 147 QTGs, a significant enrichment was obtained for Gene Ontology terms (goTermFinder, <https://www.yeastgenome.org>) (Figure 3). This is the case for function and process terms related to transcription (nucleic acid-binding transcription factor activity, DNA binding, protein-binding transcription factor activity, and transcription factor binding) ( $p$ -value < 0.05) and transport (plasma membrane, transmembrane transporter activity, regulation of transport, amino acids transport, and carbohydrate transport) ( $p$ -value < 0.05). The strong enrichment of such categories confirms that these genes are important levers for generating phenotypic variability.

In *S. cerevisiae*, the pangenome was recently defined using a population of 1,011 natural isolates. Overall, 4,940 core

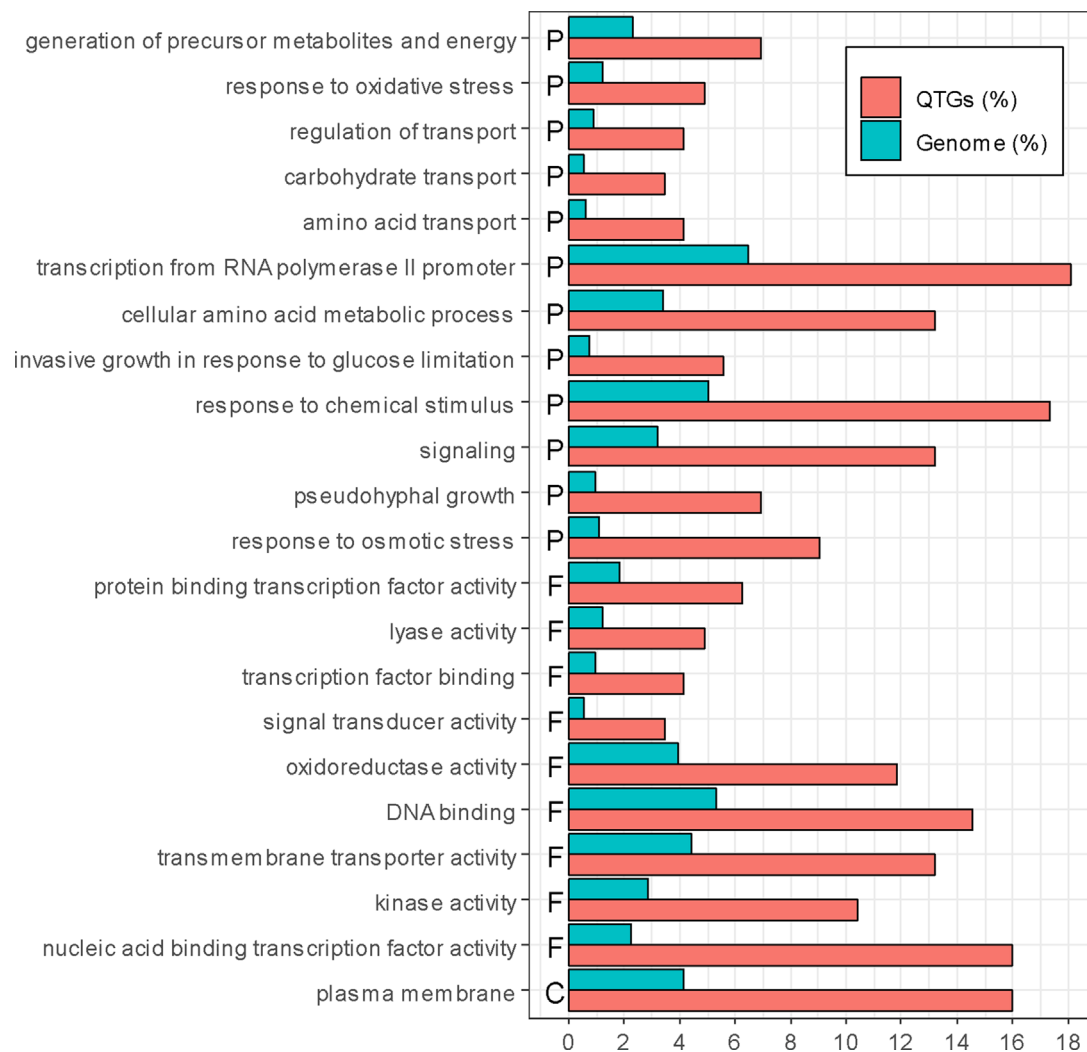




**FIGURE 1 |** Overview of identified quantitative trait genes (QTNs). **(A)** Proportion of genetic variants according to three main categories. Resistance to stress (RTS), central metabolism (MET), and organoleptic compound (OC). **(B)** Proportion of QTNs according to the context of the study in which they were identified.



**FIGURE 2 |** Distribution of the QTNs along the genome. Each dot represents the position of one QTN. No QTN hotspot was found (hypergeometric distribution, 1,000 permutations test, FDR = 5%).



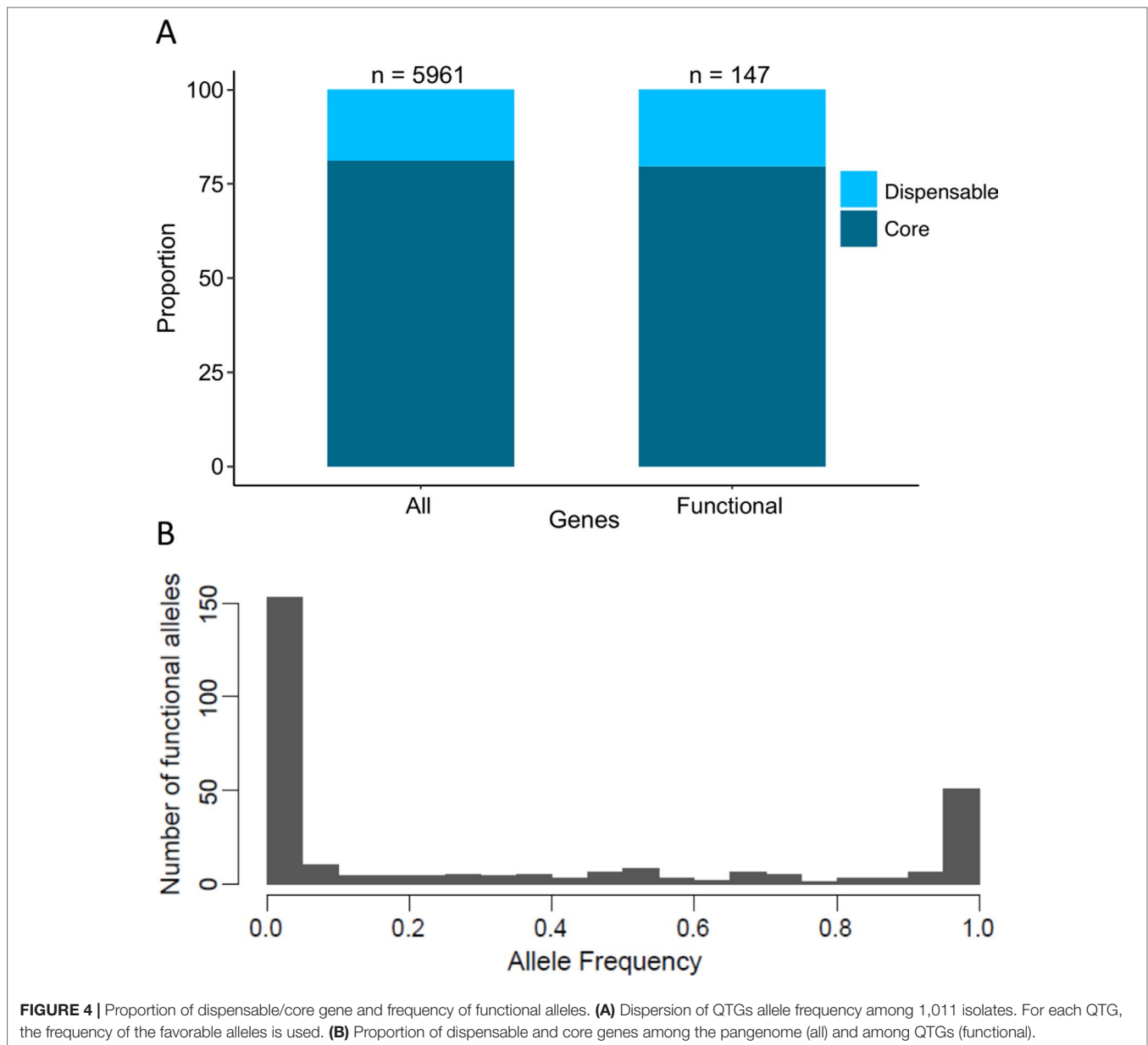
**FIGURE 3 |** Enrichment of QTNs. Only significantly over-represented or under-represented categories are represented (chi-squared test,  $p$ -value < 0.05, with Bonferroni adjustment for multiple tests). The percentage of genes in the genome and in the QTNs is indicated in blue and red, respectively. F, P, and C represent molecular function, biological process, and cellular component, respectively.

genes and 2,856 accessory genes were determined within the population (Peter et al., 2018). Interestingly, a strong enrichment has been shown for genes whose function is related to adaptation to the environment in the subset of accessory genes (Peter et al., 2018). Across the 147 identified QTNs, 117 and 30 are part of the core and accessory genomes, respectively (**Figure 4A**). This proportion clearly shows that a large fraction of the QTNs are part of the conserved core genome. Moreover, there is no bias toward the subset of accessory genes, although they are more prone to be involved to adaptation processes.

In order to focus our review at the SNP level, we listed all the possible genetic polymorphisms found in these QTNs. Since most of the articles validated genes but not at the SNP level, we took into account for each QTN all the genetic polymorphisms described by the authors. These 284 QTNs can be sorted in several categories according to the type of genetic polymorphisms that discriminate their allelic variations. Most of the identified allelic

variations correspond to one or a combination of missense mutations, also called non-synonymous substitutions (nsSNPs) (81%). Other minor cases correspond to SNPs or InDels in 5'UTR or 3'UTR regions (6%), InDels in the coding sequence (6%), synonymous SNPs (sSNPs) (4%), translocation (2%), and short tandem repeats.

The functional impact of a subset of 251 QTNs, corresponding to missense mutations, was estimated by using the predictive program SIFT (sorting tolerant from intolerant) (Ng and Henikoff, 2003). It turns out that 168 mutations are predicted to be tolerated, whereas only 83 are predicted to be deleterious, suggesting that most of the QTNs do not lead to a loss of function (**Table S2**). By contrast, 22 QTNs (approximately 8%) correspond to nonsense mutations, meaning that genetic variant results in a premature stop codon generating in most of the cases to a loss of function. In addition, one and two QTNs lead to the loss of start and stop codons, respectively.



One main feature of the *S. cerevisiae* species is the bias toward low-frequency variant. Among the 1,011 recently sequenced genomes, much of the detected genetic polymorphisms are very low-frequency variants with 93% of them having a minor allele frequency lower than 0.1 (Peter et al., 2018). This observation raises the question regarding the impact of these variants on the phenotypic diversity. Among the 284 QTNs, more than 150 have a frequency lower than 5%, meaning that these variants would be undetectable by GWAS (Figure 4B).

Population genomic studies in *S. cerevisiae* also allowed to define precisely the different subpopulations, which are related to either the ecological or geographical origins. Based on the 1,011 isolates, a total of 26 subpopulations were defined (Peter et al., 2018). Interestingly, the identified QTNs are not evenly distributed across these subpopulations, and biases toward some

specific of them are observed (Table S3). For example, 26 QTNs are private to the wine subpopulation and are only found in 9 to 50 wine isolates. Two SNPs located the *MSN2* and *MSN4* genes are exclusively found in the sake subpopulation. And finally, a mutation in the *MAL33* gene is private to the African beer population. This observation clearly suggests that most of the functional variants are private to a subpopulation. This could result to the adaptation to a specific industrial environment. However, since wine and sake subpopulations are derived from a limited number of founders, their presence could be also explained by simple genetic drift.

Overall, this set of 284 QTNs is very insightful but does not reflect the genetic architecture in its entirety. The genotypic landscape is not limited to SNPs located in protein coding regions. Indeed, SNPs as well as InDels located in promoter and

3' untranslated regions were identified for some complex traits. These genetic variants do not affect the protein sequence but can impact the transcription level, mRNA processing, translation, export, and decay. Mutations altering a functional motif in the promoter region have been identified many times (Shahsavariani et al., 2012; Zimmer et al., 2014; Cubillos, 2016; Salinas et al., 2016; Tapia et al., 2018). They constitute allele-specific expression (ASE) changes that are certainly a source of phenotypic variation (see Cubillos, 2016, for a review). In addition, structural variants such as CNVs and translocations were identified as involved in the variation of some specific industrial traits. Recently, genome-wide association analyses, performed on a large collection of *S. cerevisiae* isolates, highlighted the importance of the CNVs, which explain a more considerable proportion of the phenotypic variance and have greater effects on phenotype compared with the SNPs (Peter et al., 2018). However, the detection of the structural variants at a population scale is yet technically limited, and consequently, their global impact on the complex traits is still to be explored.

## INDUSTRIAL TRAITS IMPACTED BY NATURAL GENETIC VARIANTS

In this section, we shortly described the phenotypic impact of most of QTNs/QTNs reviewed. The phenotypes surveyed were arranged in three main subsections concerning central metabolism, resistance to toxins and stresses, and organoleptic contribution.

### Central Metabolism

In total, 71 natural alleles have been identified, and they impact 70 traits related to the central metabolism (Table 1 and Table S1).

### Nitrogen and Vitamin Metabolism

Nitrogen sources as well as the vitamin composition can vary in a wide range according to the raw vegetal material, the fertilization method, and the harvest date. Their composition may drastically affect the yeast fermentation performances in beer (Gibson et al., 2007), wine (Bell and Henschke, 2005), or bio-ethanol productions (Hahn-Hägerdal et al., 2005). Depending on the genetic background, the strain's ability to use various nitrogen sources differs between subpopulations (Ibstedt et al., 2014) and within strains of the same industrial process (Jiranek et al., 1995; Manginot et al., 1998; Gutiérrez et al., 2013; Brice et al., 2014).

The identification of genetic factors controlling nitrogen consumption has been achieved by many QTL mapping studies (Ambroset et al., 2011; Brice et al., 2014; Ibstedt et al., 2014; Jara et al., 2014), one large-scale hemizygosity analysis (Gutiérrez et al., 2013), and one ALE experiment (Gresham and Hong, 2014). These studies revealed relevant genetic variants that could be used for improving the performance of fermenting yeast. In such studies, the diverse media employed consist of either a mixture of different nitrogen sources, mimicking a natural medium (Ambroset et al., 2011; Brice et al., 2014; Jara et al., 2014), or

several distinct media containing each a single nitrogen source (Gutiérrez et al., 2013; Gresham and Hong, 2014; Ibstedt et al., 2014). When a single nitrogen source was used, the identified QTLs were chiefly due to deleterious mutations in genes involved in the pathway of the amino acid concerned (Ibstedt et al., 2014). In contrast, in mixed nitrogen media, the effects of the QTLs identified are more pleiotropic and impact a group of amino acids sharing the same biochemical structure (Jara et al., 2014).

Deleterious alleles, impairing the use of a particular nitrogen source, were identified for proline (*PUT4*), allantoin (*DAL1* and *DAL4*) (Ibstedt et al., 2014), and methionine (*ARO8*, *VBA3*, and *ADE5,7*) (Gutiérrez et al., 2013). Similar deleterious mutations were also found for asparagine (*ASP1*) (Marullo et al., 2007a) or for folic acid metabolism (*ABZ1*), having an impact on wine fermentation kinetics (Ambroset et al., 2011). These recessive mutations are rare and generally of poor interest because industrial practices require prototrophic strains. However, these alleles can be used as auxotrophic markers for achieving breeding programs in a non-GMO context (Steensels et al., 2014) as the *ura3* and *lys2* markers (Timberlake et al., 2011; Dufour et al., 2013).

More interestingly, three pleiotropic genes (*GLT1*, *ASI1*, and *AGP1*) impacting consumption of several amino acids were identified by measuring the consumption profile of amino acids (Jara et al., 2014). By implementing a multi-parental design (SGRP-X4), these authors identified four additional genes (*ASI2*, *CPS1*, *LYP1*, and *ALP1*) involved in the consumption of aromatic and basic amino acids. In the same study, the comparative RNA-seq profiling of extreme progeny clones allowed the identification of two additional genes (*PDC1* and *ARO1*) that influence the amino acid consumption in the wine fermentation (Cubillos et al., 2017). Following a similar strategy, the progenies of two enological strains were phenotyped for their fermentation capacity in a synthetic grape must containing a low assimilable nitrogen level (Brice et al., 2014). Four genes directly or indirectly linked with the nitrogen metabolism were identified (*BIO3*, *GCN1*, *ARG81*, and *MDS3*).

The expression level and the stability of proteins involved in nitrogen catabolism may also contribute to nitrogen assimilation. For example, the ASE of *ASN1*, the asparagine synthetase, modulates the consumption of aspartic and glutamic acid in wine-related fermentations (Salinas et al., 2016). Amino acid assimilation, in particular for proline, can be also induced by increasing the half-life of the membrane transporter Put4p by changing some N-terminal arginine residues involved in its ubiquitination (Omura et al., 2005). Although the optimal consumption of nitrogen source is generally considered as a suitable technological trait, the rapid and complete consumption of amino acids may negatively affect fermentation capacities (Martí-Raga et al., 2015) and reduces the chronological life span (Kwan et al., 2011) in specific conditions. All together, these studies support that multiple molecular mechanisms impact the nitrogen assimilation including the nitrogen signaling pathways, metabolic enzymes, and protein degradation as summarized in Figure 5.



**TABLE 1 |** Nature of SNP grouped by phenotypes.

Phenotype categories	Genes involved (experimentally validated)	Number and nature of the molecular cause involved							No data	References	
		Total	nsSNP	InDels	5' or 3'UTR positions	Translocation	STR	sSNP			
Central metabolism											
Nitrogen requirement, growth rate on specific nitrogen sources, ammonium and amino acid uptake, vitamin biosynthesis	<i>GCN1, DAL1, DAL4, LST4, PUT4, GAT1, MEP2, ABZ1, ALP1, ARO1, ASI2, CPS1, LYP1, PDC1, RIM15, ASN1, AGP1, BUL2, GLT1, ADE5, ARO8, VBA3 ASI1, ARG81, BIO3, MDS3, ASP1</i>	29	18	2	2	0	0	0	7	(Omura et al., 2005; Marullo et al., 2007a; Ambroset et al., 2011; Kwan et al., 2011; Gutiérrez et al., 2013; Hong and Gresham, 2014; Brice et al., 2014; Ibstedt et al., 2014; Jara et al., 2014; Salinas et al., 2016; Cubillos et al., 2017)	
Sugar catabolism, fructose, glucose, maltose, and maltotriose uptake, glucose–galactose switch, diauxic switch	<i>HXT3, MTH1, GAL80, MAL33, MAL11, HAP4, MBR1, ADR1, BMH1, YJR030c</i>	10	10	0	0	0	0	0	0	(Higgins et al., 1999; Segrè et al., 2006; Guillaume et al., 2007; Vidgren et al., 2009; Kvitek and Sherlock, 2011; Salinas et al., 2012)	
Glycerol metabolism	<i>SSK1, GPD1, HOT1, SMP1, GUT1, GAT1, YFL040W, TAO3, ADH3</i>	10	7	1	0	0	0	0	2	(Salinas et al., 2012; Hubmann et al., 2013a; Hubmann et al., 2013b; Swinnen et al., 2013; Wilkening et al., 2014; Tapia et al., 2018)	
Acetic acid production	<i>YAP1, FAS2, ASP1, ALD6</i>	4	3	0	1	0	0	0	0	(Inokoshi et al., 1994; Marullo et al., 2007a; Salinas et al., 2012; Cordente et al., 2013)	
Fermentation rate and completion, <i>fil</i> phenotype	<i>CYR1, GPR1, HAP4, MBR1, YJR030c, ABZ1, GDB1, MSB2, PDR1, PMA1, VMA13, PDR3, SSU1, MSN2, MSN4, RIM15, ADR1</i>	18	15	1	1	1	0	0	0	(Mizoguchi et al., 2002; Ambroset et al., 2011; Watanabe et al., 2011; Salinas et al., 2012; Watanabe et al., 2013; Cubillos, 2016; Martí-Raga et al., 2017; Peltier et al., 2018)	
Resistance to toxins and stresses											
Ethanol accumulation capacity, ethanol tolerance, growth on ethanol	<i>ADE1, KIN3, SSD1, UTH1, VPS70, IAI11, APJ1, MKT1, SWS2, MEX67, PRT1, VPS70, SPT5, TAO3</i>	14	12	0	2	0	0	0	0	(Yang et al., 2011; Avrahami-Moyal et al., 2012; Salinas et al., 2012; Swinnen et al., 2012b; Hubmann et al., 2013a; Hubmann et al., 2013b; Pais et al., 2013; Duitama et al., 2014; Wilkening et al., 2014; Voordeckers et al., 2015)	
High-temperature growth, temperature tolerance, low temperature adaptation, freezing tolerance	<i>CDC19, IRA1, IRA2, RSP5, END3, MKT1, RHO2, NCS2, TAO3, PRP42, SMD2, GAA1, FPK1, OPT2, PET494, QCR2, OYE2, VHS1</i>	18	14	0	2	0	0	1	1	(Sinha et al., 2006; Sinha et al., 2008; Parts et al., 2011; Benjaphokee et al., 2012; Shahsavarani et al., 2012; Li et al., 2013; Wilkening et al., 2014; García-Ríos et al., 2017; Marullo et al., 2019)	
Stress tolerance, oxidative, osmotic	<i>PRO1, MPR1, AQY1, AQY2, RCK2, MOT2, MPR1, RDS2, PUT1, SPT5</i>	12	8	4	0	0	0	0	0	(Sekine et al., 2007; Takagi, 2008; Will et al., 2010; Diezmann and Dietrich, 2011; Dhar et al., 2011; Sasano et al., 2012; Greetham et al., 2014)	
Toxins resistance: acidic, basic, phenolic compounds, SO <sub>2</sub>	<i>WAR1, ADH3, ASG1, GIS4, SKS1, COX20, CUP2, DOT5, GLO1, VMA7, HAA1, CDC23, ECM21, GPH1, IES2, MAC1, NMD4, SGT2, SPS100, LEU3, MNE1, SAP190, ADH1, MKT1, RAD5, UBP7, SSU1</i>	31	12	0	4	2	0	3	11	(Pérez-Ortín et al., 2002; Boaz et al., 2008; Demogines et al., 2008; Romano et al., 2010; Yang et al., 2011; Brion et al., 2013; Zimmer et al., 2014; Pinel et al., 2015; González-Ramos et al., 2016; Meijnen et al., 2016; Swinnen et al., 2017; Sardi et al., 2018)	

TABLE 1 | Continued

Phenotype categories	Genes involved (experimentally validated)	Number and nature of the molecular cause involved					No data	References	
		Total	nsSNP	InDels	5' or 3'UTR positions	STR			
Organoleptic compounds production		19	19	0	0	0	0	(Akada et al., 1999; Salinas et al., 2012; Steyer et al., 2012; Trindade de Carvalho et al., 2017; Eder et al., 2018)	
Acids, higher alcohols, and relative ester production	<i>IXR1, LEU9, RGS2, RGT1, FAS2, ABZ1, TOR1, NRG1, FLX1, MDH2, AGP1, ALP1, ILV6, MAE1, AGP2, FAS1, SIR2, EAT1, SNF8</i>	9	5	1	0	1	0	2	(Cordente et al., 2009; Linderholm et al., 2010; Huang et al., 2014; Noble et al., 2015)
Sulfur compounds	<i>MET10, MET5, MET2, IRC7, URE2, SKP2, MET6, CYS4, SSU1</i>	3	3	0	0	0	0	0	(Chambon et al., 1991; Steyer et al., 2013)
Other aromas	<i>PAD1, PDR8, ERG20</i>	17	13	0	0	0	3	1	(Yvert et al., 2003; Verstrepen et al., 2005; Fidalgo et al., 2006; Nogami et al., 2007; Yang et al., 2011; Li et al., 2013; Taylor and Ehrenreich, 2014; Wilkening et al., 2014)
Adhesion, flocculation, clumpiness	<i>FLO1, SFL1, GPA1, AMN1, RGA1, CDC28, FLO8, END3, IRA2, MSS11, TRR1, FLO11</i>								

### Sugar Catabolism

#### Control of Fermentative and Respiratory Switch

The expanded use of *S. cerevisiae* in biotechnology is likely due to its strong efficiency to dissimilate small sugars by both respiratory and fermentative routes (Pronk et al., 1996; Zampar et al., 2013). In the past, deleterious mutants unable to switch between fermentative and respiratory metabolism have been identified for *CAT8* (Zimmermann et al., 1977) and *ADR1* (Denis and Young, 1983). Besides these drastic mutations, the fermentation/respiration balance is controlled by many other genes, and considerable variations have been measured within strains and species (Quirós et al., 2014). Some alleles impacting this metabolic control have been identified in *HAP4* and *MBR1* (Salinas et al., 2012). Both genes are involved in mitochondrial function, and *HAP4* is a key transcription factor activating respiratory genes (Zampar et al., 2013). These particular alleles may partially activate the Krebs cycle, reducing the fermentation efficiency of yeast in winemaking conditions.

A similar alteration in fermentation/respiration balance was found in sake strains by a comparative transcriptomic approach (Watanabe et al., 2013). Two distinct loss-of-function alleles have been identified for *ADR1* in the sake strains K7 and K701. This transcription factor is activated when glucose becomes limiting (diauxic shift) and promotes ethanol catabolism by activating the transcription of *ADH2*. Therefore, strains lacking a functional *ADR1* might have an accelerated alcoholic fermentation in a sake-brewing context.

#### Sugar Uptake and Assimilation

Although *S. cerevisiae* is able to ferment many sugars, their uptake and catabolism obey priority rules. This regulation, called glucose catabolite repression, limits the speed and efficiency of the fermentation of many sugars (Gancedo, 1998). Industrial conditions offer selective constraints that have promoted the constitutive activation of non-preferred carbon sources uptake and catabolism.

An example is the fermentation of maltose and maltotriose, the two most abundant sugars in brewing wort (Stewart, 2006), which are also present in bakery doughs. The fermentation performance of beer strains is therefore defined by their capacity to transport those sugars. Depending on the species, i.e., *S. cerevisiae* (ale group) or *S. pastorianus* (lager group), different  $\alpha$ -glucoside transporters have been described (*ScAGT1*, *ScMALx1*, *SeMALx1*, *SeAGT1*, *MTT1*, and *ScMPHx*). Most of the ale beer strains have a full-length *AGT1* gene, which ensures complete wort fermentation. By contrast, the lager strains have a premature stop codon at the position 1183 that reduces their maltotriose uptake (Vidgren et al., 2009). Moreover, the *MAL* loci are present in several subtelomeric regions, resulting in CNVs that affect the transport capacity of the strains (Brown et al., 2010). The impact of sugar assimilation on fermentation kinetics has been also demonstrated for the *SUC2* gene, encoding for the invertase, that can be present in numerous subtelomeric regions especially in bakery strains (Carlson and Botstein, 1983; Ness and Aigle, 1995). Besides their integrity and their copy number,  $\alpha$ -glucoside transporters may be also differentially regulated. A documented example



industry and especially in winemaking. This demand moves in the direction of public health policy and could be in part solved by the development of new strains (Kutyna et al., 2010; Dequin et al., 2017). The most explored route consists in redirecting a part of the glycolytic flux toward the production of glycerol since this compound is organoleptically neutral. Allelic variations in *GAT1*, *YFL040W*, *GPD1*, and *ADH3* impacting glycerol yield (Salinas et al., 2012; Tapia et al., 2018) were identified by the same laboratory in a wine context. However, the highest production observed did not modify the sugar-to-ethanol conversion in a significant manner.

### Acetic Acid Production

Acetic acid can be produced by spoilage microorganisms but is also produced by yeast during the alcoholic fermentation (Vilela-Moura et al., 2010). Acetic acid production level is a quantitative trait that varies across isolates (Giudici and Zambonelli, 1992; Marullo et al., 2004). In Lambic beers and some other sour beer styles, acetic acid can be a desirable component that contributes to the complexity of the flavor and aroma profile. However, in beverage industries, acetic acid generally has a negative impact on the organoleptic quality especially for wine (Ribéreau-Gayon et al., 2006). Therefore, the allelic variants controlling acid acetic production during fermentation were mostly identified in a winemaking context.

The major gene involved in acetic acid production in wine fermentation is *ALD6*, which encodes an aldehyde dehydrogenase (Remize et al., 2000; Saint-Prix et al., 2004). Depending on the yeast isolate, this gene is differentially expressed leading to various acetic acid production levels. Two SNPs in the promoter region of the *ALD6* gene were linked to this variation in an enological context (Salinas et al., 2012). Due to its role in the redox balance homeostasis, the acetic acid production level could also correlate with cell growth. A deleterious mutation in the catalytic core of the asparaginase *ASP1* gene was linked to the slow assimilation of asparagine in synthetic must. This affects the cell growth, causing overproduction of acetic acid in asparagine-rich media (Marullo et al., 2007a). More relevant are the premature stop codons found in the *YAP1* gene (*YAP1*<sup>Q541X</sup> and *YAP1*<sup>Q573X</sup>). Those mutations were isolated by screening cerulenin-resistant mutants of the wine starter PDM. They dramatically decreased (~30%) acetic acid production during the fermentation of different grape juices (Cordente et al., 2013). Yap1p is a transcriptional factor involved in oxidative stress, and the metabolic link between these mutations and the phenotype was not perfectly understood. Interestingly, these mutants showed a high alcohol dehydrogenase activity with an increased production of acetate esters that could reduce the acetic acid production (see below).

### Fermentation Efficiency

Several allelic variants affecting the overall fermentation rate in diverse industrial contexts have been identified. In the bakery context, a noteworthy trait is the *FIL* phenotype that consists in losing stress resistance when the alcoholic fermentation is initiated. Some strains show a recessive *fil* phenotype (fermentation-induced loss of stress resistance) and conserve a suitable resistance to heating and freezing even if they are collected

in exponential growth phase during alcoholic fermentation (Thevelein et al., 1999). This particular feature can be obtained by using specific mutations in the gene *CYR1* (adenylate cyclase) at the position G1682L or in the gene *GPR1* (G-protein coupled receptor). The *fil* phenotype conferred is particularly relevant for obtaining full active strains for the fermentation from frozen doughs or active dry yeast.

Other allelic variants affecting the fermentation rate itself were identified in high-gravity matrices such as sake and wine. Genes impacted may encode for unexpected functions like the glycogen debranching protein (Gdp1p) (Cubillos, 2016). Interestingly, the positive allele of the *GPD1* gene is located in both the promoter and coding regions. Although the physiological mechanisms of these natural variants are not particularly linked to alcoholic fermentation, the allele of a wine-related strain (WE) confers a faster fermentation. Other mechanisms of adaptation were found by investigating the performance of wine yeast in the second fermentation that took place in locked bottles (*méthode champenoise*). Fermentation kinetics were measured by following the CO<sub>2</sub> pressure rise inside the bottle. Two genes encoding for components of the plasma (*PMA1*) and vacuolar (*VMA13*) membrane ATPases were identified. Positive alleles provide a faster fermentation kinetics in low pH conditions (2.8), which is a particular feature of sparkling wines. The same authors identified two other genes related to osmotic regulation (*MSB2*) and multidrug resistance (*PDR1*) (Martí-Raga et al., 2017). Interestingly, transcriptional regulators of this multidrug resistance family (PDR network) have been also linked to the fermentation resistance in a sake-brewing context. By applying a drug resistance screening, different alleles of the transcriptional factors *PDR1* (M308I) and *PDR3* (L950S, G948D, and G957D) were isolated. These alleles drastically improved the fermentation efficiency of sake strain, allowing the production of more than 21% (v/v) of ethanol in industrial trials (Mizoguchi et al., 2002).

Another relevant industrial property of fermenting yeast is the length of the lag phase that could be particularly critical for achieving the inoculation of non-sterile musts. The genetic determinism of the lag phase has been partially elucidated in wine fermentations by performing a QTL mapping between two wine-related strains (Zimmer et al., 2014). In this work, a major QTL explaining relevant differences in the lag phase duration (more than 24 h) was identified. In this specific case, the molecular cause of phenotypic discrepancy is due to a reciprocal translocation event (XV-t-XVI) involving the gene *SSU1*, which encodes for a sulfite pump. This gross chromosomal rearrangement increased the expression level of *SSU1* in the parental strain GN that achieves a rapid fermentation start in synthetic grape juice containing SO<sub>2</sub>. This work illustrates an interesting case of phenotypic convergence since two other independent chromosomal rearrangements, VIII-t-XVI (Pérez-Ortín et al., 2002) and inv-XVI (García-Ríos et al., 2019), targeting *SSU1* and conferring SO<sub>2</sub> resistance were identified. Recently, the pleiotropic effect of these translocations has been demonstrated by seeking for QTLs that interact with environmental conditions (Peltier et al., 2018). Depending on the nature of the grape juice and the amount of free SO<sub>2</sub> in the medium, the translocations associated with *SSU1* may impact the production of SO<sub>2</sub>, lag phase, and fermentation rate.



## Resistance to Toxins and Stresses

Industrial applications are characterized by a broad set of stresses such as osmotic, temperature, ethanol, pH, nutrient limitation, and presence of various toxins that affect the yeast cell growth and viability (Bauer and Pretorius, 2000; Gibson et al., 2007; Zhao and Bai, 2009; Sicard and Legras, 2011). Each particular stress activates specific and general stress responses, ensuring a better physiological adaptation (see Gasch, 2003). Although most of these stresses are common, each biotechnological process recreates particular conditions explaining the emergence of yeast strains adapted to each specific process (Sicard and Legras, 2011; Albertin et al., 2011). In this section, we point out natural genetic polymorphisms in 75 genes that impact the yeast resistance to several types of stress commonly found in biotechnological applications (Table 1 and Table S1).

### Ethanol Tolerance

Ethanol accumulated during the fermentation impacts negatively the more sensitive strains, impairing the fermentation completion. Stuck fermentations affect the ethanol production yield and also the microbiological stability of beverages due to the presence of residual sugars. The selection of ethanol tolerant strains constitutes a real challenge in particular for sake and wine production where high concentrations are reached (respectively, 20% and 17%). Several hundred genes associated with ethanol tolerance were identified by functional genetics (see Ma and Liu, 2010; Snoek et al., 2016, for a review); however, identifications of causative SNPs are more rare.

Among the *S. cerevisiae* species, the sake strains demonstrate the highest capacity of ethanol accumulation with concentration reaching approximately 20% (v/v). Several genetic causes explaining this characteristic have been identified by comparing the transcriptome between sake and laboratory strains during sake fermentation (Wu et al., 2006; Shobayashi et al., 2007). Sake strains carry deleterious alleles in genes involved in the stress response and quiescent phase entry (*RIM15* and *MSN2/MSN4*) (Watanabe et al., 2011; Watanabe et al., 2012). In addition, they lack the *PPT1* genes involved in the heat shock stress response (Noguchi et al., 2012). These mutations private to the sake group (Table S3) may explain their high ethanol accumulation capacity.

Global transcription machinery engineering (gTME) was used to generate strains more tolerant to ethanol. One of the targeted genes is *SPT5* encoding for a TATA-binding protein associated with ethanol tolerance (Alper et al., 2006). The random mutations generated globally modify the yeast transcriptional response, providing mutants with a higher ethanol tolerance (Yang et al., 2011). This strategy allowed the isolation of two *SPT5* haplotypes that enhanced the ethanol production from 8% to 10% for the yeast strain L3262. Those alleles also confer a better tolerance to hyperosmotic stress (Kim et al., 2013), another character highly desirable for high-gravity brewing (HBV) fermentation.

Besides the differential activation of these transcriptional pathways, numerous QTNs impacting ethanol tolerance have been successfully identified by reverse genetics. Since ethanol exerts a toxic effect, resistant strains can be readily screened by applying selective media where only a part of the population overcomes a desired threshold. Historically, adaptive laboratory

evolution has been successfully used (Chen et al., 2010; Stanley et al., 2010; Avrahami-Moyal et al., 2012; Voordeckers et al., 2015). By increasing ethanol concentration in turbidostatic cultures, causative SNPs in the *SSD1* and *UTH1* genes were identified (Avrahami-Moyal et al., 2012). The selective pressure imposed (up to 8% ethanol) and the laboratory background used in this study are, however, a bit far from the conditions met in industrial fermentations. In similar conditions, a long-term evolution experiment was applied for 200 generations in six independent bioreactors (Voordeckers et al., 2015). Ploidy level changes and CNVs were mainly observed. Some QTNs were also identified: alleles *VPS70*<sup>C590A</sup>, *PRT1*<sup>A1384G</sup>, *IAI1*<sup>G479T</sup>, and *MEX67*<sup>G456A</sup> conferred an adaptive advantage by impacting remarkably diverse molecular functions such as mRNA export (*Mex67p*), vacuolar protein sorting (*Vps70p*), and protein synthesis (*Prt1p*).

QTL mapping approaches were also used for exploring the tolerance to ethanol with concentrations closer to industrial conditions (up to 20%) (Swinnen et al., 2012a; Pais et al., 2013; Duitama et al., 2014). In these studies, segregants were screened on YPD plates with different ethanol concentrations and genotyped using a bulk-segregant analysis (BSA). RHA validated the effect of five genes involved in the ethanol tolerance: *MKT1*, *SWS2*, *APJ1*, *ADE1*, and *KIN3*. The negative impact of the S288c allele of *MKT1* was found in both studies and is due to a rare deleterious allele (minor allele frequency (MAF) = 0.1%) that has a strong impact on many characters of the laboratory strain (Lee et al., 2009). This type of defective allele is not present in industrial strains and cannot be used for their improvement. However, pairs of positive alleles (*KIN3* and *ADE1*) (Pais et al., 2013) and (*APJ1* and *SWS2*) (Swinnen et al., 2012a) were brought by parental strains isolated from sake and biofuel, respectively. For *APJ1*, a clear ASE effect was observed demonstrating that the expression of *APJ1* seems to be deleterious for high ethanol tolerance (Swinnen et al., 2012a).

### Cold and High Temperature Tolerance

Each industrial process is characterized by the application of temperatures that often meet yeast physiological limits. In biofuel industry, distillery yeast must tolerate a high temperature (up to 40°C) in order to ferment in association with enzymatic cocktails used for the saccharification (Olofsson et al., 2008). The wine or sake industry imposes milder conditions with temperature that does not exceed 35°C (Marullo et al., 2009). Furthermore, cold temperature tolerance (6–15°C) can be required especially in brewery and white wine context. Finally, the bakery process imposes an extremely broad range of temperature by applying freeze/thaw cycles.

The first quantitative genetics study in yeast focused on high-temperature growth (HTG) phenotype, which consists in measuring the colony size after a 48-h culture at 41°C (Steinmetz et al., 2002). The dissection of a major QTL by RHA allowed identification of three genes (*MKT1*, *END3*, and *RHO2*) explaining the quantitative phenotypic variation between the laboratory strain S288c and a clinical isolate YJM145. The identification of causative SNPs was achieved few years later (Sinha et al., 2006) by using site-directed mutagenesis. *RHO2* causative SNP is located in the 3'UTR region, which constitutes

a rare example of polymorphism outside the coding sequence. A supplemental gene impacting this phenotype, *NCS2*, was finally identified by using a backcross strategy in order to eliminate the effect of the main segregating QTL (Sinha et al., 2008). Using a bulk sequencing analysis strategy (Yang et al., 2013) confirmed the deleterious inheritance of S288c alleles for *NCS2* and *MKT1*. It also identified two other causative genes (*PRP42* and *SMD2*) encoding for proteins belonging to the same spliceosome complex, suggesting complex epistatic relations. Indeed, the authors demonstrated that the thermotolerant alleles were *PRP42*<sup>S288c</sup> and *SMD2*<sup>MUCL2817</sup>. However, when the *PRP42*<sup>S288c</sup> was introduced in the MUCL2817 genetic background, no additional effect was observed.

By implementing panmictic crosses (advanced intercross lines) between two natural isolates (West African and North American strains), Parts et al. (2011) underlined the role of RAS/cAMP signaling pathway for high-temperature growth by demonstrating the impact of the two paralogues (*IRA1* and *IRA2*) encoding for the RAS inhibitor proteins. Similar phenotypes were also elucidated at the gene level and concern the pyruvate kinase protein Cdc19p (Benjaphokee et al., 2012) or the E3 ubiquitin ligase *Rsp5p* (Shahsavarani et al., 2012).

All these studies investigated the temperature tolerance of *S. cerevisiae* strains by evaluating their growing capacities. However, in industrial fermentation, the deleterious effect of high temperature is generally coupled with many other stressful conditions including high ethanol content (Marullo et al., 2009; Mitumasu et al., 2014), presence of toxins (Taherzadeh et al., 2000; Hasunuma et al., 2011), and sterol or nitrogen depletion (Bely et al., 1990; Marullo et al., 2009).

In a recent study, two genes, *OYE2* and *VHS1*, impacting the fermentation rate at high temperature were identified in a winemaking context (Marullo et al., 2019). In both cases, an SNP was generating a codon-stop insertion impairing the completion of the fermentation above 30°C. Interestingly, for the *VHS1* gene, the truncated protein of 371 amino acids confers a more efficient fermentation above 30°C. The recently documented function of *Vhs1p* (Simpson-Lavy et al., 2017) allows establishing a link between respire-fermentative switch and the fermentation efficiency at high temperature.

Alternatively, some industrial processes (especially food-related fermentations) impose cold and negative temperatures that may have drastic consequences on yeast fitness. If the physiological and molecular mechanism of cold tolerance has been understood (Gunde-Cimerman et al., 2014), few studies have investigated the natural genetic variation impacting cold adaptation.

In a recent work, genes related to lipid remodeling in the plasma membrane or mitochondrial metabolism were identified (García-Ríos et al., 2017). The phenotype investigated was the specific ability to growth at low temperature (15°C) in a synthetic grape juice. The selective genotyping of a pool of progenies derived from two wine-related strains (P24 and P5) was used. The impact of the *FPK1* gene encoding for a protein kinase that regulates phospholipid translocation and membrane asymmetry was clearly demonstrated. A substitution R520K in the P24 strain seems to be deleterious for growth at 15°C but

not at 28°C. Interestingly, most of the QTLs mapped are located in subtelomeric regions of chromosomes XIII, XV, and XVI. Many of tested genes affect the time to achieve the fermentation, underlining the role of mitochondrial proteins (*QCR2* and *PET494*), oligopeptide transporter (*OPT2*), and aquaporin (*AQY1*), which seem particularly important for maintaining a fermentation activity at low temperature. The same team carried out an evolution experiment using the P5 strain and identified an nsSNP in the gene *GAA1*, which encodes for a protein belonging to the GPI-protein transamidase complex. The introduction of a threonine at the position 108 of this protein enhances the growth fitness at 12°C, suggesting the involvement of mannoproteins in cold adaptation (García-Ríos et al., 2014).

The impact of yeast aquaporins encoded by *AQY1* and its paralogue *AQY2* was also implicated in freezing and osmotic tolerance. Many loss-of-function mutations in these genes are present in the *Saccharomyces cerevisiae* wild population (Will et al., 2010). Moreover, differential expression was described for *AQY2* and are due to polymorphisms in the promoter region (Fay et al., 2004). In a bakery context, it is well known that functional and overexpressed aquaporins enhance freezing tolerance, while non-functional ones promote osmotic tolerance (Tanghe et al., 2002; Will et al., 2010). This suggests that *AQY* genes are submitted to balancing selection (Will et al., 2010).

Other mutations enhancing the proline accumulation also improve freezing resistance. The intracellular storage of this amino acid confers resistance to many stresses including freezing, desiccation, oxidation, and ethanol and enhances the fermentation kinetics (see Takagi, 2008 for review; Kitagaki and Takagi, 2014). Mutations in the *PRO1* were generated by either the selection of proline-analogue resistant mutants (Takagi et al., 1997) or by PCR random mutagenesis (Sekine et al., 2007). These mutations desensitize Pro1p against the feedback inhibition exerted by proline. Disruption of *PUT1* involved in the proline degradation pathway also enhances freezing resistance (Takagi, 2008) and was successfully combined with *Pro1p* (*D154N*) or (*I150T*) mutations in a self-cloned diploid strain (Kaino et al., 2008). Interestingly, a similar combination of a *Pro1p* mutation with an *Mpr1p* variant (*F65L*) enhances both freezing and air-dry resistance (Sasano et al., 2012).

## Osmotic Stress

Most of the industrial processes involving *Saccharomyces* species are characterized by a high sugar content that goes hand in hand with a severe osmotic stress. Improving the resistance to osmotic pressure is therefore essential for achieving high-gravity ethanol fermentations. The osmotic stress response in yeast has been widely investigated (Tamás and Hohmann, 2003), and several mutations affecting osmotolerance have been intensively identified (Hohmann, 2002). Natural or induced mutations have been reported in many distinct pathways, including Hog1p activation (*Pbs2p*<sup>K389M</sup> (Reiser et al., 2000) and *Sln1p*<sup>P1148S/P1196L</sup> (Fassler et al., 1997)), proline accumulation (*PRO1*), and water efflux (*AQY1-2*) (see above). More recently, a QTL mapping study identified genetic causes of osmotic shock associated with very-high-gravity ethanol fermentations using the SGRP-X4 design (Greetham et al., 2014). The alleles *RCK2*<sup>Q113/S456</sup> of the

wine strain DBVPG6765 enhance osmotolerance. This kinase performs a regulatory role in the *Hog1p* pathway and is involved in osmotic stress response since its overexpression improved growth in high osmotic conditions (Teige et al., 2001).

## Resistance to Toxins

In the last decade, the biofuel industry has moved from the first- to second-generation production. This technological progress consists in transforming pentose sugars present in the plant cell walls in ethanol, in addition to hexoses. This additional step required lignocellulose pre-treatments that release aromatic and acidic compounds that are detrimental to the growth of *S. cerevisiae* (Palmqvist and Hahn-Hägerdal, 2000; Klinker et al., 2004). First of all, yeast has to face acetic acid and other weak acids that decrease the cytosolic pH, inhibit growth, and remodel gene expression (see Palma et al., 2018, for a review).

In a QTL mapping study, the *COX20<sup>QR</sup>* allele of the cytochrome *c* oxidase assembly factor has been identified (Greetham et al., 2014). This allele confers sensitivity to mild concentrations of acetic acid and other weak acids (formic and levulinic) likely linked with the programmed cell death response. In a similar way, the fermentation kinetics in a culture medium spiked with acetic acid was measured for a progeny population derived from a cross between a biofuel strain (ethanol red) and an acetic acid-resistant strain JT22689. A major QTL was mapped on chromosome XVI and was associated with an nsSNP impacting the coding sequence of the gene *HAA1*. The resistant strain carries a unique polymorphism at the nucleic position 1571 that generates a single amino acid change S505N (*Haa1*\*) (Meijnen et al., 2016). This transcription factor plays a central role in the *S. cerevisiae* adaptation and tolerance to weak acids (Palma et al., 2018). The *Haa1*\* allele activates the expression of plasma membrane acetate exporters Tpo2p and Tpo3p. Interestingly, another punctual mutation (*S135F*) promoting acetic acid resistance has been identified by a gTEM approach (Swinnen et al., 2017).

In a second round of QTL mapping performed in a panmictic population (inbred lines), four other causative genes have been detected (*CUP2*, *VMA7*, *GLO1*, and *DOT5*) with the positive contribution of the acetic acid-resistant strain JT22689 (Meijnen et al., 2016). Interestingly, *CUP2* is a paralogue of *HAA1*, suggesting that the expression of acetate transporter is a preferential target of acetic acid resistance. The *VMA7* gene is involved in the vacuolar-pH homeostasis and was previously linked to acetic acid resistance. The last two genes, *GLO1* and *DOT5*, were never linked to weak acid resistance. They are related to osmotic and oxidative stresses, respectively.

Adaptive laboratory evolution experiments were also implemented for obtaining mutations enhancing the acetic acid resistance. After ~50 transfers of alternative microaerobic batch cultivations (with and without acetic acid), five independent evolved cultures showing a strong resistance to acetic acid were obtained. Four causal mutations in the genes *ASG1*, *ADH3*, *SKS1*, and *GIS4* were identified by genome sequencing and validated by allele replacement (González-Ramos et al., 2016). Three of them (*ADH3*, *SKS1*, and *GIS4*) were not previously associated with acetic acid tolerance, providing new clues for understanding

this complex trait. Interestingly, other genes related to weak acid resistance were found in a winemaking context. Although acetic acid does not impact cell growth in winemaking, a relevant allele impacting the weak acid resistance has been identified by eQTL mapping (Brion et al., 2013). By analyzing the whole-genome expression profile of 44 progeny clones (BY × 59A cross), these authors identified five nsSNPs in the coding sequence of *WAR1* gene. This gene encodes for a transcription factor that controls the expression of a plasma membrane ABC transporter responsible (Pdr12p) for organic acid efflux. Allele swapping experiment demonstrated that the *WAR1<sup>59A</sup>* wine yeast allele increases the expression level of *PDR12* and enhances the sorbic acid resistance.

Cellulose and hemicellulose hydrolysis released furfural and 5-hydroxy-methyl-furfural (HMF) that have a strong toxic effect on yeast growth and fermentation (Palmqvist and Hahn-Hägerdal, 2000). To date, only one allele of the main alcohol dehydrogenase (*ADH1*) favoring resistance to HMF was identified in an industrial isolate. The protein sequence reveals multiple amino acid polymorphisms close to the substrate binding pocket (S109P, L116S, and Y294C). This allele has a NADH-dependent HMF reductase activity, which is not present in any other strains and allows reduction of HMF in a nontoxic form (Boaz et al., 2008).

Toxin resistance of lignocellulosic raw material has been recently investigated by two other original approaches. First, a sophisticated strategy combining mutagenesis, genome shuffling, and phenotypic selection was implemented in order to isolate mutations, enhancing resistance to hardwood spent sulfite liquor (Pinel et al., 2015). Among a dozen of putative SNPs identified by whole-genome sequencing, these authors demonstrated that a single mutation in the coding sequence of the gene *UBP7* (2466 T > A) conferred a better tolerance to this medium. Second, a GWAS linked 76 SNPs with growth traits measured in complete hydrolysates spiked with a toxin cocktail (Sardi et al., 2018). The association was performed by keeping SNPs from having a minor allele frequency greater than 2% in a collection of 165 fully sequenced *S. cerevisiae* strains. The effect of allelic variants in the *LEU3*, *MNE1*, and *SAP190* genes was validated by RHA in different genetic backgrounds (Sardi et al., 2018).

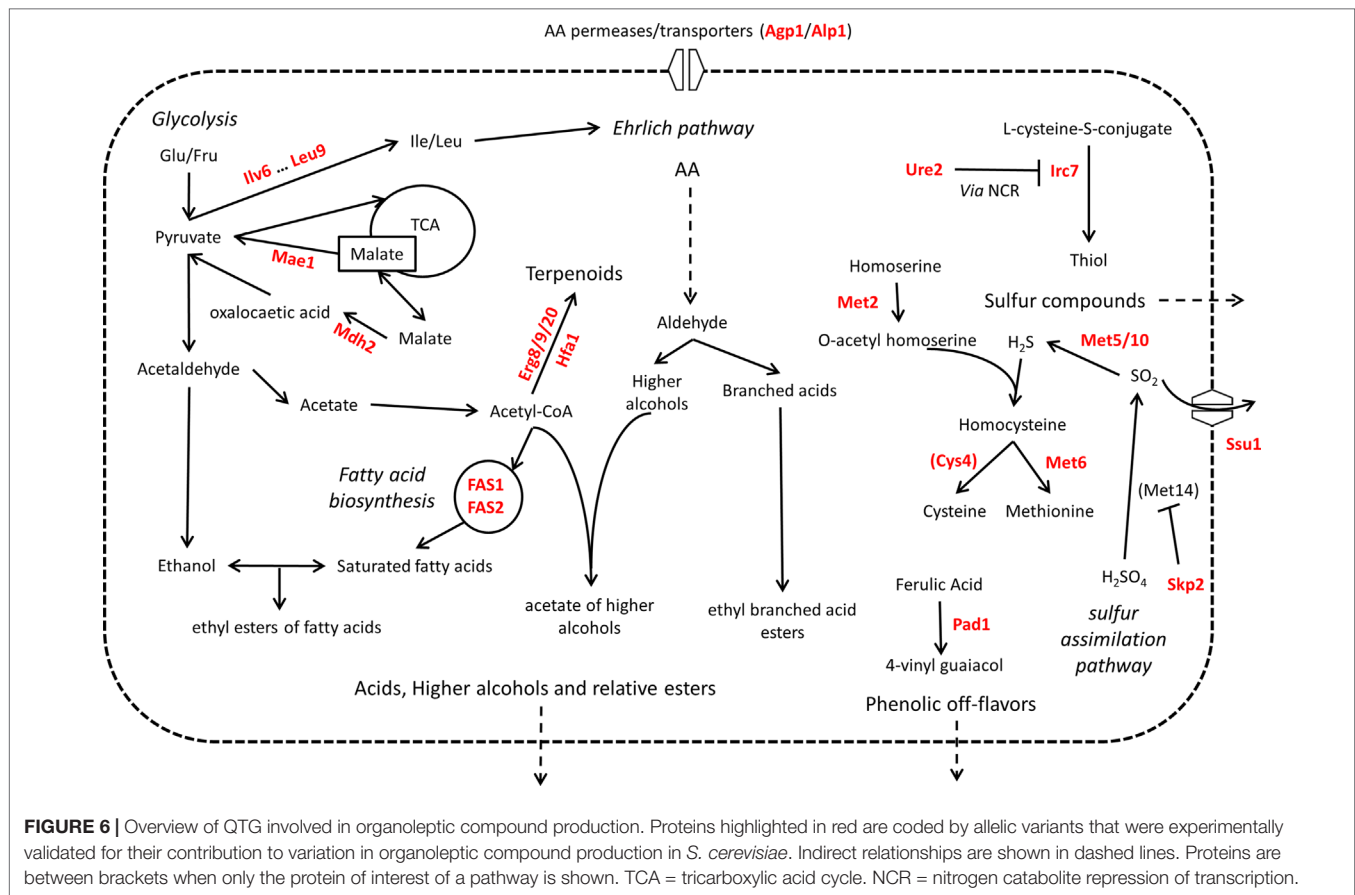
## Organoleptic Properties

When they are used in food-related fermentation processes, the fermentation efficiency is not the only technological property desired. Yeast strains are also selected depending on their impact on the composition of several organoleptic compounds. In this last subsection, we reviewed the role of 48 genes and their relative QTNs that influence the organoleptic quality of beverages (Figure 6).

### Higher Alcohols and Esters *de Novo* Synthesis

Higher alcohols and esters constitute groups of volatile compounds that are *de novo* produced during the alcoholic fermentation. Their organoleptic impact in fermentation beverage has been widely reviewed (Mason et al., 2000; Sumby et al., 2010), and the main genes and enzymatic activities controlling





their biosynthesis have been identified (Malcorps et al., 1991; Verstrepen et al., 2003a; Saerens et al., 2006). Some nsSNPs in the sequence of these enzymes modulate the biosynthesis of these compounds. For example, the brewing yeast strains expressing a long form of the *ATF1* genes (*LgATF1*) produce more acetate esters of higher alcohols (Verstrepen et al., 2003a). *ATF1* is also responsible for the majority of acetate ester production, an undesirable compound with a solvent-like off-flavor (Verstrepen et al., 2003b). To identify genetic factors responsible for the remaining acetate ester production, a QTL mapping study was carried out with or without *ATF1* deletion in parental strains (Holt et al., 2018). QTNs in *EAT1* and *SNF8* were identified, with rare alleles that prevent acetate ester production.

Moreover, many genetic variations affecting proteins in connection with esterification reactions were described. They concern the Ehrlich pathway, the acetyl-CoA production, and the lipid biosynthesis. For example, allelic variation in *ILV6* gene stimulates the production level of 2-methyl-propyl acetate by enhancing the biosynthesis of  $\alpha$ -ketoisovalerate, its related precursor (Eder et al., 2018). The activity of the FAS complex (fatty acid synthase) has a direct effect on the biosynthesis of many esters. This complex of two proteins (Fas1p and Fas2p) synthesizes fatty acid by the repeated condensation of malonyl-CoA and acetyl-CoA. Since medium fatty acids (C6–C12) are the precursors of ethyl esters, their intracellular concentrations are directly linked to those of ethyl esters. Moreover, the activity

FAS complex regulates the cytoplasmic acetyl-CoA pool; this metabolite is the substrate alcohol acetyl-CoA transferase (Atf1p and Atf2p) that produces acetic esters of higher alcohols. Therefore, mutations in the FAS complex affect both acetic and ethyl ester biosynthesis. For instance, ethyl esters biosynthesis in sake (i.e., ethyl-caproate) is modulated by the *fas2*<sup>G1250C</sup> defective allele that could be easily obtained due to their resistance to cerulenin (Aritomi et al., 2004). Interestingly, natural allelic variants in the *FAS1* gene explain a similar phenotypic discrepancy in a winemaking context (Eder et al., 2018). Other allelic versions of the *FAS2* gene alter the acetylation of higher alcohols like phenyl-ethanol acetate and isoamyl-acetate (Trindade de Carvalho et al., 2017).

The production level of esters and higher alcohols is also impacted by amino acid uptake. Indeed, several nsSNPs in the amino acid transporters encoded by the *ALP1*, *AGP1*, and *AGP2* genes control the phenotypic variance of acetate esters of higher alcohols observed among a large progeny (Eder et al., 2018). The same authors identified more distant enzymatic activities impacting the intracellular pool of acetyl-CoA (*SIR2*) or pyruvate (*MAE1*) that are building blocks of ester precursors. The framework between these precursor pathways and the esterification reactions shed light on the complex determinism of these aromas. The perturbation of folic acid biosynthesis has been also reported to modulate the production of phenyl-ethanol by identifying the deleterious allele *ABZ1* via linkage mapping



(Steyer et al., 2012). Finally, the deleterious mutation *TOR1<sup>E216\*</sup>* affecting the regulatory protein Tor1p has been also reported as reducing ester production by an unknown mechanism (Trindade de Carvalho et al., 2017).

## Sulfur Compounds

Volatile sulfur compounds are important contributors to the flavor of many foods. These molecules are characterized by a low sensory detection threshold due to the high volatility of sulfur atoms. During alcoholic fermentation, *Saccharomyces cerevisiae* releases different classes of volatile sulfur compounds that could be *de novo* produced or bio-converted from precursor molecules. Although present at very low concentration levels (close to ppb), differences within strains strongly contribute to the quality of the final product. A key node of sulfur compound production is centered on the sulfite reductase enzymatic activity involved in the reduction of sulfate into sulfide ions, which ensure the incorporation of a sulfur atom in both methionine and cysteine. A leak of sulfide ions during alcoholic fermentation results in the overproduction of hydrogen sulfide (H<sub>2</sub>S) and high undesirable off-flavors in beverage industry (Swiegers and Pretorius, 2007).

Several approaches were used to identify genetic variants that might decrease H<sub>2</sub>S production. By screening an EMS-mutagenized population, several alleles reducing H<sub>2</sub>S production were found in the *CYS4*, *MET6*, *MET5*, and *MET10* genes (Linderholm et al., 2006; Cordente et al., 2009; Linderholm et al., 2010). A linkage analysis revealed natural genetic variations in this pathway affecting the *MET1<sup>(A458T,T511I,G687D,E805K)</sup>*, *MET2<sup>R301G</sup>*, and *MET5<sup>V288X</sup>* genes that strongly modulate the H<sub>2</sub>S production during alcoholic fermentation (Huang et al., 2014). Interestingly, the *MET2<sup>R301G</sup>* allele was independently detected by another QTL mapping study as an enhancer of SO<sub>2</sub> production (Noble et al., 2015) and is quite frequent in wine European group (Table S2). This finding highlights the fact that a reduced H<sub>2</sub>S production is often coupled with a high SO<sub>2</sub> production as reported for *MET2* (Noble et al., 2015), *MET5* (Cordente et al., 2009), and *MET10* (Cordente et al., 2009; Huang et al., 2014). More interestingly, Noble et al. identified beneficial alleles in the *SKP2* gene that reduce both the SO<sub>2</sub> and H<sub>2</sub>S productions. This protein encodes an F-box factor that regulates the abundance of Met14p involved in the first steps of the sulfate reduction pathway as previously demonstrated in a brewery context (Yoshida et al., 2010). By introducing an appropriate allelic variant (*T357I*) using a backcross strategy, the authors demonstrated that wine strains producing low SO<sub>2</sub> amount can be readily selected (Blondin et al., 2014).

Another route for reducing the SO<sub>2</sub> production consists in selecting strains devoid of a translocated *SSU1* allele (either XV-t-XVI or VIII-t-XVI; see Fermentation Efficiency). Since the *Ssu1p* transporter pumps the SO<sub>2</sub> outside the cell, the less active allelic forms (non-translocated) reduce the concentration of this toxic compound in the fermented matrices. Although non-translocated alleles are not suitable in white grape juice fermentations due to the high level of free SO<sub>2</sub>, this feature can be used for red-grape juice fermentation (Peltier et al., 2018).

Sulfur volatile compounds may also contribute positively to the aroma of fermented beverages. One of the most achieved examples

is given by the bioconversion of volatile thiols (4MMP, 3MH, and 3MHA) that contribute to the typicity of Sauvignon blanc wines (Tominaga and Dubourdieu, 2000). These powerful odorant molecules are derived from cysteinylated and glutathionylated precursors present in grape juice that are converted in volatile thiols by yeast  $\beta$ -lyases (Marullo and Dubourdieu, 2010; Coetzee and du Toit, 2012). The molecular dissection of this bioconversion led to the identification of different  $\beta$ -lyases. A relevant allele of *IRC7* explains most of variations in terms of 4MMP production (Roncoroni et al., 2011). The positive allele of *IRC7* was carried by a clinical *S. cerevisiae* isolate having a full-length protein of 400 amino acids able to convert the cysteinylated precursor in its relative aroma, whereas the 60-bp truncated form is not functional. The expression level of this gene is controlled by the nitrogen catabolite repression. Indeed, the use of *ure2* mutations is useful for enhancing the bioconversion rate of volatile thiols by enhancing the enzymatic activity of  $\beta$ -lyases (Thibon et al., 2008). Since non-GMO UV-mutations of *ure2* can be readily obtained (Marullo et al., 2008), appropriate *ure2* and *IRC7<sup>400</sup>* alleles were implemented for enhancing the volatile thiols bioconversion in order to select yeast strains expressing more intense notes of exotic fruits (Dufour et al., 2013).

## Enhancement of Terpenoid Biosynthesis

Terpenoids constitute a wide class of natural molecules that can be produced by metabolic engineering for the production of antibiotics, anticancer, and other medicinal products and also for their aromas and fragrances (Ajikumar et al., 2008). These molecules can be synthesized by rerouting the sterol pathway that was first characterized in *Saccharomyces cerevisiae*. Natural genetic variations in the ergosterol pathway were identified for increasing the production of geraniol and linalool using a leaky FDP synthase (*Erg20p*) mutant (Chambon et al., 1991). This allele was introgressed into a wine yeast genetic background by repeated backcrosses (Javelot et al., 1991). The resulting strain has a satisfactory terpene production and aroma profile. Nevertheless, the alteration of the sterol synthesis pathway affected the ethanol tolerance, limiting its use in winemaking. A unique QTL mapping on terpene biosynthesis was carried out and allows the identification of one *PDR8* allele that increases nerolidol release in a synthetic media spiked with geraniol (Steyer et al., 2013).

## Off-Flavor Reduction

Phenolic off-flavors (POFs), 4-vinyl phenol and 4-vinyl guaiacol, are unwanted compounds produced by yeast during beer fermentation and also in white wine production from phenolic acids present in the must. Therefore, a desirable characteristic of fermenting yeast is the phenotype POF- that is determined by the combined action of two genes localized in the subtelomeric region of chromosome IV. The *PAD1* gene encodes for a flavin prenyl-transferase, which catalyzes the formation of a prenylated cofactor required for the ferulic acid decarboxylase encoded by *FDC1* (Clausen et al., 1994; Mukai et al., 2010). A recent whole-genome sequencing project (Gallone et al., 2016) reveals that most of brewery strains and also some wine yeast strains have loss-of-function mutations in both *PAD1(Q86\*, Y98\*, and W102\*)*

and *FDC1*(K54\*, Q154\*, c.495\_496insA, c.864delA, R309\*, and W497\*). These loss-of-function mutations are strongly correlated with the *POF*- phenotype (Mertens et al., 2017). Another *PADI* allele (*PADI*<sup>D213G</sup>) has been identified by QTL mapping (Marullo et al., 2007b) and was used for the selection of various white wine strains to avoid the production of these undesirable compounds (Marullo, 2009).

### Flocculation Properties

Flocculation is the yeast cells' capacity to co-aggregate and form flocs that can provoke cell sedimentation or the creation of a floating biofilm at the broth surface (Vidgren and Londesborough, 2011). These characteristics are important in the brewing industry where flocculation is necessary for yeast sedimentation at the end of the fermentation (Vidgren and Londesborough, 2011) or in winemaking where flor yeast development is essential to sherry wine production (Ossa et al., 1987; Fidalgo et al., 2006). Natural genetic variants that affect the flocculation capacity have been identified especially in the well-studied *FLO* genes that mostly encode for surface proteins.

The ability to form a floating biofilm is driven by an allelic version of *FLO11* in which a rearrangement within the central repeat domain leads to a more hydrophobic protein, and a 111-bp deletion in the promoter increases its expression (Fidalgo et al., 2006). *FLO11* not only concerns floating biofilm as the number of repetitions in the Flo11p central domain is also positively correlated to the flocculation strength (Wilkening et al., 2014). The same pattern is also found for Flo1p (Verstrepen et al., 2005). In the S288C genetic background, a premature stop codon in *FLO8* impairs flocculation. This gene encodes a transcription factor promoting the *FLO1* expression (Li et al., 2013). Conversely, a premature stop codon in the gene *SFL1* encoding a repressor of flocculation-related genes increases flocculation (Wilkening et al., 2014). Besides flocculation, the clumpiness of strains may be due to several allelic variants in the *AMN1*, *GPA1*, *RGA1*, *CDC28*, *FLO8*, *END3*, *IRA2*, *MSS11*, and *TRR1* genes, which have been identified in laboratory conditions. These non-sexual adhesion properties could play a critical role in industry (Pretorius, 2000).

### Moving from QTN Detection to Industrial Applications

In the previous section, we listed a catalogue of 147 genes with natural or induced genetic variants that impact quantitative traits of industrial relevance at least in one specific genetic background. Most of them correspond to alleles that could be considered as a promising reservoir of functional levers to modulate metabolic pathways and biological activity of the prime industrial microorganism, *Saccharomyces cerevisiae*. Since their number is steadily increasing, QTNs should have a profound impact for improving the selection methods of industrial strains in the future. These polymorphisms can be introduced in any desired yeast "chassis" by using GM (genetically modified) organisms as it was perfectly illustrated in the context of bioethanol production (Maurer et al., 2017). Depending on the industrial field and the local legal regulation, these QTNs may also be

exploited by implementing more classical breeding strategies (Marullo et al., 2007b; Blondin et al., 2013; Dufour et al., 2013). Although the identification of causative SNPs is now a routine task, their efficiency for improving technological properties by marker-assisted selection (no GM) or allele replacement (GM) is still unpredictable. The two most critical issues that yeast researchers are facing are the incomplete penetrance/expressivity level of identified SNPs and the gene–environment interaction modulating their effect. These issues are well recognized in agronomical science and explain why numerous markers identified in academic studies failed to be translated in the domain of application (Xu and Crouch, 2008). In this last section, we documented examples in yeast and evaluate possible solutions to overcome them.

### The Low Penetrance/Expressivity Issue

The penetrance is defined as the proportion of individuals in a population that express a phenotype associated with a specific genetic variation. Indeed, a genetic variation may affect the phenotype of some individuals but remains silent in other backgrounds. A practical illustration of this phenomenon is given by *MET10*, *EAT1*, and *SNF8* alleles, which do not have the same effect according to the genetic backgrounds (Linderholm et al., 2010; Holt et al., 2018). Therefore, deleterious modifications of a key enzyme in a well-defined pathway do not ensure a predictable phenotype. Low penetrance examples have been particularly well documented in the four parents cross designed by Ed Louis's group, suggesting that most QTLs are largely cross dependent (Cubillos et al., 2013).

The simplest explanation is given by the fact that most of the strains carry loss-of-function alleles that are preferably detected by QTL analysis. Among the 284 nsSNPs reported, 8% confer a nonsense mutation. Moreover, 33% of the referenced missense mutations have a possible deleterious effect (see Genes and Polymorphisms Impacting Quantitative Traits of Industrial Interest). Therefore, up to 60% of QTNs reviewed have a MAF lower than 5%. This is confirmed by two recent studies showing that rare variants explain a disproportionately large part of the variation of quantitative trait in yeast (Bloom et al., 2019; Fournier et al., 2019). This high proportion is explained by the yeast life history with clearly defined subpopulations that have been evolved mostly by genetic drift in separated habitats (Liti et al., 2009; Peter et al., 2018). Although easy to identify, these loss-of-function alleles are scarcely relevant since they negatively impact the phenotype of interest. This is, for instance, the case of the *OYE2*<sup>Ser77sf</sup>, *ASP1*<sup>D142H</sup>, and *ABZ1*<sup>S288c</sup> alleles that drastically reduce fermentation performances in an enological context (Marullo et al., 2007a; Ambroset et al., 2011; Marullo et al., 2019). As a consequence, the opposite alleles that were defined as favorable have a very low impact since they are present in most of the other genetic backgrounds. To overcome this problem, different approaches were used for identifying minor QTLs (Brem et al., 2005; Sinha et al., 2008; Yang et al., 2013; Holt et al., 2018; Marullo et al., 2019). The segregation of the most impacting loci is eliminated from the segregants by its deletion in the hybrid, by selecting only a part of the segregants for linkage analysis or by generating new segregants with a targeted backcross. These

strategies can help to identify QTLs with lower effects that better reflect the true differences between industrial strains.

In some cases, recessive, deleterious mutations have an industrial relevance when loss-of-function is associated with the desired trait as for the POF character (Marullo et al., 2007b; Mertens et al., 2017), and the *fil* phenotype (Thevelein et al., 1999). In natural isolates, most of these deleterious SNPs are masked in diploid progenitors since they are recessive. Therefore, before starting any QTL detection programs, it is relevant to discard meiotic progenies showing extremely “bad” phenotypes because they would transmit such deleterious SNPs.

Besides that, incomplete penetrance mostly results from genetic interactions between the causative allele identified and many other loci that impair its complete expressivity. Indeed, the Mendelian inheritance of a trait may turn out to be a quantitative in other genetic backgrounds (Hou et al., 2016). This is caused by the presence of modifier (epistatic) loci that modulate the expressivity of a major locus. Such epistatic loci have been identified in different species and in particular in yeast (Yadav and Sinha, 2018). First of all, epistasis concerns genes belonging to the same pathways and in particular between upstream regulator(s) and downstream effector(s). For instance, the favorable *GPD1*<sup>L164P</sup> allele had a reducing effect on glycerol yield for biofuel industry. This positive effect is only observed when two of its transcriptional factors (*HOT1* and *SMP1*) have the laboratory strain genotype (Hubmann et al., 2013b). Another example is given by the positive allelic combination of the fully active  $\beta$ -lyase *Irc7p*<sup>LT</sup> and the loss-of-function allele of the regulator *Ure2p*<sup>G181E</sup> that enhances the bioconversion of volatile thiols from cysteine conjugates precursors of grape juice (Dufour et al., 2013). In the same way, a strong positive epistatic interaction was found between *FLX1* and *MDH2* genes. Both genes play a role in the Krebs cycle, and the combination of *FLX1*<sup>SA</sup> and *MDH2*<sup>WE</sup> results in high levels of succinic acid during wine fermentation (Salinas et al., 2012). Besides these “obvious” metabolic connections, other interactions have been identified between functionally unrelated couples of genes. This is the case of *NCS2*–*MKT1* (Sinha et al., 2008) and *END3*–*RHO2* (Sinha et al., 2006), which strongly impacted HTG phenotype in the BY–RM11 cross. Once identified and understood, these epistatic relationships would be useful for dramatically enhancing a phenotype of interest by introducing suitable allelic combinations. Interestingly, a larger genetic modification such as aneuploidies (Sirr et al., 2015) may also play a modifier role and would be more difficult to control.

### Genetic per Environment (GxE) Interactions

Environmental conditions represent the second major factor that drastically modifies the expressivity of genetic loci. It is noteworthy that among individuals of the same species, phenotypic plasticity is frequently observed (Veerkamp et al., 1994; Pigliucci and Kolodnynska, 2002). These different non-parallel norms of reaction are due to GxE interactions. In yeast, the systematic research of genetic loci interacting with the environment has been achieved in several fundamental studies focusing on whole-genome expression level (Smith and

Kruglyak, 2008; Gagneur et al., 2009), or growth traits (Gagneur et al., 2009; Bhatia et al., 2014; Wei and Zhang, 2017; Yadav et al., 2016) measured in divergent laboratory conditions. GxE interactions observed can be divided into two broad classes (Yadav and Sinha, 2018).

First, the effect of a locus may be environment-specific, reflecting the presence of a gene/allele adapted to a particular composition of the medium. This is often observed for traits related to a specific sugar transport (*MAL13*) (Bhatia et al., 2014) or for the adaptation to a specific toxin (*SSU1*) (Pérez-Ortín et al., 2002; Zimmer et al., 2014). Second, GxE interactions may explain phenotypic trade-offs illustrated by individuals showing a contrasted fitness across a pair of environments (Yadav et al., 2015; Wei and Zhang, 2017). These antagonistic effects are mostly due to the presence of one allele that has been positively selected in one environment but shows a negative effect in another one. Such contrasted responses are frequently observed when phenotyping is done in drastically different conditions such as respiratory versus fermentative conditions. In such cases, allelic variations of key regulatory genes have been detected in *IRA2* (Smith and Kruglyak, 2008) or *HAP1* (Wei and Zhang, 2017), which are both involved in metabolic pathway switches. Since most of these studies had a fundamental focus, the reaction norms were investigated in very divergent laboratory media in order to enhance the phenotypic plasticity. Therefore, it is likely that the importance of antagonistic effects claimed is certainly biased by the drastic physiological conditions imposed.

Pleiotropic QTLs showing large effects with both beneficial and negative impacts on species adaptation constitute an interesting case of GxE interactions. The description of pleiotropic QTLs/genes/SNPs has been achieved in several studies for *IRA2* (Yadav et al., 2015), *MKT1* (Deutschbauer and Davis, 2005; Fay, 2013), *CYS4* (Kim et al., 2009), *AQY1* (Will et al., 2010), and *SSU1* (Peltier et al., 2018). In some cases, these loci showed antagonistic effects and could reflect balancing selection since the alleles involved are found in similar frequency in natural populations. The aquaporin genes (*AQY1* and *AQY2*) are a notorious case of alleles under balancing selection with a possible effect in a bakery context. Indeed, loss of aquaporin reduces freeze–thaw tolerance but increases fitness in high-sugar environments, two conditions encountered in the bread industry (Will et al., 2010). Another case is given by the translocation XV–t-XVI that influences the expression level of the sulfite pump *SSU1* (Zimmer et al., 2014). Although very beneficial for initiating the alcoholic fermentation in the presence of SO<sub>2</sub>, yeast strains presenting this translocation have a reduced fermentation rate in the late steps of wine fermentation (Peltier et al., 2018).

In agronomy, QTL programs take into account the environmental effect by achieving phenotyping in various conditions. The first benefit of this strategy is the gain of power detection of minor QTLs. The second one allows the detection of robust loci having an effect whatever the environment is, which is an obvious asset for achieving marker-assisted selection. In a recent work, QTL mapping was achieved in three natural grape juices using two independent F1-cross

derived from wine starters. The conditions applied represent extreme environmental conditions from an enological point of view (different grape juice colors, various amounts of sugar and nitrogen, and different micro-oxygenation conditions). In such conditions, up to 72% of the mapped QTLs have the same effect regardless of the environment. This observation suggests that when QTL mapping is carried out with genetic rootstock adapted to the specific industrial context, the G×E interactions are quite moderate. Therefore, most of the mapped QTLs would be robust in various enological environments (Peltier et al., 2018).

### Selecting an Appropriate Genetic Background

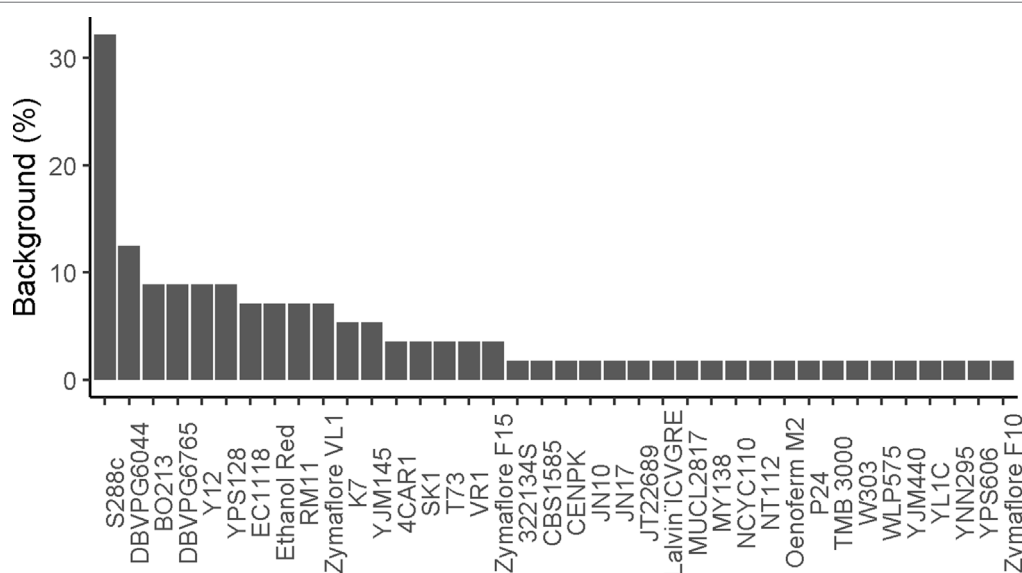
Until now, most QTNs were identified by using less than 20 yeast backgrounds (Figure 7). Usually, an elite strain with relevant features has been crossed with the laboratory background (S288c) or with one (or few) strains derived from another subpopulation. However, for achieving relevant research of biotechnological interest, the use of an “outgroup” partner may represent an important risk. This is particularly true for lab strains that have been cultivated for numerous generations in a controlled environment (growth on rich media at steady temperature). The use of various parental strains originated from distinct subpopulations (SGRP4X design) may have the same effect. Indeed, the NA, SA, WE, and WA strains belong to distinct clean lineages that have been evolved mostly by genetic drift in separated habitat. Although locally neutral in one environment, these alleles would have a deleterious effect on each other (Zörgö et al., 2013). Therefore, the selection of distant parental strains may enhance the chance to find QTNs with low expressivity by introducing a pool of not adapted alleles that *in fine* cause a drastic loss of fitness.

The cross of unrelated lineages also regroups a large pool of alleles that had never co-evolved together. This likely enhances epistatic interactions that have a negative impact on

QTL expressivity. Then, we believed that the introduction of deleterious alleles must be prevented as far as possible. A rational way consists in using a set of strains adapted to the environment of interest. Indeed, by using such adapted strains, the pool of alleles submitted to positive selection would increase, while those having deleterious impact would be reduced. In most of the studies, parental strains showed very opposite trait values, which is not an obligation. In theory, two conditions are required for achieving linkage analyses. The first is to capture enough genetic polymorphisms for building a fine grain genetic map. This even can be obtained within strains of the same subpopulations. The second is a wide phenotypic segregation that is not necessary depending on parental traits. Recently, we demonstrated that the phenotypic and genetic distance within pairs of parental strains does not affect neither the phenotypic segregation nor the efficiency of QTL detection (Peltier et al., 2018). By extrapolation, the cross of two optimal strains with a sufficient phenotypic diversity would have the benefit to fix many positive alleles and allow the segregation of other QTLs that would also contribute positively to the phenotype.

### CONCLUDING REMARKS

This review provides a first compendium of QTNs of biotechnological interest belonging to the *Saccharomyces cerevisiae* species. This emphasizes the success of quantitative yeast genetics for identifying relevant natural (or induced) genetic variations that confers technological properties. The SNPs reported here will constitute a rich reservoir of genetic variations useful for improving the technological properties of industrial strains by using breeding or genome editing strategies. However, for bridging the gap between the identification of causative SNP and their routine exploitation, the complex architecture of quantitative traits needs to be better understood.



**FIGURE 7 |** Proportion of genetic background used for QTL mapping studies. The percentage of time a genetic background was used in reported studies.



## DATA AVAILABILITY

The datasets analyzed for this study can be found in Peter et al. (2018).

## AUTHOR CONTRIBUTIONS

EP and PM wrote the first draft of the manuscript. AF and JS wrote sections of the manuscript and reviewed and edited the original draft.

## REFERENCES

- Ajikumar, P. K., Tyo, K., Carlsen, S., Mucha, O., Phon, T. H., and Stephanopoulos, G. (2008). Terpenoids: opportunities for biosynthesis of natural product drugs using engineered microorganisms. *Mol. Pharm.* 5, 167–190. doi: 10.1021/mp700151b
- Akada, R., Matsuo, K., Aritomi, K., and Nishizawa, Y. (1999). Construction of recombinant sake yeast containing a dominant FAS2 mutation without extraneous sequences by a two-step gene replacement protocol. *J. Biosci. Bioeng.* 87, 43–48. doi: 10.1016/S1389-1723(99)80006-1
- Albertin, W., Marullo, P., Aigle, M., Dillmann, C., de Vienne, D., Bely, M., et al. (2011). Population size drives industrial *Saccharomyces cerevisiae* alcoholic fermentation and is under genetic control. *Appl. Environ. Microbiol.* 77, 2772–2784. doi: 10.1128/AEM.02547-10
- Almeida, P., Gonçalves, C., Teixeira, S., Libkind, D., Bontrager, M., Masneuf-Pomarède, I., et al. (2014). A Gondwanan imprint on global diversity and domestication of wine and cider yeast *Saccharomyces uvarum*. *Nat. Commun.* 5, 4044. doi: 10.1038/ncomms5044
- Alper, H., Moxley, J., Nevoigt, E., Fink, G. R., and Stephanopoulos, G. (2006). Engineering yeast transcription machinery for improved ethanol tolerance and production. *Science* 314, 1565–1568. doi: 10.1126/science.1131969
- Ambroset, C., Petit, M., Brion, C., Sanchez, I., Delobel, P., Guerin, C., et al. (2011). Deciphering the molecular basis of wine yeast fermentation traits using a combined genetic and genomic approach. *G3 (Bethesda)* 1, 263–281. doi: 10.1534/g3.111.000422
- Aritomi, K., Hirokawa, I., Hoshida, H., Shigii, M., Nishizawa, Y., Susumu, K., et al. (2004). Self-cloning yeast strains containing novel FAS2 mutations produce a higher amount of ethyl caproate in Japanese sake. *Biosci. Biotechnol. Biochem.* 68, 206–214. doi: 10.1271/bbb.68.206
- Avrahami-Moyal, L., Engelberg, D., Wenger, J. W., Sherlock, G., and Braun, S. (2012). Turbidostat culture of *Saccharomyces cerevisiae* W303-1A under selective pressure elicited by ethanol selects for mutations in SSD1 and UTH1. *FEMS Yeast Res.* 12, 521–533. doi: 10.1111/j.1567-1364.2012.00803.x
- Barbosa, R., Pontes, A., Santos, R. O., Montandon, G. G., de Ponzzes-Gomes, C. M., Morais, P. B., et al. (2018). Multiple rounds of artificial selection promote microbe secondary domestication—the case of Cachaça yeasts. *Genome Biol. Evol.* 10, 1939–1955. doi: 10.1093/gbe/evy132
- Bauer, F. F., and Pretorius, I. S. (2000). Yeast stress response and fermentation efficiency: how to survive the making of wine—a review. *S Afr. J. Enol. Vitic.* 21, 27–51. doi: 10.21548/21-1-3557
- Beck, T., Hastings, R. K., Gollapudi, S., Free, R. C., and Brookes, A. J. (2014). GWAS Central: a comprehensive resource for the comparison and interrogation of genome-wide association studies. *Eur. J. Hum. Genet.* 22, 949–952. doi: 10.1038/ejhg.2013.274
- Bell, S.-J., and Henschke, P. A. (2005). Implications of nitrogen nutrition for grapes, fermentation and wine. *Aust. J. Grape Wine Res.* 11, 242–295. doi: 10.1111/j.1755-0238.2005.tb00028.x
- Bely, M., Sablayrolles, J. M., and Barre, P. (1990). Description of alcoholic fermentation kinetics: its variability and significance. *Am. J. Enol. Vitic.* 41, 319–324.
- Benjaphokee, S., Koedrit, P., Auesukaree, C., Asvarak, T., Sugiyama, M., Kaneko, Y., et al. (2012). CDC19 encoding pyruvate kinase is important for high-temperature tolerance in *Saccharomyces cerevisiae*. *New Biotechnol.* 29, 166–176. doi: 10.1016/j.nbt.2011.03.007

## SUPPLEMENTARY MATERIALS

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2019.00683/full#supplementary-material>.

**TABLE S1** | Description of the QTGs reported.

**TABLE S2** | Allelic frequencies of the QTNs reported.

**TABLE S3** | Distribution of the allelic frequencies among isolate origins.

- Berthels, N. J., Cordero Otero, R. R., Bauer, F. F., Thevelein, J. M., and Pretorius, I. S. (2004). Discrepancy in glucose and fructose utilisation during fermentation by *Saccharomyces cerevisiae* wine yeast strains. *FEMS Yeast Res.* 4, 683–689. doi: 10.1016/j.femsyr.2004.02.005
- Bhatia, A., Yadav, A., Zhu, C., Gagneur, J., Radhakrishnan, A., Steinmetz, L. M., et al. (2014). Yeast growth plasticity is regulated by environment-specific multi-QTL interactions. *G3 (Bethesda)* 4, 769–777. doi: 10.1534/g3.113.009142
- Blondin, B., NOBLE, J., Sanchez, I., and Agronomique, I.N.D.L.R., and Supagro, C.I.D.S.E.S.A. (Montpellier) (2013). Méthode de contrôle de la production de sulfites, d'hydrogène sulfureux et d'acétaldéhyde par des levures.
- Blondin, B., NOBLE, J., and Sanchez, I. (2014). Méthode de contrôle de la production de sulfites, d'hydrogène sulfureux et d'acétaldéhyde par des levures (Google Patents).
- Bloom, J. S., Boocock, J., Treusch, S., Sadhu, M. J., Day, L., Oates-Barker, H., et al. (2019). Rare variants contribute disproportionately to quantitative trait variation in yeast. *BioRxiv*, 607291. doi: 10.1101/607291
- Boaz, L., M., A. J. R., Peter, R., Bärbel, H.-H., and Marie, G.-G. (2008). Identification of an NADH-dependent 5-hydroxymethylfurfural-reducing alcohol dehydrogenase in *Saccharomyces cerevisiae*. *Yeast* 25, 191–198. doi: 10.1002/yea.1578
- Borneman, A. R., Forgan, A. H., Kolouchova, R., Fraser, J. A., and Schmidt, S. A. (2016). Whole genome comparison reveals high levels of inbreeding and strain redundancy across the spectrum of commercial wine strains of *Saccharomyces cerevisiae*. *G3 (Bethesda)* 6, 957–971. doi: 10.1534/g3.115.025692
- Brachi, B., Morris, G. P., and Borevitz, J. O. (2011). Genome-wide association studies in plants: the missing heritability is in the field. *Genome Biol.* 12, 232. doi: 10.1186/gb-2011-12-10-232
- Brem, R. B., Storey, J. D., Whittle, J., and Kruglyak, L. (2005). Genetic interactions between polymorphisms that affect gene expression in yeast. *Nature* 436, 701–703. doi: 10.1038/nature03865
- Brice, C., Sanchez, I., Bigey, F., Legras, J.-L., and Blondin, B. (2014). A genetic approach of wine yeast fermentation capacity in nitrogen-starvation reveals the key role of nitrogen signaling. *BMC Genomics* 15, 495. doi: 10.1186/1471-2164-15-495
- Brion, C., Ambroset, C., Sanchez, I., Legras, J.-L., and Blondin, B. (2013). Differential adaptation to multi-stressed conditions of wine fermentation revealed by variations in yeast regulatory networks. *BMC Genomics* 14, 681. doi: 10.1186/1471-2164-14-681
- Bro, C., Knudsen, S., Regenber, B., Olsson, L., and Nielsen, J. (2005). Improvement of galactose uptake in *Saccharomyces cerevisiae* through overexpression of phosphoglucosyltransferase: example of transcript analysis as a tool in inverse metabolic engineering. *Appl. Environ. Microbiol.* 71, 6465–6472. doi: 10.1128/AEM.71.11.6465-6472.2005
- Brown, C. A., Murray, A. W., and Verstrepen, K. J. (2010). Rapid expansion and functional divergence of subtelomeric gene families in yeasts. *Curr. Biol.* 20, 895–903. doi: 10.1016/j.cub.2010.04.027
- Carlson, M., and Botstein, D. (1983). Organization of the SUC gene family in *Saccharomyces*. *Mol. Cell. Biol.* 3, 351–359. doi: 10.1128/MCB.3.3.351
- Chambon, C., Ladeveze, V., Servouse, M., Blanchard, L., Javelot, C., Vladescu, B., et al. (1991). Sterol pathway in yeast. Identification and properties of mutant strains defective in mevalonate diphosphate decarboxylase and farnesyl diphosphate synthetase. *Lipids* 26, 633–636. doi: 10.1007/BF02536428
- Chen, X., Zhao, F., and Xu, S. (2010). Mapping environment-specific quantitative trait loci. *Genetics* 186, 1053–1066. doi: 10.1534/genetics.110.120311

- Ching, A., Caldwell, K. S., Jung, M., Dolan, M., Smith, O. S., Tingey, S., et al. (2002). SNP frequency, haplotype structure and linkage disequilibrium in elite maize inbred lines. *BMC Genet.* 3, 19. doi: 10.1186/1471-2156-3-19
- Clausen, M., Lamb, C. J., Megnet, R., and Doerner, P. W. (1994). PAD1 encodes phenylacrylic acid decarboxylase which confers resistance to cinnamic acid in *Saccharomyces cerevisiae*. *Gene* 142, 107–112. doi: 10.1016/0378-1119(94)90363-8
- Clomburg, J. M., and Gonzalez, R. (2013). Anaerobic fermentation of glycerol: a platform for renewable fuels and chemicals. *Trends Biotechnol.* 31, 20–28. doi: 10.1016/j.tibtech.2012.10.006
- Coetzee, C., and du Toit, W. J. (2012). A comprehensive review on Sauvignon blanc aroma with a focus on certain positive volatile thiols. *Food Res. Int.* 45, 287–298. doi: 10.1016/j.foodres.2011.09.017
- Cordente, A. G., Heinrich, A., Pretorius, I. S., and Swiegers, J. H. (2009). Isolation of sulfite reductase variants of a commercial wine yeast with significantly reduced hydrogen sulfide production. *FEMS Yeast Res.* 9, 446–459. doi: 10.1111/j.1567-1364.2009.00489.x
- Cordente, A. G., Cordero-Bueso, G., Pretorius, I. S., and Curtin, C. D. (2013). Novel wine yeast with mutations in YAP1 that produce less acetic acid during fermentation. *FEMS Yeast Res.* 13, 62–73. doi: 10.1111/1567-1364.12010
- Cubillos, F. A. (2016). Exploiting budding yeast natural variation for industrial processes. *Curr. Genet.* 62, 745–751. doi: 10.1007/s00294-016-0602-6
- Cubillos, F. A., Parts, L., Salinas, F., Bergström, A., Scovacicchi, E., Zia, A., et al. (2013). High-resolution mapping of complex traits with a four-parent advanced intercross yeast population. *Genetics* 195, 1141–1155. doi: 10.1534/genetics.113.155515
- Cubillos, F. A., Brice, C., Molinet, J., Tisné, S., Abarca, V., Tapia, S. M., et al. (2017). Identification of nitrogen consumption genetic variants in yeast through QTL mapping and bulk segregant RNA-Seq analyses. *G3 (Bethesda)* 7, 1693–1705. doi: 10.1534/g3.117.042127
- Demogines, A., Smith, E., Kruglyak, L., and Alani, E. (2008). Identification and dissection of a complex DNA repair sensitivity phenotype in baker's yeast. *PLoS Genet.* 4, e1000123. doi: 10.1371/journal.pgen.1000123
- Denis, C. L., and Young, E. T. (1983). Isolation and characterization of the positive regulatory gene ADR1 from *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* 3, 360–370. doi: 10.1128/MCB.3.3.360
- Dequin, S., Escudier, J.-L., Bely, M., Noble, J., Albertin, W., Masneuf-Pomarède, I., et al. (2017). How to adapt winemaking practices to modified grape composition under climate change conditions. *OENO One* 51, 205. doi: 10.20870/oeno-one.2017.51.2.1584
- Deutschbauer, A. M., and Davis, R. W. (2005). Quantitative trait loci mapped to single-nucleotide resolution in yeast. *Nat. Genet.* 37, 1333–1340. doi: 10.1038/ng1674
- Dhar, R., Sägeser, R., Weikert, C., Yuan, J., and Wagner, A. (2011). Adaptation of *Saccharomyces cerevisiae* to saline stress through laboratory evolution. *J. Evol. Biol.* 24, 1135–1153. doi: 10.1111/j.1420-9101.2011.02249.x
- Diezmann, S., and Dietrich, F. S. (2011). Oxidative stress survival in a clinical *Saccharomyces cerevisiae* isolate is influenced by a major quantitative trait nucleotide. *Genetics* 188, 709–722. doi: 10.1534/genetics.111.128256
- Doniger, S. W., Kim, H. S., Swain, D., Corcuera, D., Williams, M., Yang, S. P., et al. (2008). A catalog of neutral and deleterious polymorphism in yeast. *PLoS Genet.* 4, e1000183. doi: 10.1371/journal.pgen.1000183
- Dufour, M., Zimmer, A., Thibon, C., and Marullo, P. (2013). Enhancement of volatile thiol release of *Saccharomyces cerevisiae* strains using molecular breeding. *Appl. Microbiol. Biotechnol.* 97, 5893–5905. doi: 10.1007/s00253-013-4739-7
- Duitama, J., Sánchez-Rodríguez, A., Goovaerts, A., Pulido-Tamayo, S., Hubmann, G., Foulquié-Moreno, M. R., et al. (2014). Improved linkage analysis of quantitative trait loci using bulk segregants unveils a novel determinant of high ethanol tolerance in yeast. *BMC Genomics* 15, 207. doi: 10.1186/1471-2164-15-207
- Eder, M., Sanchez, I., Brice, C., Camarasa, C., Legras, J.-L., and Dequin, S. (2018). QTL mapping of volatile compound production in *Saccharomyces cerevisiae* during alcoholic fermentation. *BMC Genomics* 19, 166. doi: 10.1186/s12864-018-4562-8
- Fassler, J. S., Gray, W. M., Malone, C. L., Tao, W., Lin, H., and Deschenes, R. J. (1997). Activated alleles of yeast SLN1 increase Mcm1-dependent reporter gene expression and diminish signaling through the Hog1 osmosensing pathway. *J. Biol. Chem.* 272, 13365–13371. doi: 10.1074/jbc.272.20.13365
- Fay, J. C. (2013). The molecular basis of phenotypic variation in yeast. *Curr. Opin. Genet. Dev.* 23, 672–677. doi: 10.1016/j.gde.2013.10.005
- Fay, J. C., McCullough, H. L., Sniegowski, P. D., and Eisen, M. B. (2004). Population genetic variation in gene expression is associated with phenotypic variation in *Saccharomyces cerevisiae*. *Genome Biol.* 5, R26. doi: 10.1186/gb-2004-5-4-r26
- Fidalgo, M., Barrales, R. R., Ibeas, J. I., and Jimenez, J. (2006). Adaptive evolution by mutations in the FLO11 gene. *Proc. Natl. Acad. Sci. U.S.A.* 103, 11228–11233. doi: 10.1073/pnas.0601713103
- Fournier, T., Saada, O. A., Hou, J., Peter, J., Caudal, E., and Schacherer, J. (2019). Extensive impact of low-frequency variants on the phenotypic landscape at population-scale. *BioRxiv*, 609917. doi: 10.1101/609917
- Gagneur, J., Sinha, H., Perocchi, F., Bourgon, R., Huber, W., and Steinmetz, L. M. (2009). Genome-wide allele- and strand-specific expression profiling. *Mol. Syst. Biol.* 5, 274. doi: 10.1038/msb.2009.31
- Gallone, B., Steensels, J., Prah, T., Soriaga, L., Saels, V., Herrera-Malaver, B., et al. (2016). Domestication and divergence of *Saccharomyces cerevisiae* beer yeasts. *Cell* 166, 1397–1410.e16. doi: 10.1016/j.cell.2016.08.020
- Gancedo, J. M. (1998). Yeast carbon catabolite repression. *Microbiol. Mol. Biol. Rev.* 62, 334–361.
- García-Ríos, E., López-Malo, M., and Guillamón, J. (2014). Global phenotypic and genomic comparison of two *Saccharomyces cerevisiae* wine strains reveals a novel role of the sulfur assimilation pathway in adaptation at low temperature fermentations. *BMC Genomics* 15, 1059. doi: 10.1186/1471-2164-15-1059
- García-Ríos, E., Morard, M., Parts, L., Liti, G., and Guillamón, J. M. (2017). The genetic architecture of low-temperature adaptation in the wine yeast *Saccharomyces cerevisiae*. *BMC Genomics* 18, 159. doi: 10.1186/s12864-017-3572-2
- García-Ríos, E., Nuévalos, M., Barrio, E., Puig, S., and Guillamón, J. M. (2019). A new chromosomal rearrangement improves the adaptation of wine yeasts to sulfite. *Environ. Microbiol.* 21, 1771–1781. doi: 10.1111/1462-2920.14586
- Gasch, A. P. (2003). “The environmental stress response: a common yeast response to diverse environmental stresses,” in *Yeast stress responses*. Eds. S. Hohmann and W. H. Mager Berlin, Heidelberg: Springer, 11–57. doi: 10.1007/3-540-45611-2\_2
- Gibson, B. R., Lawrence, S. J., Leclaire, J. P. R., Powell, C. D., and Smart, K. A. (2007). Yeast responses to stresses associated with industrial brewery handling. *FEMS Microbiol. Rev.* 31, 535–569. doi: 10.1111/j.1574-6976.2007.00076.x
- Giudici, P., and Zambonelli, C. (1992). Biometric and genetic study on acetic acid production for breeding of wine yeast. *Am. J. Enol. Vitic.* 43, 370–374.
- González-Ramos, D., Gorter de Vries, A. R., Grijseels, S. S., van Berkum, M. C., Swinnen, S., van den Broek, M., et al. (2016). A new laboratory evolution approach to select for constitutive acetic acid tolerance in *Saccharomyces cerevisiae* and identification of causal mutations. *Biotechnol. Biofuels* 9, 173. doi: 10.1186/s13068-016-0583-1
- Grant, D., Nelson, R. T., Cannon, S. B., and Shoemaker, R. C. (2010). SoyBase, the USDA-ARS soybean genetics and genomics database. *Nucleic Acids Res.* 38, D843–D846. doi: 10.1093/nar/gkp798
- Grethart, D., Wimalasena, T. T., Leung, K., Marvin, M. E., Chandelia, Y., Hart, A. J., et al. (2014). The genetic basis of variation in clean lineages of *Saccharomyces cerevisiae* in response to stresses encountered during bioethanol fermentations. *PLoS One* 9, e103233. doi: 10.1371/journal.pone.0103233
- Gresham, D., and Hong, J. (2014). The functional basis of adaptive evolution in chemostats. *FEMS Microbiol. Rev.* 39, 1–19. doi: 10.1111/1574-6976.12082
- Griffiths, A. J., Miller, J. H., Suzuki, D. T., Lewontin, R. C., Gelbart, W. M., Griffiths, A. J., et al. (2000). *An introduction to genetic analysis*. (San Francisco, CA: W. H. Freeman).
- Guillaume, C., Delobel, P., Sablayrolles, J. M., and Blondin, B. (2007). Molecular basis of fructose utilization by the wine yeast *Saccharomyces cerevisiae*: a mutated HXT3 allele enhances fructose fermentation. *Appl. Environ. Microbiol.* 73, 2432–2439. doi: 10.1128/AEM.02269-06
- Gunde-Cimerman, N., Plementitis, A., and Buzzini, P. (2014). “Cold-adapted yeasts,” in *Cold-adapted yeasts*. Eds. P. Buzzini and R. Margesin (Berlin, Heidelberg: Springer Berlin Heidelberg), 225–242. doi: 10.1007/978-3-642-39681-6\_10
- Gutiérrez, A., Beltran, G., Warringer, J., and Guillamón, J. M. (2013). Genetic basis of variations in nitrogen source utilization in four wine commercial yeast strains. *PLoS One* 8, e67166. doi: 10.1371/journal.pone.0067166
- Hahn-Hägerdal, B., Karhumaa, K., Larsson, C. U., Gorwa-Grauslund, M., Görgens, J., and van Zyl, W. H. (2005). Role of cultivation media in the development

- of yeast strains for large scale industrial use. *Microb. Cell Fact.* 4, 31. doi: 10.1186/1475-2859-4-31
- Hasunuma, T., Sanda, T., Yamada, R., Yoshimura, K., Ishii, J., and Kondo, A. (2011). Metabolic pathway engineering based on metabolomics confers acetic and formic acid tolerance to a recombinant xylose-fermenting strain of *Saccharomyces cerevisiae*. *Microb. Cell Fact.* 10, 2. doi: 10.1186/1475-2859-10-2
- Higgins, V. J., Braidwood, M., Bissinger, P., Dawes, I. W., and Attfield, P. V. (1999). Leu343Phe substitution in the Malx3 protein of *Saccharomyces cerevisiae* increases the constitutivity and glucose insensitivity of MAL gene expression. *Curr. Genet.* 35, 491–498. doi: 10.1007/s002940050444
- Hohmann, S. (2002). Osmotic stress signaling and osmoadaptation in yeasts osmotic. *Microbiol. Mol. Biol. Rev.* 66, 300–372. doi: 10.1128/MMBR.66.2.300-372.2002
- Holt, S., Carvalho, B. T., Foulquié-Moreno, M. R., and Thevelein, J. M. (2018). Polygenic analysis in absence of major effector ATF1 unveils novel components in yeast flavor ester biosynthesis. *MBio* 9, e01279–18. doi: 10.1128/mBio.01279-18
- Hong, J., and Gresham, D. (2014). Molecular specificity, convergence and constraint shape adaptive evolution in nutrient-poor environments. *PLoS Genet.* 10, e1004041. doi: 10.1371/journal.pgen.1004041
- Hou, J., Sigwalt, A., Fournier, T., Pflieger, D., Peter, J., de Montigny, J., et al. (2016). The hidden complexity of Mendelian traits across natural yeast populations. *Cell Rep.* 16, 1106–1114. doi: 10.1016/j.celrep.2016.06.048
- Huang, C., Roncoroni, M., and Gardner, R. C. (2014). MET2 affects production of hydrogen sulfide during wine fermentation. *Appl. Microbiol. Biotechnol.* 98, 7125–7135. doi: 10.1007/s00253-014-5789-1
- Hubmann, G., Mathé, L., Foulquié-Moreno, M. R., Duitama, J., Nevoigt, E., and Thevelein, J. M. (2013a). Identification of multiple interacting alleles conferring low glycerol and high ethanol yield in *Saccharomyces cerevisiae* ethanolic fermentation. *Biotechnol. Biofuels* 6, 87. doi: 10.1186/1754-6834-6-87
- Hubmann, G., Foulquié-Moreno, M. R., Nevoigt, E., Duitama, J., Meurens, N., Pais, T. M., et al. (2013b). Quantitative trait analysis of yeast biodiversity yields novel gene tools for metabolic engineering. *Metab. Eng.* 17, 68–81. doi: 10.1016/j.ymben.2013.02.006
- Ibstedt, S., Stenberg, S., Bages, S., Gjuvslund, A. B., Salinas, F., Kourtchenko, O., et al. (2014). Concerted evolution of life stage performances signals recent selection on yeast nitrogen use. *Mol. Biol. Evol.* 32, 153–161. doi: 10.1093/molbev/msu285
- Inokoshi, J., Tomoda, H., Hashimoto, H., Watanabe, A., Takeshima, H., and Omura, S. (1994). Cerulenin-resistant mutants of *Saccharomyces cerevisiae* with an altered fatty acid synthase gene. *Mol. Gen. Genet.* 244, 90–96. doi: 10.1007/BF00280191
- Jara, M., Cubillos, F., García, V., Salinas, F., Aguilera, O., Liti, G., et al. (2014). Mapping genetic variants underlying differences in the central nitrogen metabolism in fermenter yeasts. *PLoS One* 9, e86533. doi: 10.1371/journal.pone.0086533
- Javelot, C., Girard, P., Colonna-Ceccaldi, B., and Vladescu, B. (1991). Introduction of terpene-producing ability in a wine strain of *Saccharomyces cerevisiae*. *J. Biotechnol.* 21, 239–251. doi: 10.1016/0168-1656(91)90045-W
- Jiranek, V., Langridge, P., and Henschke, P. A. (1995). Amino acid and ammonium utilization by *Saccharomyces cerevisiae* wine yeasts from a chemically defined medium. *Am. J. Enol. Vitic.* 46, 75–83.
- Kaino, T., Tateiwa, T., Mizukami-Murata, S., Shima, J., and Takagi, H. (2008). Self-cloning baker's yeasts that accumulate proline enhance freeze tolerance in doughs. *Appl. Environ. Microbiol.* 74, 5845–5849. doi: 10.1128/AEM.00998-08
- Kim, H. S., Huh, J., and Fay, J. C. (2009). Dissecting the pleiotropic consequences of a quantitative trait nucleotide. *FEMS Yeast Res.* 9, 713–722. doi: 10.1111/j.1567-1364.2009.00516.x
- Kim, N.-R., Yang, J., Kwon, H., An, J., Choi, W., and Kim, W. (2013). Mutations of the TATA-binding protein confer enhanced tolerance to hyperosmotic stress in *Saccharomyces cerevisiae*. *Appl. Microbiol. Biotechnol.* 97, 8227–8238. doi: 10.1007/s00253-013-4985-8
- Kitagaki, H., and Takagi, H. (2014). Mitochondrial metabolism and stress response of yeast: applications in fermentation technologies. *J. Biosci. Bioeng.* 117 (4), 383–393. doi: 10.1016/j.jbiosc.2013.09.011
- Klinke, H. B., Thomsen, A. B., and Ahring, B. K. (2004). Inhibition of ethanol-producing yeast and bacteria by degradation products produced during pre-treatment of biomass. *Appl. Microbiol. Biotechnol.* 66, 10–26. doi: 10.1007/s00253-004-1642-2
- Kutyna, D. R., Varela, C., Henschke, P., Chambers, P. J., and Stanley, G. A. (2010). Microbiological approaches to lowering ethanol concentration in wine. *Trends Food Sci. Technol.* 21, 293–302. doi: 10.1016/j.tifs.2010.03.004
- Kvitek, D. J., and Sherlock, G. (2011). Reciprocal sign epistasis between frequently experimentally evolved adaptive mutations causes a rugged fitness landscape. *PLoS Genet.* 7, e1002056. doi: 10.1371/journal.pgen.1002056
- Kwan, E. X., Foss, E., Kruglyak, L., and Bedalov, A. (2011). Natural polymorphism in BUL2 links cellular amino acid availability with chronological aging and telomere maintenance in yeast. *PLoS Genet.* 7, e1002250. doi: 10.1371/journal.pgen.1002250
- Lee, S.-I., Dudley, A. M., Drubin, D., Silver, P. A., Krogan, N. J., Pe'er, D., et al. (2009). Learning a prior on regulatory potential from eQTL data. *PLoS Genet.* 5, e1000358. doi: 10.1371/journal.pgen.1000358
- Legras, J.-L., Galeote, V., Bigey, F., Camarasa, C., Marsit, S., Nidelet, T., et al. (2018). Adaptation of *S. cerevisiae* to fermented food environments reveals remarkable genome plasticity and the footprints of domestication. *Mol. Biol. Evol.* 35, 1712–1727. doi: 10.1093/molbev/msy066
- Li, J., Wang, L., Wu, X., Fang, O., Wang, L., Lu, C., et al. (2013). Polygenic molecular architecture underlying non-sexual cell aggregation in budding yeast. *DNA Res.* 20, 55–66. doi: 10.1093/dnares/dss033
- Linderholm, A., Dietzel, K., Hirst, M., and Bisson, L. F. (2010). Identification of MET10-932 and characterization as an allele reducing hydrogen sulfide formation in wine strains of *Saccharomyces cerevisiae*. *Appl. Environ. Microbiol.* 76, 7699–7707. doi: 10.1128/AEM.01666-10
- Linderholm, A. L., Olineka, T. L., Hong, Y., and Bisson, L. F. (2006). Allele diversity among genes of the sulfate reduction pathway in wine strains of *Saccharomyces cerevisiae*. *Am. J. Enol. Vitic.* 57, 431–440.
- Liti, G., and Louis, E. J. (2012). Advances in quantitative trait analysis in yeast. *PLoS Genet.* 8, e1002912. doi: 10.1371/journal.pgen.1002912
- Liti, G., Carter, D. M., Moses, A. M., Warringer, J., Parts, L., James, S. A., et al. (2009). Population genomics of domestic and wild yeasts. *Nature* 458, 337–341. doi: 10.1038/nature07743
- Ma, M., and Liu, Z. L. (2010). Mechanisms of ethanol tolerance in *Saccharomyces cerevisiae*. *Appl. Microbiol. Biotechnol.* 87, 829–845. doi: 10.1007/s00253-010-2594-3
- Malcorps, P., Cheval, J. M., Jamil, S., and Dufour, J.-P. (1991). A new model for the regulation of ester synthesis by alcohol acetyltransferases in *Saccharomyces cerevisiae*. *J. Am. Soc. Brew. Chem.* 49, 47–53. doi: 10.1094/ASBCJ-49-0047
- Manginot, C., Roustan, J., and Sablayrolles, J. (1998). Nitrogen demand of different yeast strains during alcoholic fermentation. Importance of the stationary phase. *Enzyme Microb. Technol.* 23, 511–517. doi: 10.1016/S0141-0229(98)00080-5
- Marsit, S., Leducq, J.-B., Durand, É., Marchant, A., Filteau, M., and Landry, C. R. (2017). Evolutionary biology through the lens of budding yeast comparative genomics. *Nat. Rev. Genet.* 18, 581–598. doi: 10.1038/nrg.2017.49
- Martí-Raga, M., Sancho, M., Guzmán, J. M., Mas, A., and Beltrán, G. (2015). The effect of nitrogen addition on the fermentative performance during sparkling wine production. *Food Res. Int.* 67, 126–135. doi: 10.1016/j.foodres.2014.10.033
- Martí-Raga, M., Peltier, E., Mas, A., Beltrán, G., and Marullo, P. (2017). Genetic causes of phenotypic adaptation to the second fermentation of sparkling wines in *Saccharomyces cerevisiae*. *G3 (Bethesda)*. g3.116.037283. 7, 399–412. doi: 10.1534/g3.116.037283
- Marullo, P. (2009). Procédé de sélection de levures oenologiques (FRANCE: Google Patents).
- Marullo, P., and Dubourdieu, D. (2010). “Yeast selection for wine flavour modulation,” in *Managing wine quality*, vol. 2. Ed. A. G. Reynolds (Cambridge: Woodhead Publishing Limited), 296–320. doi: 10.1533/9781845699987.2.293
- Marullo, P., Bely, M., Masneuf-Pomarede, I., Aigle, M., and Dubourdieu, D. (2004). Inheritable nature of enological quantitative traits is demonstrated by meiotic segregation of industrial wine yeast strains. *FEMS Yeast Res.* 4, 711–719. doi: 10.1016/j.femsyr.2004.01.006
- Marullo, P., Aigle, M., Bely, M., Masneuf-Pomarede, I., Durrrens, P., Dubourdieu, D., et al. (2007a). Single QTL mapping and nucleotide-level resolution of a physiologic trait in wine *Saccharomyces cerevisiae* strains. *FEMS Yeast Res.* 7, 941–952. doi: 10.1111/j.1567-1364.2007.00252.x



- Marullo, P., Yvert, G., Bely, M., Aigle, M., and Dubourdieu, D. (2007b). Efficient use of DNA molecular markers to construct industrial yeast strains. *FEMS Yeast Res.* 7, 1295–1306. doi: 10.1111/j.1567-1364.2007.00281.x
- Marullo, P., Thibon, C., Tominaga, T., and Dubourdieu, D. (2008). Utilisation de levures mutantes ure2 pour augmenter la libération de thiols volatils aromatiques par la levure pendant la fermentation (Google Patents).
- Marullo, P., Mansour, C., Dufour, M., Albertin, W., Sicard, D., Bely, M., et al. (2009). Genetic improvement of thermo-tolerance in wine *Saccharomyces cerevisiae* strains by a backcross approach. *FEMS Yeast Res.* 9, 1148–1160. doi: 10.1111/j.1567-1364.2009.00550.x
- Marullo, P., Durrens, P., Peltier, E., Bernard, M., Mansour, C., Dubourdieu, D., et al. (2019). Natural allelic variations of *Saccharomyces cerevisiae* impact stuck fermentation due to the combined effect of ethanol and temperature; a QTL-mapping study. *BioRxiv*. doi: 10.1101/576835
- Mason, A. B., Dufour, J., and Dufour, J. (2000). Alcohol acetyltransferases and the significance of ester synthesis in yeast. *Yeast*, 16, 1287–1298. doi: 10.1002/1097-0061(200010)16:14<1287::AID-YEA613>3.0.CO;2-I
- Maurer, M. J., Sutardja, L., Pinel, D., Bauer, S., Muehlbauer, A. L., Ames, T. D., et al. (2017). Quantitative trait loci (QTL)-guided metabolic engineering of a complex trait. *ACS Synth. Biol.* 6, 566–581. doi: 10.1021/acssynbio.6b00264
- Meijnen, J.-P., Randazzo, P., Foulquié-Moreno, M. R., van den Brink, J., Vandecruys, P., Stojiljkovic, M., et al. (2016). Polygenic analysis and targeted improvement of the complex trait of high acetic acid tolerance in the yeast *Saccharomyces cerevisiae*. *Biotechnol. Biofuels* 9, 5. doi: 10.1186/s13068-015-0421-x
- Mertens, S., Steensels, J., Gallone, B., Souffriau, B., Malcorps, P., and Verstrepen, K. J. (2017). Rapid screening method for phenolic off-flavor (POF) production in yeast. *J. Am. Soc. Brew. Chem.* 75, 318–323. doi: 10.1094/ASBCJ-2017-4142-01
- Mitsumasa, K., Liu, Z.-S., Tang, Y.-Q., Akamatsu, T., Taguchi, H., and Kida, K. (2014). Development of industrial yeast strain with improved acid- and thermo-tolerance through evolution under continuous fermentation conditions followed by haploidization and mating. *J. Biosci. Bioeng.* 118, 689–695. doi: 10.1016/j.jbiosc.2014.05.012
- Mizoguchi, H., Yamauchi, T., Watanabe, M., Yamanaka, H., Nishimura, A., and Hanamoto, H. (2002). Different missense mutations in PDR1 and PDR3 genes from clotrimazole-resistant sake yeast are responsible for pleiotropic drug resistance and improved fermentative activity. *J. Biosci. Bioeng.* 93, 221–227. doi: 10.1016/S1389-1723(02)80018-4
- Mukai, N., Masaki, K., Fujii, T., Kawamukai, M., and Iefuji, H. (2010). PAD1 and FDC1 are essential for the decarboxylation of phenylacrylic acids in *Saccharomyces cerevisiae*. *J. Biosci. Bioeng.* 109, 564–569. doi: 10.1016/j.jbiosc.2009.11.011
- Ness, F., and Aigle, M. (1995). RTM1: a member of a new family of telomeric repeated genes in yeast. *Genetics* 140, 945–956.
- Ng, P. C., and Henikoff, S. (2003). SIFT: predicting amino acid changes that affect protein function. *Nucleic Acids Res.* 31, 3812–3814. doi: 10.1093/nar/gkg509
- Noble, J., Sanchez, I., and Blondin, B. (2015). Identification of new *Saccharomyces cerevisiae* variants of the MET2 and SKP2 genes controlling the sulfur assimilation pathway and the production of undesirable sulfur compounds during alcoholic fermentation. *Microb. Cell Fact.* 14, 68. doi: 10.1186/s12934-015-0245-1
- Nogami, S., Ohya, Y., and Yvert, G. (2007). Genetic complexity and quantitative trait loci mapping of yeast morphological traits. *PLoS Genet.* 3, e31. doi: 10.1371/journal.pgen.0030031
- Noguchi, C., Watanabe, D., Zhou, Y., Akao, T., and Shimoi, H. (2012). Association of constitutive hyperphosphorylation of Hsf1p with a defective ethanol stress response in *Saccharomyces cerevisiae* sake yeast strains. *Appl. Environ. Microbiol.* 78, 385–392. doi: 10.1128/AEM.06341-11
- Ohnuki, S., Okada, H., Friedrich, A., Kanno, Y., Goshima, T., Hasuda, H., et al. (2017). Phenotypic diagnosis of lineage and differentiation during sake yeast breeding. *G3 (Bethesda)* 7, 2807–2820. doi: 10.1534/g3.117.044099
- Olofsson, K., Bertilsson, M., Lidén, G., Saidi, B., Braun, R., Jönsson, L., et al. (2008). A short review on SSF: An interesting process option for ethanol production from lignocellulosic feedstocks. *Biotechnol. Biofuels* 1, 7. doi: 10.1186/1754-6834-1-7
- Omura, F., Fujita, A., Miyajima, K., and Fukui, N. (2005). Engineering of yeast Put4 permease and its application to lager yeast for efficient proline assimilation. *Biosci. Biotechnol. Biochem.* 69, 1162–1171. doi: 10.1271/bbb.69.1162
- Ossa, E. M. D., Pérez, L., and Caro, I. (1987). Variations of the major volatiles through aging of sherry. *Am. J. Enol. Vitic.* 38, 293–297.
- Pais, T. M., Foulquié-Moreno, M. R., Hubmann, G., Duitama, J., Swinnen, S., Goovaerts, A., et al. (2013). Comparative polygenic analysis of maximal ethanol accumulation capacity and tolerance to high ethanol levels of cell proliferation in yeast. *PLoS Genet.* 9, e1003548. doi: 10.1371/journal.pgen.1003548
- Palma, M., Guerreiro, J. F., and Sá-Correia, I. (2018). Adaptive response and tolerance to acetic acid in *Saccharomyces cerevisiae* and *Zygosaccharomyces bailii*: a physiological genomics perspective. *Front. Microbiol.* 9, 274. doi: 10.3389/fmicb.2018.00274
- Palmqvist, E., and Hahn-Hägerdal, B. (2000). Fermentation of lignocellulosic hydrolysates. II: inhibitors and mechanisms of inhibition. *Bioresour. Technol.* 74, 25–33. doi: 10.1016/S0960-8524(99)00161-3
- Parts, L., Cubillos, F. A., Warringer, J., Jain, K., Salinas, F., Bumpstead, S. J., et al. (2011). Revealing the genetic structure of a trait by sequencing a population under selection. *Genome Res.* 21, 1131–1138. doi: 10.1101/gr.116731.110
- Peltier, E., Sharma, V., Raga, M. M., Roncoroni, M., Bernard, M., Gibon, Yves, et al. (2018). Dissection of the molecular bases of genotype x environment interactions: a study of phenotypic plasticity of *Saccharomyces cerevisiae* in grape juices. *BMC Genomics* 19, 772. doi: 10.1186/s12864-018-5145-4
- Pérez-Ortín, J. E., Querol, A., Puig, S., and Barrio, E. (2002). Molecular characterization of a chromosomal rearrangement involved in the adaptive evolution of yeast strains. *Genome Res.* 12, 1533–1539. doi: 10.1101/gr.436602
- Peter, J., and Schacherer, J. (2016). Population genomics of yeasts: towards a comprehensive view across a broad evolutionary scale. *Yeast* 33, 73–81. doi: 10.1002/yea.3142
- Peter, J., De Chiara, M., Friedrich, A., Yue, J.-X., Pflieger, D., Bergström, A., et al. (2018). Genome evolution across 1,011 *Saccharomyces cerevisiae* isolates. *Nature* 556, 339–344. doi: 10.1038/s41586-018-0030-5
- Pigliucci, M., and Kolodnynska, A. (2002). Phenotypic plasticity to light intensity in *Arabidopsis thaliana*: invariance of reaction norms and phenotypic integration. *Evol. Ecol.* 16, 27–47. doi: 10.1023/A:1016073525567
- Pinel, D., Colatrinano, D., Jiang, H., Lee, H., and Martin, V. J. J. (2015). Deconstructing the genetic basis of spent sulphite liquor tolerance using deep sequencing of genome-shuffled yeast. *Biotechnol. Biofuels* 8, 53. doi: 10.1186/s13068-015-0241-z
- Pretorius, I. S. (2000). Tailoring wine yeast for the new millennium: novel approaches to the ancient art of winemaking. *Yeast* 16, 675–729. doi: 10.1002/1097-0061(20000615)16:8<675::AID-YEA585>3.0.CO;2-B
- Pronk, J. T., Steensma, H. Y., and van Dijken, J. P. (1996). Pyruvate metabolism in *Saccharomyces cerevisiae*. *Yeast* 12, 1607–1633. doi: 10.1002/(SICI)1097-0061(199612)12:16<1607::AID-YEA70>3.0.CO;2-4
- Quiros, M., Rojas, V., Gonzalez, R., and Morales, P. (2014). Selection of non-*Saccharomyces* yeast strains for reducing alcohol levels in wine by sugar respiration. *Int. J. Food Microbiol.* 181, 85–91. doi: 10.1016/j.jfoodmicro.2014.04.024
- Reiser, V., Salah, S. M., and Ammerer, G. (2000). Polarized localization of yeast Pbs2 depends on osmotic stress, the membrane protein Sho1 and Cdc42. *Nat. Cell Biol.* 2, 620–627. doi: 10.1038/35023568
- Remize, F., Andrieu, E., and Dequin, S. (2000). Engineering of the pyruvate dehydrogenase bypass in *Saccharomyces cerevisiae*: role of the cytosolic Mg<sup>2+</sup> and mitochondrial K<sup>+</sup> acetaldehyde dehydrogenases Ald6p and Ald4p in acetate formation during alcoholic fermentation. *Appl. Environ. Microbiol.* 66, 3151–3159. doi: 10.1128/AEM.66.8.3151-3159.2000
- Ribéreau-Gayon, P., Glories, Y., Maujean, A., and Dubourdieu, D. (2006). “Alcohols and other volatile compounds,” in *Handbook of enology* (New York, USA: John Wiley & Sons, Ltd), 51–64. doi: 10.1002/0470010398.ch2
- Romano, G. H., Gurvich, Y., Lavi, O., Ulitsky, I., Shamir, R., and Kupiec, M. (2010). Different sets of QTLs influence fitness variation in yeast. *Mol. Syst. Biol.* 6, 346. doi: 10.1038/msb.2010.1
- Roncoroni, M., Santiago, M., Hooks, D. O., Moroney, S., Harsch, M. J., Lee, S., et al. (2011). The yeast IRC7 gene encodes a b-lyase responsible for production of the varietal thiol 4-mercapto-4-methylpentan-2-one in wine. *Food Microbiol.* 28, 926–935. doi: 10.1016/j.fm.2011.01.002
- Sachidanandam, R., Weissman, D., Schmidt, S. C., Kakol, J. M., Stein, L. D., Marth, G., et al. (2001). A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* 409, 928–933. doi: 10.1038/35057149
- Saerens, S. M. G., Verstrepen, K. J., Van Laere, S. D. M., Voet, A. R. D., Van Dijk, P., Delvaux, F. R., et al. (2006). The *Saccharomyces cerevisiae* EHT1 and



- EEB1 genes encode novel enzymes with medium-chain fatty acid ethyl ester synthesis and hydrolysis capacity. *J. Biol. Chem.* 281, 4446–4456. doi: 10.1074/jbc.M512028200
- Saint-Prix, F., Bönquist, L., and Dequin, S. (2004). Functional analysis of the ALD gene family of *Saccharomyces cerevisiae* during anaerobic growth on glucose: the NADP<sup>+</sup>-dependent Ald6p and Ald5p isoforms play a major role in acetate formation. *Microbiology* 150, 2209–2220. doi: 10.1099/mic.0.26999-0
- Salinas, F., Cubillos, F. A., Soto, D., García, V., Bergström, A., Warringer, J., et al. (2012). The genetic basis of natural variation in oenological traits in *Saccharomyces cerevisiae*. *PLoS One* 7, e49640. doi: 10.1371/journal.pone.0049640
- Salinas, F., de Boer, C. G., Abarca, V., García, V., Cuevas, M., Araos, S., et al. (2016). Natural variation in non-coding regions underlying phenotypic diversity in budding yeast. *Sci. Rep.* 6, 21849. doi: 10.1038/srep21849
- Sardi, M., Paithane, V., Place, M., Robinson, D. E., Hose, J., Wohlbach, D. J., et al. (2018). Genome-wide association across *Saccharomyces cerevisiae* strains reveals substantial variation in underlying gene requirements for toxin tolerance. *PLoS Genet.* 14, e1007217. doi: 10.1371/journal.pgen.1007217
- Sasano, Y., Haitani, Y., Hashida, K., Ohtsu, I., Shima, J., and Takagi, H. (2012). Enhancement of the proline and nitric oxide synthetic pathway improves fermentation ability under multiple baking-associated stress conditions in industrial baker's yeast. *Microb. Cell Fact.* 11, 40. doi: 10.1186/1475-2859-11-40
- Schacherer, J., Shapiro, J., Ruderfer, D. M., and Kruglyak, L. (2009). Comprehensive polymorphism survey elucidates population structure of *Saccharomyces cerevisiae*. *Nature* 458, 342–345. doi: 10.1038/nature07670
- Scozzari, R., Massia, A., Trombetta, B., Bellusci, G., Myres, N. M., Novelletto, A., et al. (2014). An unbiased resource of novel SNP markers provides a new chronology for the human Y chromosome and reveals a deep phylogenetic structure in Africa. *Genome Res.* 24, 535–544. doi: 10.1101/gr.160788.113
- Segrè, A. V., Murray, A. W., and Leu, J.-Y. (2006). High-resolution mutation mapping reveals parallel experimental evolution in yeast. *PLoS Biol.* 4, e256. doi: 10.1371/journal.pbio.0040256
- Sekine, T., Kawaguchi, A., Hamano, Y., and Takagi, H. (2007). Desensitization of feedback inhibition of the *Saccharomyces cerevisiae*  $\gamma$ -glutamyl kinase enhances proline accumulation and freezing tolerance. *Appl. Environ. Microbiol.* 73, 4011–4019. doi: 10.1128/AEM.00730-07
- Shahsavari, H., Sugiyama, M., Kaneko, Y., Chuenchit, B., and Harashima, S. (2012). Superior thermotolerance of *Saccharomyces cerevisiae* for efficient bioethanol fermentation can be achieved by overexpression of RSP5 ubiquitin ligase. *Biotechnol. Adv.* 30, 1289–1300. doi: 10.1016/j.biotechadv.2011.09.002
- Sharma, A., Lee, J. S., Dang, C. G., Sudrajat, P., Kim, H. C., Yeon, S. H., et al. (2015). Stories and challenges of genome wide association studies in livestock—a review. *Asian-Australas. J. Anim. Sci.* 28, 1371–1379. doi: 10.5713/ajas.14.0715
- Shobayashi, M., Ukena, E., Fujii, T., and Ilefujii, H. (2007). Genome-wide expression profile of sake brewing yeast under shaking and static conditions. *Biosci. Biotechnol. Biochem.* 71, 323–335. doi: 10.1271/bbb.60190
- Sicard, D., and Legras, J. L. (2011). Bread, beer and wine: yeast domestication in the *Saccharomyces sensu stricto* complex. *C. R. Biol.* 334, 229–236. doi: 10.1016/j.crv.2010.12.016
- Simpson-Lavy, K., Xu, T., Johnston, M., and Kupiec, M. (2017). The Std1 activator of the Snf1/AMPK kinase controls glucose response in yeast by a regulated protein aggregation. *Mol. Cell* 68, 1120–1133.e3. doi: 10.1016/j.molcel.2017.11.016
- Sinha, H., Nicholson, B. P., Steinmetz, L. M., and McCusker, J. H. (2006). Complex genetic interactions in a quantitative trait locus. *PLoS Genet.* 2, 140–147. doi: 10.1371/journal.pgen.0020013
- Sinha, H., David, L., Pascon, R. C., Clauder-Münster, S., Krishnakumar, S., Nguyen, M., et al. (2008). Sequential elimination of major-effect contributors identifies additional quantitative trait loci conditioning high-temperature growth in yeast. *Genetics* 180, 1661–1670. doi: 10.1534/genetics.108.092932
- Sirr, A., Cromie, G. A., Jeffery, E. W., Gilbert, T. L., Ludlow, C. L., Scott, A. C., et al. (2015). Allelic variation, aneuploidy, and nongenetic mechanisms suppress a monogenic trait in yeast. *Genetics* 199, 247–262. doi: 10.1534/genetics.114.170563
- Smith, E. N., and Kruglyak, L. (2008). Gene–environment interaction in yeast gene expression. *PLoS Biol.* 6, e83. doi: 10.1371/journal.pbio.0060083
- Snoek, T., Verstrepen, K. J., and Voordeckers, K. (2016). *Curr. Genet.* 62, 475–480. doi: 10.1007/s00294-015-0561-3
- Stanley, D., Fraser, S., Chambers, P. J., Rogers, P., and Stanley, G. A. (2010). Generation and characterisation of stable ethanol-tolerant mutants of *Saccharomyces cerevisiae*. *J. Ind. Microbiol. Biotechnol.* 37, 139–149. doi: 10.1007/s10295-009-0655-3
- Steensels, J., Snoek, T., Meersman, E., Nicolino, M. P., Voordeckers, K., and Verstrepen, K. J. (2014). Improving industrial yeast strains: exploiting natural and artificial diversity. *FEMS Microbiol. Rev.* 38, 1–49. doi: 10.1111/1574-6976.12073
- Steinmetz, L. M., Sinha, H., Richards, D. R., Spiegelman, J. I., Oefner, P. J., McCusker, J. H., et al. (2002). Dissecting the architecture of a quantitative trait locus in yeast. *Nature* 416, 326–330. doi: 10.1038/416326a
- Stewart, G. G. (2006). Studies on the uptake and metabolism of wort sugars during brewing fermentations. *MBA Tech. Quart* 43, 265–269. doi: 10.1094/TQ-43-0265
- Steyer, D., Ambroset, C., Brion, C., Claudel, P., Delobel, P., Sanchez, I., et al. (2012). QTL mapping of the production of wine aroma compounds by yeast. *BMC Genomics* 13, 573. doi: 10.1186/1471-2164-13-573
- Steyer, D., Erny, C., Claudel, P., Riveill, G., Karst, F., and Legras, J.-L. (2013). Genetic analysis of geraniol metabolism during fermentation. *Food Microbiol.* 33, 228–234. doi: 10.1016/j.fm.2012.09.021
- Storici, F., and Resnick, M. A. (2006). The delitto perfetto approach to *in vivo* site-directed mutagenesis and chromosome rearrangements with synthetic oligonucleotides in yeast. *Meth. Enzymol.* 409, 329–345. doi: 10.1016/S0076-6879(05)09019-1
- Stucki, D., Malla, B., Hostettler, S., Huna, T., Feldmann, J., Yeboah-Manu, D., et al. (2012). Two new rapid SNP-typing methods for classifying *Mycobacterium tuberculosis* complex into the main phylogenetic lineages. *PLoS One* 7, e41253. doi: 10.1371/journal.pone.0041253
- Sumby, K. M., Grbin, P. R., and Jiranek, V. (2010). Microbial modulation of aromatic esters in wine: current knowledge and future prospects. *Food Chem.* 121, 1–16. doi: 10.1016/j.foodchem.2009.12.004
- Swiegers, J. H., and Pretorius, I. S. (2007). Modulation of volatile sulfur compounds by wine yeast. *Appl. Microbiol. Biotechnol.* 74, 954–960. doi: 10.1007/s00253-006-0828-1
- Swinen, S., Schaerlaekens, K., Pais, T., Claesen, J., Hubmann, G., Yang, Y., et al. (2012a). Identification of novel causative genes determining the complex trait of high ethanol tolerance in yeast using pooled-segregant whole-genome sequence analysis. *Genome Res.* 22, 975–984. doi: 10.1101/gr.131698.111
- Swinen, S., Thevelein, J. M., and Nevoigt, E. (2012b). Genetic mapping of quantitative phenotypic traits in *Saccharomyces cerevisiae*. *FEMS Yeast Res.* 12, 215–227. doi: 10.1111/j.1567-1364.2011.00777.x
- Swinen, S., Klein, M., Carrillo, M., McInnes, J., Nguyen, H. T. T., and Nevoigt, E. (2013). Re-evaluation of glycerol utilization in *Saccharomyces cerevisiae*: characterization of an isolate that grows on glycerol without supporting supplements. *Biotechnol. Biofuels* 6, 157. doi: 10.1186/1754-6834-6-157
- Swinen, S., Henriques, S. F., Shrestha, R., Ho, P.-W., Sá-Correia, I., and Nevoigt, E. (2017). Improvement of yeast tolerance to acetic acid through Haa1 transcription factor engineering: towards the underlying mechanisms. *Microb. Cell Fact.* 16, 7. doi: 10.1186/s12934-016-0621-5
- Taherzadeh, M., Gustafsson, L., and Niklasson, C. (2000). Physiological effects of 5-hydroxymethylfurfural on *Saccharomyces cerevisiae*. *Appl. Microbiol. Biotechnol.* 53 (6), 701–708. doi: 10.1007/s002530000328
- Takagi, H. (2008). Proline as a stress protectant in yeast: physiological functions, metabolic regulations, and biotechnological applications. *Appl. Microbiol. Biotechnol.* 81, 211–223. doi: 10.1007/s00253-008-1698-5
- Takagi, H., Iwamoto, F., and Nakamori, S. (1997). Isolation of freeze-tolerant laboratory strains of *Saccharomyces cerevisiae* from proline-analogue-resistant mutants. *Appl. Microbiol. Biotechnol.* 47, 405–411. doi: 10.1007/s002530050948
- Tamás, M. J., and Hohmann, S. (2003). “The osmotic stress response of *Saccharomyces cerevisiae*,” in *Yeast stress responses*. Eds. S. Hohmann and W. H. Mager (Berlin, Heidelberg: Springer Berlin Heidelberg), 121–200. doi: 10.1007/3-540-45611-2\_4
- Tanghe, A., Dijk, P. V., Dumortier, F., Teunissen, A., Hohmann, S., and Thevelein, J. M. (2002). Aquaporin expression correlates with freeze tolerance in baker's yeast, and overexpression improves freeze tolerance in industrial strains. *Appl. Environ. Microbiol.* 68, 5981–5989. doi: 10.1128/AEM.68.12.5981-5989.2002
- Tapia, S. M., Cuevas, M., Abarca, V., Delgado, V., Rojas, V., García, V., et al. (2018). GPD1 and ADH3 natural variants underlie glycerol yield differences in wine fermentation. *Front. Microbiol.* 9, 1460. doi: 10.3389/fmicb.2018.01460

- Taylor, M. B., and Ehrenreich, I. M. (2014). Genetic interactions involving five or more genes contribute to a complex trait in yeast. *PLoS Genet.* 10, e1004324. doi: 10.1371/journal.pgen.1004324
- Teige, M., Scheikl, E., Reiser, V., Ruis, H., and Ammerer, G. (2001). Rck2, a member of the calmodulin-protein kinase family, links protein synthesis to high osmolarity MAP kinase signaling in budding yeast. *Proc. Natl. Acad. Sci. U.S.A.* 98, 5625–5630. doi: 10.1073/pnas.091610798
- Thevelein, J. M., Gorwa-Grauslund, M. F., Dumortier, F., Van Dijck, P., Teunissen, A., Lemaire, K., et al. (1999). Yeast strains with fil phenotype, stress resistant under conditions of growth and/or fermentation.
- Thibon, C., Marullo, P., Claisse, O., Cullin, C., Dubourdieu, D., and Tominaga, T. (2008). Nitrogen catabolic repression controls the release of volatile thiols by *Saccharomyces cerevisiae* during wine fermentation. *FEMS Yeast Res.* 8, 1076–1086. doi: 10.1111/j.1567-1364.2008.00381.x
- Timberlake, W. E., Frizzell, M. A., Richards, K. D., and Gardner, R. C. (2011). A new yeast genetic resource for analysis and breeding. *Yeast* 28, 63–80. doi: 10.1002/yea.1821
- Tominaga, T., and Dubourdieu, D. (2000). Identification of cysteinylated aroma precursors of certain volatile thiols in passion fruit juice. *J. Agric. Food Chem.* 48, 2874–2876. doi: 10.1021/jf990980a
- Trindade de Carvalho, B., Holt, S., Souffriau, B., Lopes Brandão, R., Foulquié-Moreno, M. R., and Thevelein, J. M. (2017). Identification of novel alleles conferring superior production of rose flavor phenylethyl acetate using polygenic analysis in yeast. *MBio* 8, e01173–17. doi: 10.1128/mBio.01173-17
- Veerkamp, R. F., Simm, G., and Oldham, J. D. (1994). Effects of interaction between genotype and feeding system on milk production, feed intake, efficiency and body tissue mobilization in dairy cows. *Livest. Prod. Sci.* 39, 229–241. doi: 10.1016/0301-6226(94)90202-X
- Verstrepen, K. J., Laere, S. D. M., Vanderhaegen, B. M. P., Derdelinckx, G., Dufour, J., Pretorius, I. S., et al. (2003a). Expression levels of the yeast alcohol acetyltransferase genes. *Society* 69, 5228–5237. doi: 10.1128/AEM.69.9.5228-5237.2003
- Verstrepen, K. J., Derdelinckx, G., Dufour, J.-P., Winderickx, J., Thevelein, J. M., Pretorius, I. S., et al. (2003b). Flavor-active esters: adding fruitiness to beer. *J. Biosci. Bioeng.* 96, 110–118. doi: 10.1263/jbb.96.110
- Verstrepen, K. J., Jansen, A., Lewitter, F., and Fink, G. R. (2005). Intragenic tandem repeats generate functional variability. *Nat. Genet.* 37, 986–990. doi: 10.1038/ng1618
- Vidgren, V., and Londesborough, J. (2011). 125th anniversary review: yeast flocculation and sedimentation in brewing. *J. Inst. Brew.* 117, 475–487. doi: 10.1002/j.2050-0416.2011.tb00495.x
- Vidgren, V., Huuskonen, A., Virtanen, H., Ruohonen, L., and Londesborough, J. (2009). Improved fermentation performance of a lager yeast after repair of its AGT1 maltose and maltotriose transporter genes. *Appl. Environ. Microbiol.* 75, 2333–2345. doi: 10.1128/AEM.01558-08
- Vilela-Moura, A., Schuller, D., Mendes-Faia, A., and Côte-Real, M. (2010). Effects of acetic acid, ethanol, and SO<sub>2</sub> on the removal of volatile acidity from acidic wines by two *Saccharomyces cerevisiae* commercial strains. *Appl. Microbiol. Biotechnol.* 87, 1317–1326. doi: 10.1007/s00253-010-2558-7
- Voordeckers, K., Kominek, J., Das, A., Espinosa-Cantú, A., De Maeyer, D., Arslan, A., et al. (2015). Adaptation to high ethanol reveals complex evolutionary pathways. *PLoS Genet.* 11, e1005635. doi: 10.1371/journal.pgen.1005635
- Watanabe, D., Wu, H., Noguchi, C., Zhou, Y., Akao, T., and Shimoi, H. (2011). Enhancement of the initial rate of ethanol fermentation due to dysfunction of yeast stress response components Msn2p and/or Msn4p. *Appl. Environ. Microbiol.* 77, 934–941. doi: 10.1128/AEM.01869-10
- Watanabe, D., Araki, Y., Zhou, Y., Maeya, N., Akao, T., and Shimoi, H. (2012). A loss-of-function mutation in the PAS kinase Rim15p is related to defective quiescence entry and high fermentation rates of *Saccharomyces cerevisiae* sake yeast strains. *Appl. Environ. Microbiol.* 78, 4008–4016. doi: 10.1128/AEM.00165-12
- Watanabe, D., Hashimoto, N., Mizuno, M., Zhou, Y., Akao, T., and Shimoi, H. (2013). Accelerated alcoholic fermentation caused by defective gene expression related to glucose derepression in *Saccharomyces cerevisiae*. *Biosci. Biotechnol. Biochem.* 77, 2255–2262. doi: 10.1271/bbb.130519
- Wei, X., and Zhang, J. (2017). The genomic architecture of interactions between natural genetic polymorphisms and environments in yeast growth. *Genetics* 205, 925–937. doi: 10.1534/genetics.116.195487
- Wilkening, S., Lin, G., Fritsch, E. S., Tekkedil, M. M., Anders, S., Kuehn, R., et al. (2014). An evaluation of high-throughput approaches to QTL mapping in *Saccharomyces cerevisiae*. *Genetics* 196, 853–865. doi: 10.1534/genetics.113.160291
- Will, J. L., Kim, H. S., Clarke, J., Painter, J. C., Fay, J. C., and Gasch, A. P. (2010). Incipient balancing selection through adaptive loss of aquaporins in natural *Saccharomyces cerevisiae* population. *PLoS Genet.* 6, e1000893. doi: 10.1371/journal.pgen.1000893
- Wu, H., Zheng, X., Araki, Y., Sahara, H., Takagi, H., and Shimoi, H. (2006). Global gene expression analysis of yeast cells during sake brewing. *Appl. Environ. Microbiol.* 72, 7353–7358. doi: 10.1128/AEM.01097-06
- Xu, Y., and Crouch, J. H. (2008). Marker-assisted selection in plant breeding: from publications to practice. *Crop Sci.* 48, 391–407. doi: 10.2135/cropsci2007.04.0191
- Yadav, A., and Sinha, H. (2018). Gene–gene and gene–environment interactions in complex traits in yeast. *Yeast* 35, 403–416. doi: 10.1002/yea.3304
- Yadav, A., Radhakrishnan, A., Bhanot, G., and Sinha, H. (2015). Differential regulation of antagonistic pleiotropy in synthetic and natural populations suggests its role in adaptation. *G3 (Bethesda)* 5, 699–709. doi: 10.1534/g3.115.017020
- Yadav, A., Dhole, K., and Sinha, H. (2016). Genetic regulation of phenotypic plasticity and canalisation in yeast growth. *PLoS One* 11, e0162326. doi: 10.1371/journal.pone.0162326
- Yang, J., Bae, J. Y., Lee, Y. M., Kwon, H., Moon, H.-Y., Kang, H. A., et al. (2011). Construction of *Saccharomyces cerevisiae* strains with enhanced ethanol tolerance by mutagenesis of the TATA-binding protein gene and identification of novel genes associated with ethanol tolerance. *Biotechnol. Bioeng.* 108, 1776–1787. doi: 10.1002/bit.23141
- Yang, Y., Foulquié-Moreno, M. R., Clement, L., Erdei, E., Tanghe, A., Schaelekens, K., et al. (2013). QTL analysis of high thermotolerance with superior and downgraded parental yeast strains reveals new minor QTLs and converges on novel causative alleles involved in RNA processing. *PLoS Genet.* 9, e1003693. doi: 10.1371/journal.pgen.1003693
- Yoshida, S., Imoto, J., Minato, T., Oouchi, R., Kamada, Y., Tomita, M., et al. (2010). A novel mechanism regulates H<sub>2</sub>S and SO<sub>2</sub> production in *Saccharomyces cerevisiae*. *Yeast* 28, 109–121. doi: 10.1002/yea.1823
- Youns-Clark, K., Buckler, E., Casstevens, T., Chen, C., DeClerck, G., Derwent, P., et al. (2011). Gramene database in 2010: updates and extensions. *Nucleic Acids Res.* 39, D1085–D1094. doi: 10.1093/nar/gkq1148
- Yvert, G., Brem, R. B., Whittle, J., Akey, J. M., Foss, E., Smith, E. N., et al. (2003). Trans-acting regulatory variation in *Saccharomyces cerevisiae* and the role of transcription factors. *Nat. Genet.* 35, 57–64. doi: 10.1038/ng1222
- Zampar, G. G., Kümmel, A., Ewald, J., Jol, S., Niebel, B., Picotti, P., et al. (2013). Temporal system-level organization of the switch from glycolytic to gluconeogenic operation in yeast. *Mol. Syst. Biol.* 9, 651. doi: 10.1038/msb.2013.11
- Zhao, X. Q., and Bai, F. W. (2009). Mechanisms of yeast stress tolerance and its manipulation for efficient fuel ethanol production. *J. Biotechnol.* 144, 23–30. doi: 10.1016/j.jbiotec.2009.05.001
- Zimmer, A., Durand, C., Loira, N., Durrens, P., Sherman, D. J., and Marullo, P. (2014). QTL dissection of lag phase in wine fermentation reveals a new translocation responsible for *Saccharomyces cerevisiae* adaptation to sulfite. *PLoS One* 9, 37–39. doi: 10.1371/journal.pone.0086298
- Zimmermann, F. K., Kaufmann, I., Rasenberger, H., and Haufmann, P. (1977). Genetics of carbon catabolite repression in *Saccharomyces cerevisiae*: genes involved in the derepression process. *Mol. Gen. Genet.* 151, 95–103. doi: 10.1007/BF00446918
- Zörgö, E., Chwialkowska, K., Gjuvsland, A. B., Garré, E., Sunnerhagen, P., Liti, G., et al. (2013). Ancient evolutionary trade-offs between yeast ploidy states. *PLoS Genet.* 9, e1003388. doi: 10.1371/journal.pgen.1003388

**Conflict of Interest Statement:** EP and PM are employed by LAFFORT. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Peltier, Friedrich, Schacherer and Marullo. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# A Unique *Saccharomyces cerevisiae* × *Saccharomyces uvarum* Hybrid Isolated From Norwegian Farmhouse Beer: Characterization and Reconstruction

Kristoffer Krogerus<sup>1,2\*</sup>, Richard Preiss<sup>3,4</sup> and Brian Gibson<sup>1</sup>

<sup>1</sup> VTT Technical Research Centre of Finland Ltd., Espoo, Finland, <sup>2</sup> Department of Biotechnology and Chemical Technology, School of Chemical Technology, Aalto University, Espoo, Finland, <sup>3</sup> Department of Molecular and Cellular Biology, University of Guelph, Guelph, ON, Canada, <sup>4</sup> Escarpment Laboratories, Guelph, ON, Canada

## OPEN ACCESS

### Edited by:

Ed Louis,  
University of Leicester,  
United Kingdom

### Reviewed by:

Liti Gianni,  
INSERM U1081 Institut de Recherche  
sur le Cancer et le Vieillessement,  
France  
Barbara Dunn,  
Stanford University, United States

### \*Correspondence:

Kristoffer Krogerus  
kristoffer.krogerus@gmail.com

### Specialty section:

This article was submitted to  
Evolutionary and Genomic  
Microbiology,  
a section of the journal  
Frontiers in Microbiology

**Received:** 20 June 2018

**Accepted:** 04 September 2018

**Published:** 24 September 2018

### Citation:

Krogerus K, Preiss R and Gibson B  
(2018) A Unique *Saccharomyces*  
*cerevisiae* × *Saccharomyces uvarum*  
Hybrid Isolated From Norwegian  
Farmhouse Beer: Characterization  
and Reconstruction.  
Front. Microbiol. 9:2253.  
doi: 10.3389/fmicb.2018.02253

An unknown interspecies *Saccharomyces* hybrid, “Muri,” was recently isolated from a “kveik” culture, a traditional Norwegian farmhouse brewing yeast culture (Preiss et al., 2018). Here we used whole genome sequencing to reveal the strain as an allodiploid *Saccharomyces cerevisiae* × *Saccharomyces uvarum* hybrid. Phylogenetic analysis of its sub-genomes revealed that the *S. cerevisiae* and *S. uvarum* parent strains of Muri appear to be most closely related to English ale and Central European cider and wine strains, respectively. We then performed phenotypic analysis on a number of brewing-relevant traits in a range of *S. cerevisiae*, *S. uvarum* and hybrid strains closely related to the Muri hybrid. The Muri strain possesses a range of industrially desirable phenotypic properties, including broad temperature tolerance, good ethanol tolerance, and efficient carbohydrate use, therefore making it an interesting candidate for not only brewing applications, but potentially various other industrial fermentations, such as biofuel production and distilling. We identified the two *S. cerevisiae* and *S. uvarum* strains that were genetically and phenotypically most similar to the Muri hybrid, and then attempted to reconstruct the Muri hybrid by generating *de novo* interspecific hybrids between these two strains. The *de novo* hybrids were compared with the original Muri hybrid, and many appeared phenotypically more similar to Muri than either of the parent strains. This study introduces a novel approach to studying hybrid strains and strain development by combining genomic and phenotypic analysis to identify closely related parent strains for construction of *de novo* hybrids.

**Keywords:** yeast, beer, hybrid, kveik, dextrin, genome

## INTRODUCTION

Yeast has a central role in beer production and is responsible for the conversion of wort carbohydrates into ethanol and CO<sub>2</sub>, as well as the synthesis of flavor compounds. Beer is traditionally fermented with domesticated strains of *Saccharomyces cerevisiae* and *Saccharomyces pastorianus* (Gibson et al., 2017). The demand from consumers for craft and specialty beers that



have rich and unique aroma has increased in recent years (Aquilani et al., 2015), which has led breweries to explore alternative yeasts. One such group of yeasts that have recently gained attention are “kveik” yeasts: a range of traditional yeast cultures that have been used and maintained by Norwegian farmhouse brewers (Garshol, 2013, 2016; Fowle, 2017; Preiss et al., 2018). The genetic and phenotypic diversity of a range of kveik strains was recently explored by Preiss et al. (2018). The vast majority of the isolates were identified as strains of *S. cerevisiae*, which formed a genetically distinct group possessing properties relevant to brewing. One isolate, however, was found to be an unknown *Saccharomyces* interspecies hybrid. Here, we aimed to identify and characterize this hybrid strain.

The use of interspecific yeast hybrids for beer fermentation is widespread, with lager yeast, i.e., *S. pastorianus*, being used for the majority of global beer production. This *S. cerevisiae* × *Saccharomyces eubayanus* hybrid combines the efficient wort sugar utilization of the *S. cerevisiae* parent with the cold tolerance of the *S. eubayanus* parent (Gibson and Liti, 2015). Interspecific hybridization not only allows for the combination of phenotypic traits from diverse parent strains, but hybrids often exhibit superior phenotypic qualities relative to parent strains, i.e., heterosis or hybrid vigor. The relatively harsh and stressful environment that yeast is exposed to during beer fermentation may have selected for interspecific hybrids, which have been shown to exhibit increased stress tolerance (Lopandic, 2018). In addition to the lager yeast, several other natural interspecific *Saccharomyces* hybrids have been isolated in brewing environments. Hybrids between *S. cerevisiae* and *Saccharomyces kudriavzevii* are used for the fermentation of several Belgian Trappist beers (González et al., 2008), while *Saccharomyces bayanus* (*S. eubayanus* × *S. uvarum*) hybrids have been isolated as contaminants from beer (Rainieri et al., 2006; Nguyen et al., 2011). Natural hybrids between *S. cerevisiae* and *S. uvarum* are also used in winemaking (Christine et al., 2007), but limited reports exist describing use of such hybrids in brewing.

In addition to natural *Saccharomyces* hybrids, *de novo* interspecific *Saccharomyces* hybrids can also readily be generated. These have been studied for their potential in a range of industrial applications, including biofuel production (Snoek et al., 2015; Peris et al., 2017), brewing (Krogerus et al., 2015; Mertens et al., 2015) and winemaking (Bellon et al., 2011; Origone et al., 2018). *De novo* hybrids have exhibited various improved traits compared to their parent strains, including faster fermentation rates, more complete sugar use, greater stress tolerance, and increases in aroma compound production (Bellon et al., 2011; Dunn et al., 2013; Steensels et al., 2014a; Krogerus et al., 2015; Mertens et al., 2015; Snoek et al., 2015). For brewing, much of the recent research has focused on the generation and characterization of new lager yeast, i.e., *S. cerevisiae* × *S. eubayanus*, hybrids (Hebly et al., 2015; Krogerus et al., 2015, 2016, 2017, 2018; Mertens et al., 2015; Alexander et al., 2016). However, alternative *Saccharomyces* interspecific hybrid combinations have also shown promise in brewing conditions (Sato et al., 2002; Nikulin et al., 2018). In addition to their industrial applications, *de novo* hybrids have also acted as useful models for studying adaptation and molding of

hybrid genomes (Dunn et al., 2013; Peris et al., 2017; Smukowski Heil et al., 2017; Krogerus et al., 2018). Such studies could be useful for lager yeast in particular, as much of their natural evolutionary history still remains obscure despite their industrial importance (Baker et al., 2015; Okuno et al., 2015).

Recent whole genome sequencing studies have revealed multiple domestication events for *S. cerevisiae* (Gallone et al., 2016; Gonçalves et al., 2016; Peter et al., 2018). Commercially used brewing strains, for example, tend to cluster into one of two independently domesticated ‘Beer’ groups. While these studies have focused on *S. cerevisiae*, and limited data on interspecific *Saccharomyces* hybrids is available, Gonçalves et al. (2016) also revealed that the *S. cerevisiae* sub-genomes of lager yeasts group among the “Beer 1” or “Ale beer” yeasts. However, no single strain has yet been identified as the potential *S. cerevisiae* ale parent of lager yeast (Monerawela and Bond, 2018). In addition to providing valuable data on the diversity and history of brewing strains, whole genome sequence data, in combination with comprehensive phenotype data, are valuable resources for parent selection during breeding and hybridization. Gallone et al. (2016) demonstrated how a strain lacking phenolic off-flavor (POF) formation could be obtained through mating of parent strains carrying heterozygous loss-of-function polymorphisms in *FDC1*. Heterosis has also been shown to be positively correlated with sequence divergence during breeding of domesticated strains (Plech et al., 2014).

Recently, Preiss et al. (2018) described the isolation of an unknown *Saccharomyces* interspecies hybrid, i.e., the Muri strain, from a Norwegian farmhouse brewing (kveik) culture. This strain was identified as a *S. bayanus*-type hybrid based on internally transcribed spacer (ITS) sequencing. Here we aimed to identify, characterize, and ultimately reconstruct this hybrid. Whole genome sequencing was used to reveal the strain as an allodiploid *S. cerevisiae* × *S. uvarum* hybrid. We then performed phylogenetic analysis of its sub-genomes, in an attempt to identify *S. cerevisiae* and *S. uvarum* strains closely related to the parent strains of Muri. In addition, we compared a number of brewing-relevant phenotypic traits in *S. cerevisiae*, *S. uvarum* and hybrid strains closely related to the Muri hybrid. These data were used to identify two *S. cerevisiae* and *S. uvarum* strains that were genetically and phenotypically most similar to the Muri hybrid. We then attempted to reconstruct the Muri hybrid by generating *de novo* interspecific hybrids between these two strains. The *de novo* hybrids were compared with the original Muri hybrid, and appeared phenotypically more similar to Muri than either of the parent strains. This study introduces a novel approach to studying hybrid strains and strain development by combining genomic and phenotypic analysis to identify closely related parent strains for construction of *de novo* hybrids.

## MATERIALS AND METHODS

### Yeast Strains

A list of strains used in this study can be found in **Table 1**. The *de novo* yeast hybrids between *S. cerevisiae* A241 and *S. uvarum* C995 were constructed by spore-to-spore mating as described



**TABLE 1** | Yeast strains used in the study.

Code	Alternative name	Species	Origin	Reference
Muri		<i>S. cerevisiae</i> × <i>S. uvarum</i>	Beer, Norway	Preiss et al., 2018
A240	Beer032	<i>S. cerevisiae</i>	Beer, England	Gallone et al., 2016
A241	Beer059	<i>S. cerevisiae</i>	Beer, England	Gallone et al., 2016
A242	Beer062	<i>S. cerevisiae</i>	Beer, England	Gallone et al., 2016
A243	Beer091	<i>S. cerevisiae</i>	Beer, Germany	Gallone et al., 2016
C992	ZP556	<i>S. uvarum</i>	Oak tree, Canada	Almeida et al., 2014
C993	NCAIM 01116	<i>S. uvarum</i>	Grapes, Russia	Almeida et al., 2014
C995	ZP646	<i>S. uvarum</i>	Cider, Germany	Almeida et al., 2014
C996	CBS8614	<i>S. cerevisiae</i> × <i>S. kudriavzevii</i> × <i>S. uvarum</i>	Cider, France	Almeida et al., 2014
C997	CBS8615	<i>S. cerevisiae</i> × <i>S. uvarum</i>	Wine, Italy	Almeida et al., 2014
C998	NCAIM 00676	<i>S. eubayanus</i> × <i>S. uvarum</i>	Fermented beverage, Hungary	Almeida et al., 2014
3B	A241 × C995 3B	<i>S. cerevisiae</i> × <i>S. uvarum</i>	<i>de novo</i> hybrid	This study
4B	A241 × C995 4B	<i>S. cerevisiae</i> × <i>S. uvarum</i>	<i>de novo</i> hybrid	This study
6A	A241 × C995 6A	<i>S. cerevisiae</i> × <i>S. uvarum</i>	<i>de novo</i> hybrid	This study
13C	A241 × C995 13C	<i>S. cerevisiae</i> × <i>S. uvarum</i>	<i>de novo</i> hybrid	This study

previously (Krogerus et al., 2016). Potential hybrids were first identified by the ability to both grow at 37°C (*S. cerevisiae*-specific) and form blue-colored colonies (*S. uvarum*-specific) when grown in the presence of X- $\alpha$ -Gal (#16555 Sigma-Aldrich). Single cell isolates of potential hybrids were obtained by re-streaking single colonies three times on YP-Glucose agar. Hybrid status was confirmed through PCR using species-specific (*S. cerevisiae* and *S. uvarum*) primers (Muir et al., 2011; Pengelly and Wheals, 2013) and ITS-PCR followed by *HaeIII* digestion (Pham et al., 2011). The ploidy of the Muri strain was determined by flow cytometry as described previously (Krogerus et al., 2017).

## Growth Assays

Growth of the yeast strains at various temperatures (4, 12, 37, and 40 °C) was tested on YP-Glucose agar plates. Overnight pre-cultures of all the strains were grown in YP-Glucose at 25°C. The yeast was then pelleted and resuspended in 50 mM citrate buffer (pH 7.2) to deflocculate the yeast. The cell concentration was measured with a NucleoCounter YC-100 (ChemoMetec, Denmark), after which suspensions were diluted to contain approximately 10<sup>5</sup>, 10<sup>4</sup>, and 10<sup>3</sup> cells ml<sup>-1</sup>. Five-microliter aliquots of the suspensions of each strain were spotted onto agar plates. Plates were sealed with Parafilm and incubated for up to 21 days, after which growth was scored based on colony size. Growth in the presence of ethanol was tested in YP-Glucose media supplemented with 11% ethanol. One-microliter of media was inoculated with the overnight pre-cultures to a starting OD<sub>600</sub> of 0.05. Cultures were incubated at 25°C for 4 days, after which the optical densities of the cultures were measured.

## 100 mL-Scale Wort Fermentations

100 mL-scale fermentations were carried out in 250 mL Erlenmeyer flasks capped with glycerol-filled airlocks. Yeast strains were grown overnight in 50 mL YP-Maltose at 24°C. The pre-cultured yeast was then inoculated into 100 mL of 15°P all-malt wort at a rate of 10 × 10<sup>6</sup> viable cells mL<sup>-1</sup>. Fermentations

were carried out in duplicate at 12 and 20°C for 16 and 10 days, respectively, and these were monitored daily by mass lost as CO<sub>2</sub>. Samples for sugar, ethanol, and yeast-derived flavor compounds analysis were drawn from the beer when fermentations were ended. Yeast dry mass was determined from centrifuged and twice washed samples that were dried overnight at 105°C.

## Chemical Analysis of Wort and Beer

Concentrations of fermentable sugars (maltose and maltotriose) were measured by HPLC using a Waters 2695 Separation Module and Waters System Interphase Module liquid chromatograph coupled with a Waters 2414 differential refractometer (Waters Co., Milford, MA, United States). An Aminex HPX-87H organic acid analysis column (300 × 7.8 mm, Bio-Rad) was equilibrated with 5 mM H<sub>2</sub>SO<sub>4</sub> (Titrisol, Merck, Germany) in water at 55°C and samples were eluted with 5 mM H<sub>2</sub>SO<sub>4</sub> in water at a 0.3 mL/min flow rate.

The alcohol level (% v/v) of samples was determined from the centrifuged and degassed fermentation samples using an Anton Paar Density Meter DMA 5000 M with an AlcoLyzer Beer ME module (Anton Paar GmbH, Austria).

Yeast-derived higher alcohols and esters were determined by headspace gas chromatography with flame ionization detector (HS-GC-FID) analysis. 4 mL samples were filtered (0.45  $\mu$ m), incubated at 60°C for 30 min and then 1 mL of gas phase was injected (split mode; 225°C; split flow of 30 mL min<sup>-1</sup>) into a gas chromatograph equipped with an FID detector and headspace autosampler (Agilent 7890 Series; Palo Alto, CA, United States). Analytes were separated on a HP-5 capillary column (50 m × 320  $\mu$ m × 1.05  $\mu$ m column, Agilent, United States). The carrier gas was helium (constant flow of 1.4 mL min<sup>-1</sup>). The temperature program was 50°C for 3 min, 10°C min<sup>-1</sup>–100°C, 5°C min<sup>-1</sup>–140°C, 15°C min<sup>-1</sup>–260°C and then isothermal for 1 min. Compounds were identified by comparison with authentic standards and were quantified using standard curves. 1-Butanol was used as internal standard.

## Flocculation Assay

Flocculation of the yeast strains was evaluated using a modified Helm's assay essentially as described by D'Hautcourt and Smart (1999). Yeast strains were grown overnight in 50 mL YP-Glucose at 24°C. The yeast was washed twice with 0.5 M EDTA (pH 7) to break the cell aggregates and then diluted to an OD<sub>600</sub> of 0.4. Flocculation was assayed by first washing yeast pellets with 4 mM CaCl<sub>2</sub> solution and resuspending them in 1 mL of flocculation solution containing 4 mM CaCl<sub>2</sub>, 6.8 g/L sodium acetate, 4.05 g/L acetic acid, and 4% (v/v) ethanol (pH 4.5). Yeast cells in control tubes were resuspended in 0.5 M EDTA (pH 7). After a sedimentation period of 10 min, samples (200 µL) were taken from just below the meniscus and dispersed in 10 mM EDTA (800 µL). The absorbance at 600 nm was measured, and percentage of flocculation was determined from the difference in absorbance between control and flocculation tubes. The assay was performed in triplicate.

## Melibiose Activity of Yeast

Melibiose activity was tested based on the ability to form blue-colored colonies when grown in the presence of X-α-Gal (#16555 Sigma-Aldrich) (ASBC, 2011).

## Dextrin Fermentation

The ability to ferment dextrin was tested in minimal growth media with dextrin as the sole carbon source. Strains were grown overnight in YP-Glucose, after which 2 mL microcentrifuge tubes containing 1 mL of dextrin media (0.67% YNB without amino acids, 1% dextrin from potato starch) were inoculated with 20 µL of the overnight cultures. The tubes were incubated at room temperature for 3 weeks, after which the refractive index of the culture media was measured with a Quick-Brix 90 digital refractometer (Mettler-Toledo AG, Switzerland). A decrease in refractive index indicated fermentation of dextrin. *S. cerevisiae* WLP590 (White Labs Inc, United States) and *S. pastorianus* VTT-A63015 (VTT culture collection, Finland) were included as positive and negative control strains, respectively. No change in refractive index was observed for the negative control strain. In addition, the presence of the *STA1* gene was tested with PCR using primers SD-5A and SD-6B (Yamauchi et al., 1998).

## Phenolic Off-Flavor Assay

The ability to produce POF was tested using the absorbance-based method described in Mertens et al. (2017). The test was performed in 2 mL microcentrifuge tubes containing 1 mL of media instead of 96-well plates as described in Mertens et al. (2017).

## Multiplex PCR With Species-Specific Primers

Amplification of the *S. cerevisiae*-specific *MEX67* gene (amplicon size 150 bp), *S. eubayanus*-specific *FSY1* gene (amplicon size 228 bp) and *S. uvarum*-specific *DBP6* gene (amplicon size 275 bp) was performed with PCR using the ScerF2, ScerR2, SeubF3, SeubR2, SbayF1, and SbayR1 primers described by Muir et al. (2011) and Pengelly and Wheals (2013).

## PCR-RFLP of COX2 to Determine Origin of mtDNA in Hybrids

Amplification of the mitochondrial COX2 gene was performed with PCR using the COII-3 and COII-5 primers described by Belloch et al. (2000). The amplicon size (656 bp) of both the *S. cerevisiae*- and *S. uvarum*-derived COX2 were of equal size, and they could therefore not be differentiated based on size. Digestion with the *HaeIII* restriction enzyme (New England Biolabs, United States) did not affect the *S. cerevisiae*-derived COX2 amplicon, but yielded a 75 bp smaller fragment (581 bp) for the *S. uvarum*-derived COX2 amplicon.

## Genome Sequencing and Analysis

The "Muri" strain was sequenced by Biomedicum Genomics (Helsinki, Finland). In brief, DNA was initially isolated using Qiagen 100/G Genomic tips (Qiagen, Netherlands), after which an Illumina TruSeq LT paired-end 150 bp library was prepared for each strain and sequencing was carried out with a NextSeq500 instrument. Paired-end reads from the NextSeq500 sequencing were quality-analyzed with FastQC (Andrews, 2010) and trimmed and filtered with Cutadapt (Martin, 2011). Reads were aligned to a concatenated reference genome of *S. cerevisiae* S288C (R64-2-1), *S. eubayanus* FM1318 (Baker et al., 2015) and *S. uvarum* CBS7001 (Scannell et al., 2011) using SpeedSeq (Chiang et al., 2015). Quality of alignments was assessed with QualiMap (García-Alcalde et al., 2012). Variant analysis was performed on aligned reads using FreeBayes (Garrison and Marth, 2012). Prior to variant analysis, alignments were filtered to a minimum MAPQ of 50 with SAMtools (Li et al., 2009). Variants at sites where read depth was below 10 were also excluded. Interchromosomal translocations were detected based on split reads with Manta (Chen et al., 2016), and visualized with the "circlize" package in R (Gu et al., 2014). The median coverage over 10,000 bp windows was calculated with BEDTools (Quinlan and Hall, 2010) and visualized in R. Gene ontology enrichment was performed with YeastMine (Balakrishnan et al., 2012). The raw sequencing reads are available in the NCBI's Short Read Archive under BioProject PRJNA475668 in the NCBI BioProject database<sup>1</sup>.

## Phylogenetic Analysis

Prior to phylogenetic analysis, consensus genotypes of the *S. cerevisiae* and *S. uvarum* sub-genomes of the Muri strain were called from the identified variants using BCFtools (Li, 2011). Regions where the sequencing coverage was below 10 were excluded from the consensus genotypes. Genome assemblies of the 157 *S. cerevisiae* strains described in Gallone et al. (Gallone et al., 2016) were retrieved from NCBI (BioProject PRJNA323691). Consensus genotypes of 61 *S. uvarum* and hybrid strains described in Almeida et al. (2014) were kindly provided by José Paulo Sampaio. Multiple sequence alignment of the consensus genotype of the *S. cerevisiae* sub-genome of Muri and the 157 *S. cerevisiae* assemblies was performed with the NASP pipeline (Roe et al., 2016) using *S. cerevisiae* S288C (R64-2-1) as the reference genome. A matrix of single nucleotide

<sup>1</sup><https://www.ncbi.nlm.nih.gov/bioproject/>

polymorphisms (SNP) in the 159 strains was extracted from the aligned sequences. The SNPs were annotated with SnpEff (Cingolani et al., 2012) and filtered as follows: only sites that were in the coding sequence of genes, present in all 159 strains and with a minor allele frequency greater than 1% (one strain) were retained. The filtered matrix contained 3753194 SNPs (129434 sites). A maximum likelihood phylogenetic tree was estimated using IQ-TREE (Nguyen et al., 2015). IQ-TREE was run using the “GTR + F + R4” model and 1000 ultrafast bootstrap replicates (Minh et al., 2013). The resulting maximum likelihood tree was visualized in FigTree and rooted with *S. paradoxus* CBS432 (Yue et al., 2017). The above steps from multiple sequence alignment onward were repeated with the consensus genotypes of the *S. uvarum* strains (Almeida et al., 2014) and the *S. uvarum* sub-genome of Muri using *S. uvarum* CBS7001 as the reference genome (Scannell et al., 2011). The filtered matrix contained 2189200 SNPs (352638 sites).

## Data Visualization and Analysis

Data and statistical analyses were performed with R<sup>2</sup>. Z-scores ( $z$ ) of the phenotypic traits were calculated as  $z = (x - \mu) / \sigma$ , where  $x$  is value of a trait for a particular strain,  $\mu$  is the mean value of that trait in all strains, and  $\sigma$  is the standard deviation of that trait in all strains. Heat maps with hierarchical clustering and optimal leaf ordering of the strains were generated based on the  $z$ -scores with the “seriation” package (Hahsler et al., 2008). Principal component analysis (PCA) was also performed on the set of  $z$ -scores. Prior to PCA, the  $z$ -scores from the concentrations of aroma compounds were scaled based on their flavor threshold as  $\mu / C_{\text{threshold}}$ , where  $\mu$  is the mean concentration of that compound in all strains, and  $C_{\text{threshold}}$  is the aroma threshold of that compound (Meilgaard, 1982). This weighting was performed so that aroma compounds with concentrations much below the aroma threshold would have less impact on the PCA. Flow cytometry data was analyzed with “flowCore” (Hahne et al., 2009) and “mixtools” (Benaglia et al., 2009) packages. Plots were produced in R and FigTree.

## Data Availability

The Illumina reads generated in this study have been submitted to NCBI-SRA under BioProject number PRJNA475668 in the NCBI BioProject database (see footnote 1).

## RESULTS

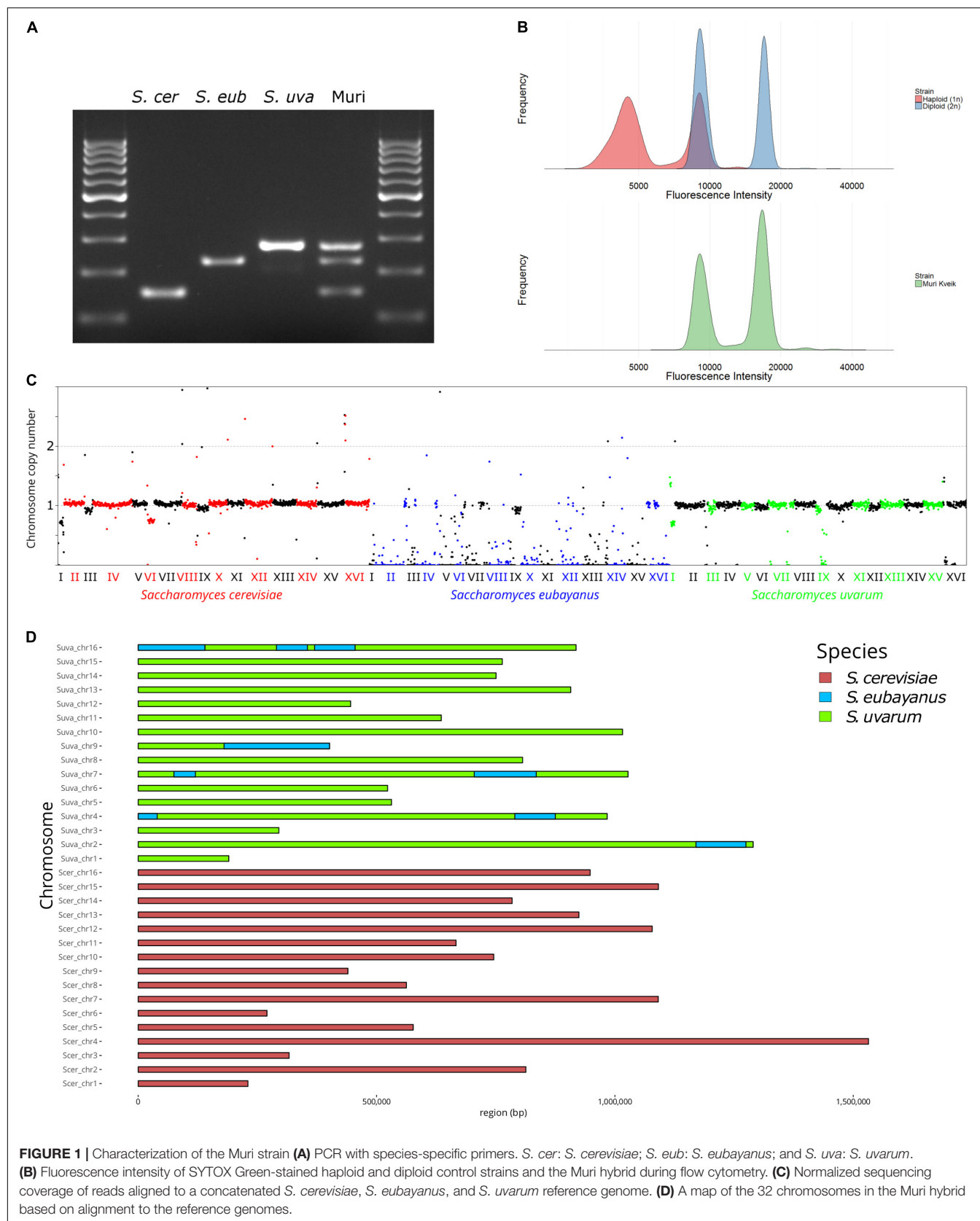
### Analysis of the Muri Genome

The genetic and phenotypic diversity of a range of Norwegian farmhouse brewing strains, i.e., kveik strains, was recently explored by Preiss et al. (2018). Sequencing of the ITS region identified the vast majority of the isolates as *S. cerevisiae*. However, one isolate, i.e., the Muri strain that is investigated in this study, was identified as a *S. bayanus*-type hybrid.

To further characterize the genetic background of the Muri strain, we initially tested the species-specific multiplex PCR primer set described by Muir et al. (2011) and Pengelly and Wheals (2013). These primers yielded bands for *S. cerevisiae*, *S. eubayanus*, and *S. uvarum* (Figure 1A). Flow cytometry analysis further revealed that the Muri strain is approximately diploid (Figure 1B). We then sequenced the genome of the Muri strain with 150 bp paired-end Illumina sequencing. Based on the results from the species-specific primers, we first aligned the trimmed sequencing reads to a concatenated reference genome of *S. cerevisiae* S288C, *S. eubayanus* FM1318, and *S. uvarum* CBS7001 (Figure 1C). The alignment (263 × coverage and 97.3% mapped reads) suggested that the Muri strain is an allodiploid *S. cerevisiae* × *S. uvarum* hybrid, containing introgressions from *S. eubayanus* (Supplementary Data S2). Using Manta, we identified that the majority of these *S. eubayanus* introgressions were in the *S. uvarum* sub-genome (Figure 1D and Supplementary Figure S1). Interestingly, *S. uvarum* chromosome 9 appears to be a chimeric chromosome, where the right arm has been replaced by that of *S. eubayanus* chromosome 9 (the breakpoints identified by Manta are at positions 196846 and 180023 bp in the *S. uvarum* and *S. eubayanus* sequences, respectively). In addition, there were substantial contributions from *S. eubayanus* chromosomes 2, 4, 7, and 16 (Figure 1D and Supplementary Data S2). The mitochondria appeared to be of *S. uvarum* origin based on sequencing coverage when reads were aligned to the mitochondrial DNA (mtDNA) of *S. cerevisiae*, *S. eubayanus*, and *S. uvarum* (Supplementary Table S1 and Supplementary Figure S4C).

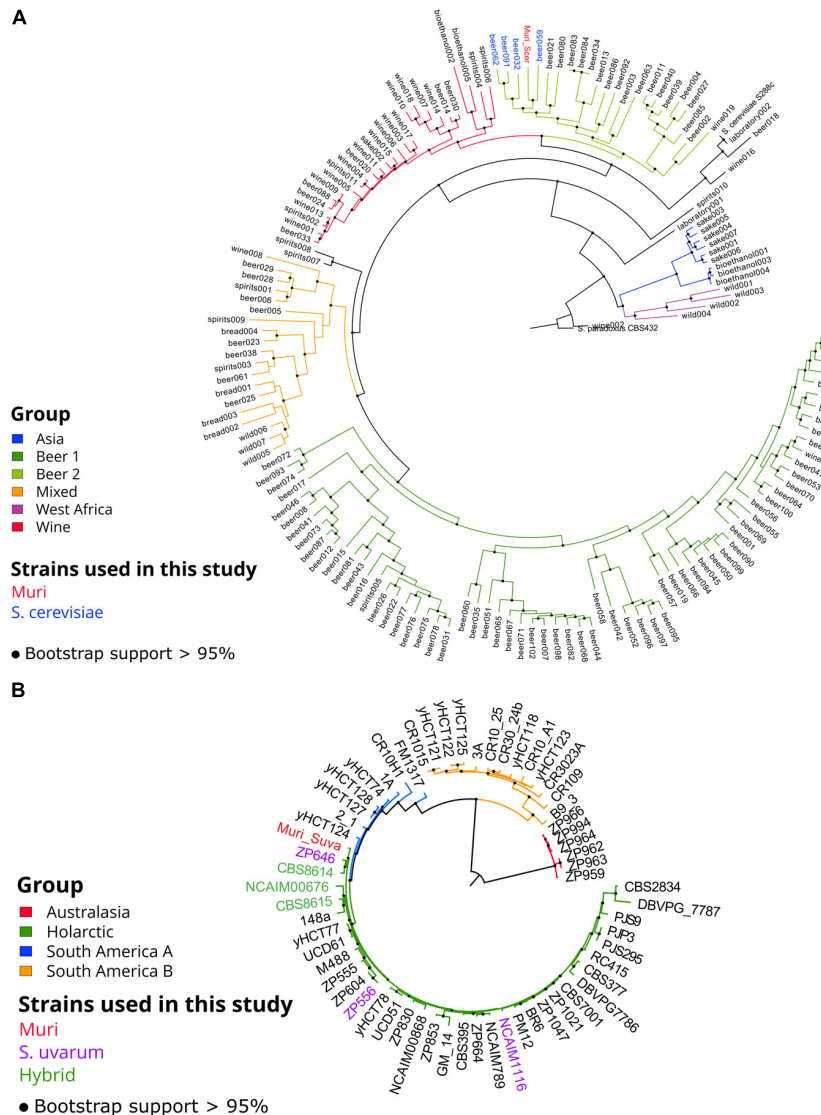
Single nucleotide polymorphism (SNP) analysis of the aligned sequencing reads of the Muri strain revealed 48983 and 26439 SNPs in the haploid *S. cerevisiae* and *S. uvarum* sub-genomes, respectively, compared to the reference genomes. Allele frequency distributions of the SNPs suggest a single allele at each site, supporting the flow cytometry results of two haploid sub-genomes (Supplementary Figure S2). Consensus sequences of both sub-genomes were then produced from these SNPs (regions where coverage was below 10× were excluded). In order to identify *S. cerevisiae* and *S. uvarum* strains closely related to the parent strains of the Muri hybrid, we performed multiple sequence alignment, SNP identification and phylogenetic analysis using the NASP pipeline and IQ-TREE with the consensus sequences and 157 *S. cerevisiae* genome assemblies obtained from Gallone et al. (2016; Bioproject PRJNA323691), as well as 61 *S. uvarum* consensus sequences obtained from Almeida et al. (2014; Bioproject PRJNA230139). The inferred maximum-likelihood tree of the *S. cerevisiae* genomes, based on 129434 polymorphic sites, suggests that the *S. cerevisiae* sub-genome of the Muri strain belongs to a lineage of beer yeasts [“Beer 2” from Gallone et al. (2016) or “Mosaic Beer” from Peter et al. (2018)], with its closest relatives being ale yeasts (e.g., Beer059 and Beer032) from the United Kingdom (Figure 2A). These ale yeasts are characterized by good ethanol tolerance and production, as well as high flocculation (Gallone et al., 2016). In addition, many of these strains carry the *STA1* gene, encoding an extracellular glucoamylase that can cause super-attenuation during beer fermentation (Yamashita et al., 1985), as

<sup>2</sup><http://www.r-project.org/>



**FIGURE 1 |** Characterization of the Muri strain **(A)** PCR with species-specific primers. *S. cer*: *S. cerevisiae*; *S. eub*: *S. eubayanus*; and *S. uva*: *S. uvarum*. **(B)** Fluorescence intensity of SYTOX Green-stained haploid and diploid control strains and the Muri hybrid during flow cytometry. **(C)** Normalized sequencing coverage of reads aligned to a concatenated *S. cerevisiae*, *S. eubayanus*, and *S. uvarum* reference genome. **(D)** A map of the 32 chromosomes in the Muri hybrid based on alignment to the reference genomes.





**FIGURE 2 |** Phylogeny of Muri's sub-genomes and the 157 *S. cerevisiae* strains described in Gallone et al. (2016) and 61 *S. uvarum* and hybrid strains described in Almeida et al. (2014). **(A)** Maximum likelihood phylogenetic tree based on SNPs at 129434 sites in 158 *S. cerevisiae* genomes (rooted with *S. paradoxus* as outgroup). **(B)** Maximum likelihood phylogenetic tree based on SNPs at 352638 sites in 62 *S. uvarum* and hybrid genomes (rooted with the Australasian strains as outgroup). Black dots on nodes indicate bootstrap support values >95%. Branches are colored according to lineage, and the names of strains used in this study (see Table 1) are colored (Muri, red; *S. cerevisiae*, blue; *S. uvarum*, purple; and Hybrids, green). Branch lengths represent the number of substitutions per site. Root branch lengths have been artificially shortened to improve the clarity of the trees.

was revealed by a BLAST search of the genome assemblies (data not shown) and using PCR (**Supplementary Figure S3A**). The *STA1* gene was also identified in the Muri strain (**Supplementary Figure S3**). Clustering within the “Beer 2” yeasts, which have an estimated common ancestor at the end of the 17<sup>th</sup> century (Gallone et al., 2016), would suggest a recent hybridization event for Muri. Likewise, the inferred maximum likelihood tree of the *S. uvarum* genomes, based on 352638 polymorphic sites, suggests the *S. uvarum* sub-genome of the Muri strain belongs to the Holarctic lineage (Almeida et al., 2014), with its closest relatives being domesticated Central European strains used in cider and wine fermentation (**Figure 2B**). Interestingly, the

*S. uvarum* sub-genome of the Muri hybrid was closely related to the *S. uvarum* sub-genome of the CBS8614 triple hybrid (*S. cerevisiae* × *S. kudriavzevii* × *S. uvarum*). This triple hybrid was isolated from homemade apple cider produced in Brittany, France (Masneuf et al., 1998; Groth et al., 1999).

## Phenotypic Analysis of Muri and Closely Related Strains

Four *S. cerevisiae* strains, three *S. uvarum* strains, and three interspecific hybrids that were genetically closely related to the Muri hybrid were obtained (**Table 1**). A comparative phenotypic

analysis was performed on these eleven strains. Thirty-five traits were analyzed, including growth at various temperatures and on various carbon sources, fermentation performance in wort, and formation of aroma-active compounds. Hierarchical clustering with optimal leaf ordering based on the z-scores observed for the 35 traits grouped the 11 strains into three groups: one with the *S. cerevisiae* strains, one with the majority of the hybrid strains, and one with the *S. uvarum* strains and hybrid C997 (Figure 3). PCA on the same dataset also clustered the strains into the same three groups, with the exception that hybrid C997 was a clear outlier (Figure 4). The *S. cerevisiae* strains were separated from the *S. uvarum* strains along the first principal component, explaining 55% of the variance, while hybrids, with the exception of C997, grouped between them.

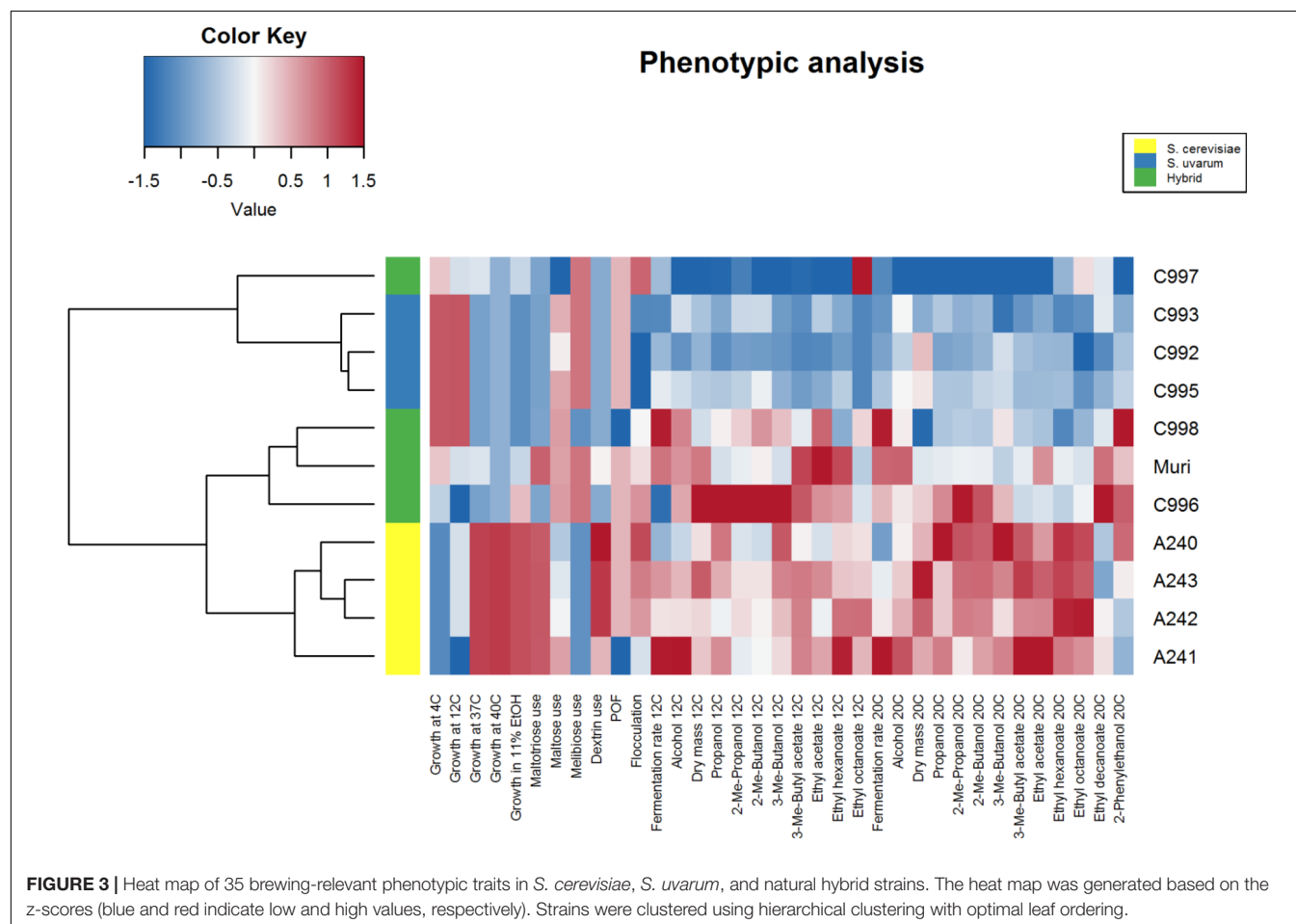
As was expected based on their brewing origin, the *S. cerevisiae* strains were associated with high flocculation, good fermentation performance in wort, high production of aroma compounds, and good growth at high temperatures and in 11% ethanol. The *S. uvarum* strains on the other hand were associated with good growth at lower temperatures and with melibiose as the sole carbon source. The hybrid strains tended to exhibit intermediate scores for the traits. Interestingly, the *S. cerevisiae* and *S. uvarum* strains that appeared most similar to the Muri hybrid based on the PCA (i.e., the shortest Euclidean distance to Muri in the first

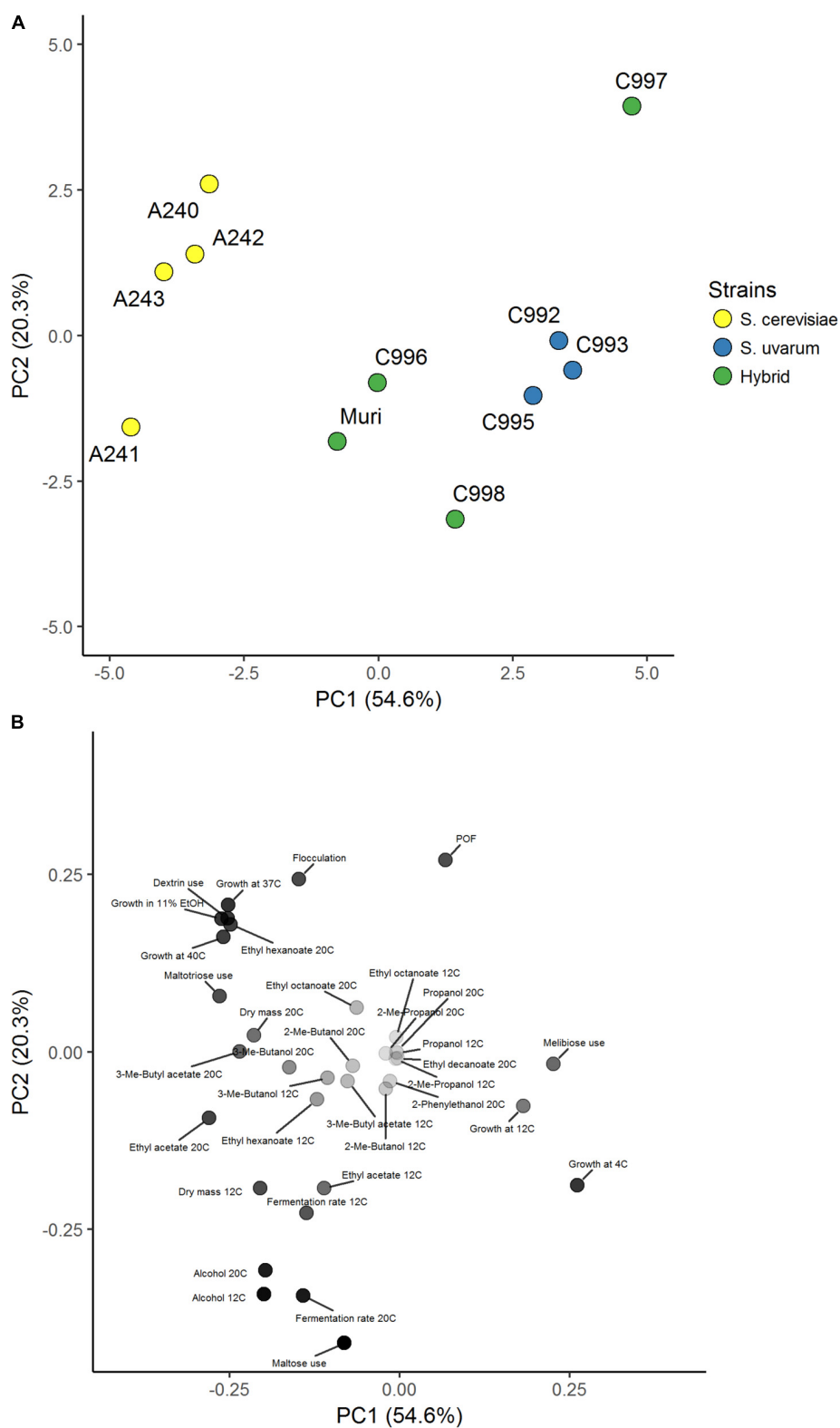
two principal components), A241 (Beer059) and C995 (ZP646), respectively, were also the strains that phylogenetically clustered closest to Muri's sub-genomes (Figure 2).

## Recreating the Muri Strain Through Interspecific Hybridization

In an attempt to recreate the Muri hybrid, *de novo* interspecific yeast hybrids between *S. cerevisiae* A241 and *S. uvarum* C995 were generated by mating. Spore-to-spore mating was chosen over rare mating as the hybridization approach, since the resulting hybrids tend to be allodiploid. High sporulation efficiency was obtained for both parent strains, and a total of 12 confirmed hybrids were obtained from 60 attempted crosses. Hybrid confirmation was performed on single cell isolates (obtained from re-streaking colonies three times) using ITS-PCR with *HaeIII* digestion and species-specific primers (Supplementary Figure S4). The four fastest growing hybrids were selected for further analysis (listed in Table 1).

In order to test how phenotypically similar the *de novo* hybrids were to the Muri hybrid, the comparative phenotypic analysis performed above was repeated with the four *de novo* hybrids, their parent strains, and the Muri hybrid. Hierarchical clustering with optimal leaf ordering based on the z-scores observed for



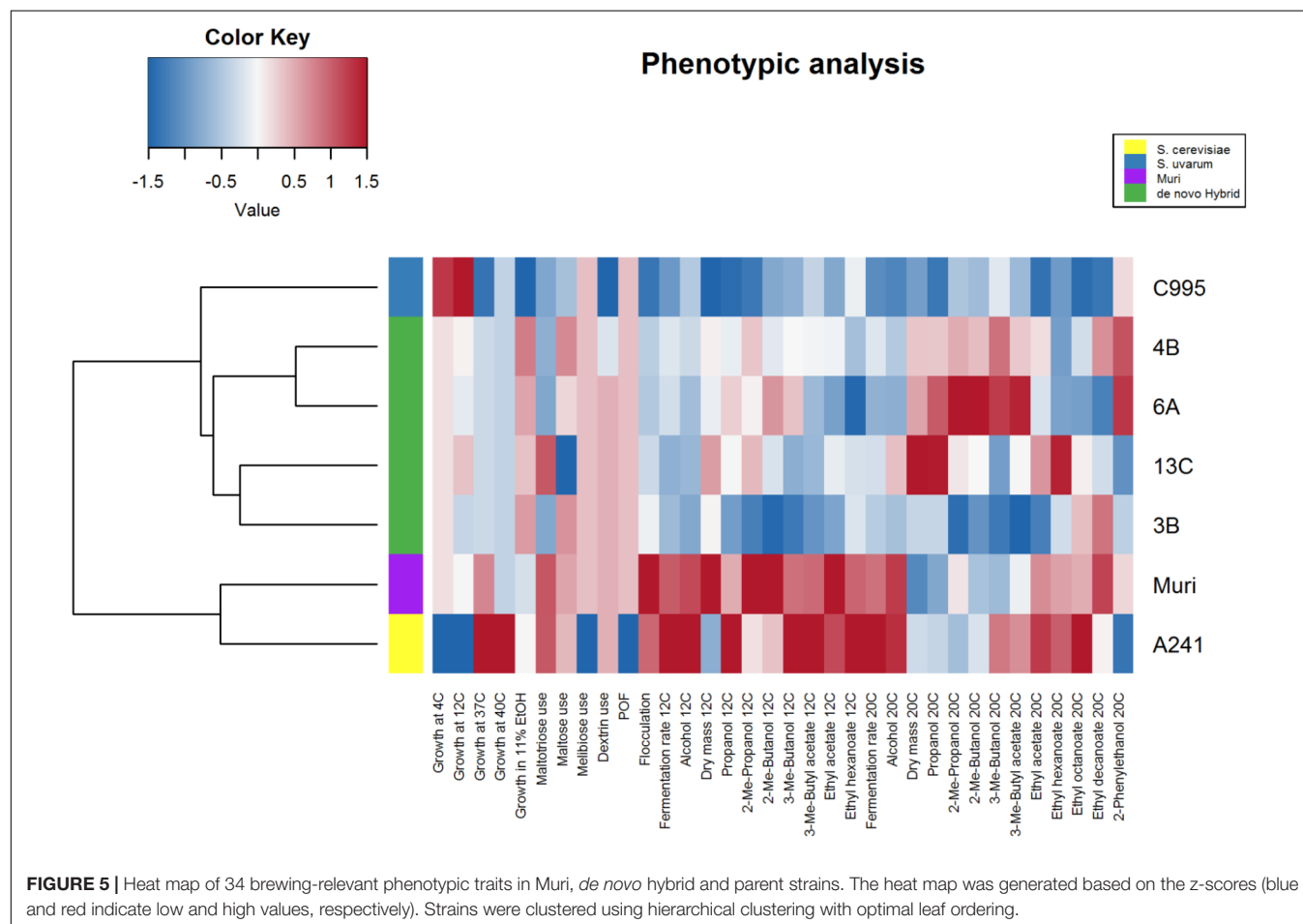


**FIGURE 4 |** Principal component analysis of 35 brewing-relevant phenotypic traits in *S. cerevisiae*, *S. uvarum*, and natural hybrid strains. **(A)** Score plot of the 11 yeast strains in the first two principal components explaining 75% of the variance. Strains are colored based on species (*S. cerevisiae*: yellow, *S. uvarum*: blue, and hybrid: green). **(B)** Loadings plot of the 35 phenotypic traits. The aroma compounds were weighted based on the weighting factors in **Supplementary Tables S4, S5**. Points are colored based on distance to origin from white (zero distance) to black.

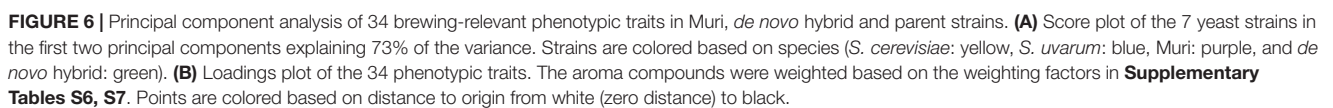
34 traits (ethyl octanoate concentrations from fermentations at 12°C were excluded because these compounds weren't detected for multiple strains) grouped the *de novo* hybrids close to the *S. uvarum* C995 parent, while Muri was grouped with the *S. cerevisiae* A241 parent (Figure 5). PCA on the same dataset revealed that the *de novo* hybrids grouped with the Muri hybrid (Figure 6). The two parent strains were separated from each other along the first principal component, explaining 57% of the variance, while both Muri and the *de novo* hybrids grouped between them. The hybrids were separated from the parent strains along the second principal component, explaining 16% of the variance. The hybrid strains tended to show intermediate scores in the majority of the traits, while best-parent heterosis was observed for biomass formation and ability to grow in the presence of 11% ethanol. The *de novo* hybrids had also inherited the *STA1* gene and the ability to ferment dextrin from *S. cerevisiae* A241 (Supplementary Figure S3B). Across the studied set of phenotypic traits, the *de novo* hybrids 4B and 13C were more similar to Muri than either of the parent strains (Supplementary Table S2).

While the *de novo* hybrids performed similarly to Muri, there were differences in their performances, particularly during the low-temperature wort fermentations (Figure 5). Interestingly, despite growing well at 4 and 12°C, the *de novo* hybrids fermented

wort slowly at 12°C. Furthermore, slight variation between the four *de novo* hybrids was also observed. Hybrid 13C, for example, was the only of the *de novo* hybrids which used maltotriose as efficiently as the A241 parent strain (Figure 5). We detected 23953 and 1490 heterozygous variants in the *S. cerevisiae* A241 (sequencing reads obtained from NCBI-SRA SRR5688171) and *S. uvarum* C995 (sequencing reads obtained from NCBI-SRA SRR1119163) parent strains, respectively. Of these, 10033 and 980, respectively, were detected in Muri as well. Therefore, one would expect spore siblings, and any resulting hybrids, to vary genetically and phenotypically. In addition, some diversity between the Muri hybrid and the *de novo* hybrids is also expected based on the sequence divergence between Muri's sub-genomes and the parent strains A241 and C995. In the *S. cerevisiae* sub-genome of Muri, a total of 2623 missense and nonsense mutations were identified that were not present in A241, while A241 contained 2128 missense and nonsense mutations not present in Muri. Gene ontology enrichment of the list of genes that these mutations affected revealed that, compared to A241, Muri appeared to be affected by mutations in genes related to regulation (Supplementary Table S3). Similarly, 3413 unique missense and nonsense mutations that were not present in C995 were identified in the *S. uvarum* sub-genome of Muri. The inheritance of mtDNA also varied between the four *de novo*







hybrids (**Supplementary Figure S4C**). Hybrid 6A had inherited mtDNA from the *S. uvarum* parent strain, while the other three strains had inherited mtDNA from the *S. cerevisiae* parent strain.

## DISCUSSION

While natural *S. cerevisiae* × *S. eubayanus* and *S. cerevisiae* × *S. kudriavzevii* interspecies hybrids are frequently used in beer fermentations (González et al., 2008; Gibson and Liti, 2015), limited reports exist describing the use of *S. cerevisiae* × *S. uvarum* hybrids in brewing. Here we characterize and attempt to reconstruct a unique *S. cerevisiae* × *S. uvarum* natural hybrid, Muri, that was isolated from a Norwegian farmhouse brewing culture. Whole genome sequencing and phylogenetic analysis revealed that the Muri hybrid's sub-genomes appeared to be closely related to domesticated *S. cerevisiae* and *S. uvarum* strains of British and Central European origin isolated from beer, cider and wine. Since the yeast was reportedly revived from an old yeast stock at the farmhouse (Garshol, 2013), we cannot exclude the possibility that the hybrid or one of its parents is a wild or contaminant yeast. However, the occurrence of wild *Saccharomyces* yeasts in Norway remains unexplored and no such strains are available for comparison. Interestingly, the *S. cerevisiae* sub-genome of the Muri strain does not appear to be related to other “kveik” isolates, which appear to be of “Beer 1” lineage rather than “Beer 2” (Preiss et al., 2018). It is therefore possible that the hybridization event to form Muri has occurred elsewhere, rather than at the farmhouse, or that not all kveik yeasts share the same ancestry. The allodiploid nature of Muri and the lack of structural rearrangements between the *S. cerevisiae* and the non-*S. cerevisiae* sub-genomes, are in contrast to that of other industrially used interspecific hybrids such as lager yeasts, which have been shown to exhibit considerable chromosomal copy number variations and rearrangements (van den Broek et al., 2015). This, together with “Beer 2” lineage of the *S. cerevisiae* sub-genome, suggests a more recent hybridization event in Muri compared to lager yeasts.

The Muri hybrid exhibited a range of phenotypic properties desirable for brewing. These included tolerance to both low and high temperatures, tolerance to a high ethanol concentration, efficient use of maltotriose, and formation of desirable aroma-active esters. In addition, the Muri hybrid possessed the *STA1* gene, encoding an extracellular glucoamylase enzyme (Yamashita et al., 1985), allowing it to utilize dextrin. This is a fairly unique property in brewing yeast and generally linked with beer spoilage (Meier-Dörnberg et al., 2018). However, it does allow for higher ethanol yield and the production of low-carbohydrate beer. The phenotype of Muri also makes it a potential candidate for other industrial fermentations, such as biofuel production or distilling, where good stress tolerance and high ethanol yield are desirable (Steensels et al., 2014b). Muri is not, however, a suitable candidate for lager beer production, as it possesses functional *PAD1* and *FDC1* genes. This allows it to produce POF, which are undesirable in lager beer. Interestingly, *S. cerevisiae* A241 (Beer059) did not produce POF as a result of premature stop codons in both *PAD1*

and *FDC1* (Gallone et al., 2016). It would therefore be possible to construct POF-negative *de novo* hybrids with rare mating and sporulation using a fertile allotetraploid intermediate, as has been demonstrated with *S. cerevisiae* × *S. eubayanus* hybrids (Krogerus et al., 2017).

In an attempt to reconstruct the Muri hybrid, we generated *de novo* hybrids between *S. cerevisiae* A241 and *S. uvarum* C995. As was expected based on previous research (Bellon et al., 2011; Dunn et al., 2013; Steensels et al., 2014a; Krogerus et al., 2015; Mertens et al., 2015; Snoek et al., 2015), these hybrids inherited traits from both parent strains. These hybrids also appeared to successfully replicate the phenotype of Muri, with the exception of efficient wort fermentation at 12°C. It is likely that this is a result of impaired maltose transport, and the absence of cold-tolerant maltose permeases (Vidgren et al., 2010). This variation could also result from heterozygosity in the parent strains and sequence divergence relative to Muri's sub-genomes. In addition, the origin of mtDNA in the *de novo* hybrids could influence their fitness (Wolters et al., 2018). The mtDNA in Muri appears to be from *S. uvarum*, while the majority of the *de novo* hybrids had inherited mtDNA from the *S. cerevisiae* parent. However, no obvious associations between the mitotype and phenotype were observed among the *de novo* hybrids. Recent studies with laboratory-generated *S. cerevisiae* × *S. uvarum* hybrids have revealed that transmission of mtDNA tends to be uniparental and a *S. cerevisiae* origin appears more common (Origone et al., 2018; Verspohl et al., 2018). However, mtDNA transmission appears strain-dependent.

A further potential cause of deviation between Muri and the *de novo* hybrids, is the impact of *S. eubayanus* introgressions. These introgressions are common in Holarctic *S. uvarum* strains, but the introgressed regions differ significantly depending on substrate origin and appear more common in domesticated strains (Almeida et al., 2014; Albertin et al., 2018). For example, an introgressed region (40 kb) from the left arm of *S. eubayanus* chromosome IV containing *FSY1*, a gene that enables efficient fructose transport, was detected in Muri, but not in *S. uvarum* C995 or any other *S. uvarum* strain studied by Almeida et al. (2014). The *S. eubayanus*-specific primers described by Pengelly and Wheals (2013) were designed based on the *S. eubayanus* allele of *FSY1*, which explains the PCR band for *S. eubayanus* that was detected in Muri (**Figure 1A**). The phenotypic impacts of *S. eubayanus* introgressions in *S. uvarum* have not yet been elucidated, but Albertin et al. (2018) speculate that they could be the most important source of genetic and phenotypic variability in Holarctic *S. uvarum* strains. These impacts should therefore be clarified in future studies. Furthermore, the use of long read sequencing technologies (e.g., PacBio or Nanopore) could also be applied to the Muri hybrid to generate end-to-end genome assemblies (Yue et al., 2017) in order to improve the detection of structural rearrangements and features not present in the reference genomes.

The approach used here, i.e., combining phylogenetic and phenotypic analysis to aid in reconstructing a natural hybrid, could be particularly useful if applied to lager yeast. While no single strain has yet been identified as the potential *S. cerevisiae* ale parent of lager yeast, recent whole genome studies suggest

that the last common ancestor of the *S. cerevisiae* sub-genome of lager yeasts is found among the “Beer 1” or “Ale beer” yeasts, close to the German wheat beer strains (Gonçalves et al., 2016; Monerawela and Bond, 2018). Such *de novo* hybrids could be used in evolutionary engineering studies to investigate which environmental conditions cause genomic changes that mimic those that have occurred in natural lager yeast. Saaz-type lager yeast, for example, have retained a larger fraction of the *S. eubayanus* sub-genome than the *S. cerevisiae* sub-genome (Dunn and Sherlock, 2008; Okuno et al., 2015), and it is still unclear how the environment has impacted its evolution. Evolutionary engineering studies with *de novo* interspecific hybrids have shown that either of the parental sub-genomes may be preferentially retained during stabilization (Piotrowski et al., 2012; Dunn et al., 2013; Lopandic et al., 2016; Smukowski Heil et al., 2017; Krogerus et al., 2018), and a stressful environment may cause more drastic changes. Exposing *S. cerevisiae* × *S. uvarum* hybrids to high temperatures resulted in the loss of the heat-sensitive *S. uvarum* sub-genome (Piotrowski et al., 2012), while exposing *S. cerevisiae* × *S. eubayanus* hybrids to high ethanol concentrations resulted in a greater loss of the *S. eubayanus* sub-genome (Krogerus et al., 2018). Evolutionary engineering of *de novo* hybrids constructed from the last common ancestors of natural lager yeasts’ sub-genomes could therefore help elucidate how various environmental stresses and nutrient limitations have shaped their genomes.

## CONCLUSION

In conclusion, we show that the Muri hybrid that was isolated from a Norwegian farmhouse beer is an allotriploid *S. cerevisiae* × *S. uvarum* hybrid. Phylogenetic analysis of the sub-genomes of this hybrid indicated that its *S. cerevisiae* parent was of brewing origin. The Muri strain possesses a range of industrially desirable phenotypic properties, making it an interesting candidate for not only brewing applications, but potentially various other industrial fermentations, such as biofuel production and distilling. In addition, we show that it is possible to mimic the phenotype of this natural hybrid, by constructing

*de novo* hybrids using parent strains closely related to Muri’s sub-genomes. This novel approach to studying natural hybrid strains has uses in both strain development and elucidating the evolutionary history of natural hybrids.

## AUTHOR CONTRIBUTIONS

KK, BG, and RP designed the experiments. KK conducted the experiments described in this study, and analyzed all data. RP contributed the Muri hybrid. KK and BG wrote the manuscript. All authors read and approved the final manuscript.

## FUNDING

This work was supported by the Alfred Kordelin Foundation, Svenska Kulturfonden - The Swedish Cultural Foundation in Finland, Suomen Kulttuurirahasto, and the Academy of Finland (Academy Project 276480).

## ACKNOWLEDGMENTS

We thank Kevin Verstrepen and Jan Steensels for providing the four *S. cerevisiae* Beer0XX strains, and Jose Paulo Sampaio and Pedro Almeida for sharing the consensus genotypes of the *S. uvarum* strains (Almeida et al., 2014). We also thank Frederico Magalhães for valuable comments during the study, Jarkko Nikulin for performing the phenolic off-flavor assay, Virve Vidgren for performing DNA extractions, Eero Mattila for wort preparation and other assistance in the VTT Pilot Brewery, and Aila Siltala for skilled technical assistance.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2018.02253/full#supplementary-material>

## REFERENCES

- Albertin, W., Chernova, M., Durrens, P., Guichoux, E., Sherman, D. J., Masneuf-Pomarede, I., et al. (2018). Many interspecific chromosomal introgressions are highly prevalent in Holarctic *Saccharomyces uvarum* strains found in human-related fermentations. *Yeast* 35, 141–156. doi: 10.1002/yea.3248
- Alexander, W. G., Peris, D., Pfannenstiel, B. T., Opulente, D. A., Kuang, M., and Hittinger, C. T. (2016). Efficient engineering of marker-free synthetic allotetraploids of *Saccharomyces*. *Fungal Genet. Biol.* 89, 10–17. doi: 10.1016/j.fgb.2015.11.002
- Almeida, P., Gonçalves, C., Teixeira, S., Libkind, D., Bontrager, M., Masneuf-Pomarede, I., et al. (2014). A Gondwanan imprint on global diversity and domestication of wine and cider yeast *Saccharomyces uvarum*. *Nat. Commun.* 5:4044. doi: 10.1038/ncomms5044
- Andrews, S. (2010). *FastQC: A Quality Control Tool for High Throughput Sequence Data*. Available at: <http://www.bioinformatics.babraham.ac.uk/projects/>
- Aquilani, B., Laureti, T., Poponi, S., and Secondi, L. (2015). Beer choice and consumption determinants when craft beers are tasted: An exploratory study of consumer preferences. *Food Qual. Prefer.* 41, 214–224. doi: 10.1016/j.foodqual.2014.12.005
- ASBC (2011). *ASBC Methods of Analysis*. St. Paul, MI: American Society of Brewing Chemists.
- Baker, E., Wang, B., Bellora, N., Peris, D., Hulfachor, A. B., Koshalek, J. A., et al. (2015). The genome sequence of *saccharomyces eubayanus* and the domestication of lager-brewing yeasts. *Mol. Biol. Evol.* 32, 2818–2831. doi: 10.1093/molbev/msv168
- Balakrishnan, R., Park, J., Karra, K., Hitz, B. C., Binkley, G., Hong, E. L., et al. (2012). YeastMine-An integrated data warehouse for *Saccharomyces cerevisiae* data as a multipurpose tool-kit. *Database* 2012:bar062. doi: 10.1093/database/bar062
- Belloch, C., Querol, A., García, M. D., and Barrio, E. (2000). Phylogeny of the genus *Kluyveromyces* inferred from the mitochondrial cytochrome-c oxidase II gene. *Int. J. Syst. Evol. Microbiol.* 50, 405–416. doi: 10.1099/00207713-50-1-405
- Bellon, J. R., Eglinton, J. M., Siebert, T. E., Pollnitz, A. P., Rose, L., De Barros Lopes, M., et al. (2011). Newly generated interspecific wine yeast hybrids

- introduce flavour and aroma diversity to wines. *Appl. Microbiol. Biotechnol.* 91, 603–612. doi: 10.1007/s00253-011-3294-3
- Benaglia, T., Chauveau, D., Hunter, D. R., and Young, D. (2009). Mixtools: an R package for analyzing finite mixture models. *J. Stat. Softw.* 32, 1–29. doi: 10.18637/jss.v032.i06
- Chen, X., Schulz-Trieglaff, O., Shaw, R., Barnes, B., Schlesinger, F., Källberg, M., et al. (2016). Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. *Bioinformatics* 32, 1220–1222. doi: 10.1093/bioinformatics/btv710
- Chiang, C., Layer, R. M., Faust, G. G., Lindberg, M. R., Rose, D. B., Garrison, E. P., et al. (2015). SpeedSeq: ultra-fast personal genome analysis and interpretation. *Nat. Methods* 12, 1–5. doi: 10.1038/nmeth.3505
- Christine, L. J., Marc, L., Catherine, D., Claude, E., Jean-Luc, L., Michel, A., et al. (2007). Characterization of natural hybrids of *Saccharomyces cerevisiae* and *Saccharomyces bayanus* var. *uvarum*. *FEMS Yeast Res.* 7, 540–549. doi: 10.1111/j.1567-1364.2007.00207.x
- Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., et al. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly* 6, 80–92. doi: 10.4161/fly.19695
- D'Hautcourt, O., and Smart, K. A. (1999). Measurement of brewing yeast flocculation. *J. Am. Soc. Brew. Chem.* 57, 123–128. doi: 10.1094/ASBCJ-57-0129
- Dunn, B., Paulish, T., Stanbery, A., Piotrowski, J., Koniges, G., Kroll, E., et al. (2013). Recurrent rearrangement during adaptive evolution in an interspecific yeast hybrid suggests a model for rapid introgression. *PLoS Genet.* 9:e1003366. doi: 10.1371/journal.pgen.1003366
- Dunn, B., and Sherlock, G. (2008). Reconstruction of the genome origins and evolution of the hybrid lager yeast *Saccharomyces pastorianus*. *Genome Res.* 18, 1610–1623. doi: 10.1101/gr.076075.108
- Fowle, Z. (2017). Kveik: The Hottest New Centuries-old Beer Yeast You've Never Heard of. *Draft Mag*. Available at: <http://draftmag.com/kveik-the-hottest-new-centuries-old-beer-yeast-youve-never-heard-of/>
- Gallone, B., Steensels, J., Baele, G., Maere, S., Verstrepen, K. J., Prahl, T., et al. (2016). Domestication and divergence of *Saccharomyces cerevisiae* beer yeasts. *Cell* 166, 1397 e16–1410 e16. doi: 10.1016/j.cell.2016.08.020
- García-Alcalde, F., Okonechnikov, K., Carbonell, J., Cruz, L. M., Götz, S., Tarazona, S., et al. (2012). Qualimap: evaluating next-generation sequencing alignment data. *Bioinformatics* 28, 2678–2679. doi: 10.1093/bioinformatics/bts503
- Garrison, E., and Marth, G. (2012). Haplotype-based variant detection from short-read sequencing. arXiv:1207.3907v2 [Preprint].
- Garshol, L. M. (2013). Kveik: Norwegian Farmhouse Yeast. Available at: <http://www.garshol.priv.no/blog/264.html>
- Garshol, L. M. (2016). *Gårdsøl - Det Norske Ølet*. Oslo: Cappelen Damm.
- Gibson, B., Geertman, J.-M. A., Hittinger, C. T., Krogerus, K., Libkind, D., Louis, E. J., et al. (2017). New yeasts—new brews: modern approaches to brewing yeast design and development. *FEMS Yeast Res.* 17:fox038. doi: 10.1093/femsyr/fox038
- Gibson, B., and Liti, G. (2015). *Saccharomyces pastorianus*: Genomic insights inspiring innovation for industry. *Yeast* 32, 17–27. doi: 10.1002/yea.3033
- Gonçalves, M., Pontes, A., Almeida, P., Barbosa, R., Serra, M., Libkind, D., et al. (2016). Distinct domestication trajectories in top-fermenting beer yeasts and wine yeasts. *Curr. Biol.* 26, 2750–2761. doi: 10.1016/j.cub.2016.08.040
- González, S. S., Barrio, E., and Querol, A. (2008). Molecular characterization of new natural hybrids of *Saccharomyces cerevisiae* and *S. kudriavzevii* in brewing. *Appl. Environ. Microbiol.* 74, 2314–2320. doi: 10.1128/AEM.01867-07
- Groth, C., Hansen, J., and Piskur, J. (1999). A natural chimeric yeast containing genetic material from three species. *Int. J. Syst. Bacteriol.* 49(Pt 4), 1933–1938. doi: 10.1099/00207713-49-4-1933
- Gu, Z., Gu, L., Eils, R., Schlesner, M., and Brors, B. (2014). Circlize implements and enhances circular visualization in R. *Bioinformatics* 30, 2811–2812. doi: 10.1093/bioinformatics/btu393
- Hahne, F., LeMeur, N., Brinkman, R. R., Ellis, B., Haaland, P., Sarkar, D., et al. (2009). flowCore: a Bioconductor package for high throughput flow cytometry. *BMC Bioinformatics* 10:106. doi: 10.1186/1471-2105-10-106
- Hahsler, M., Hornik, K., and Buchta, C. (2008). Getting Things in Order?: An Introduction to the R Package seriation. *J. Stat. Softw.* 25, 1–27. doi: 10.1088/0031-9120/44/5/F04
- Hebly, M., Brickwedde, A., Bolat, I., Driessen, M. R. M., de Hulster, E. A. F., van den Broek, M., et al. (2015). *S. cerevisiae* × *S. eubayanus* interspecific hybrid, the best of both worlds and beyond. *FEMS Yeast Res.* 15, 1–14. doi: 10.1093/femsyr/fov005
- Krogerus, K., Arvas, M., De Chiara, M., Magalhães, F., Mattinen, L., Oja, M., et al. (2016). Ploidy influences the functional attributes of de novo lager yeast hybrids. *Appl. Microbiol. Biotechnol.* 100, 7203–7222. doi: 10.1007/s00253-016-7588-3
- Krogerus, K., Holmström, S., and Gibson, B. (2018). Enhanced wort fermentation with de novo lager hybrids adapted to high-ethanol environments. *Appl. Environ. Microbiol.* 84, e2302–e2317. doi: 10.1128/AEM.02302-17
- Krogerus, K., Magalhães, F., Vidgren, V., and Gibson, B. (2015). New lager yeast strains generated by interspecific hybridization. *J. Ind. Microbiol. Biotechnol.* 42, 769–778. doi: 10.1007/s10295-015-1597-6
- Krogerus, K., Seppänen-Laakso, T., Castillo, S., and Gibson, B. (2017). Inheritance of brewing-relevant phenotypes in constructed *Saccharomyces cerevisiae* × *Saccharomyces eubayanus* hybrids. *Microb. Cell Fact.* 16:66. doi: 10.1186/s12934-017-0679-8
- Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* 27, 2987–2993. doi: 10.1093/bioinformatics/btr509
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352
- Lopandic, K. (2018). *Saccharomyces* interspecies hybrids as model organisms for studying yeast adaptation to stressful environments. *Yeast* 35, 21–38. doi: 10.1002/yea.3294
- Lopandic, K., Pfliegler, W. P., Tiefenbrunner, W., Gangl, H., Sipiczki, M., and Sterflinger, K. (2016). Genotypic and phenotypic evolution of yeast interspecies hybrids during high-sugar fermentation. *Appl. Microbiol. Biotechnol.* 100, 6331–6343. doi: 10.1007/s00253-016-7481-0
- Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* 17:10. doi: 10.14806/ej.17.1.200
- Masneuf, I., Hansen, J., Groth, C., Piskur, J., and Dubourdieu, D. (1998). New hybrids between *Saccharomyces sensu stricto* yeast species found among wine and cider production strains. *Appl. Environ. Microbiol.* 64, 3887–3892.
- Meier-Dörnberg, T., Kory, O. I., Jacob, F., Michel, M., and Hutzler, M. (2018). *Saccharomyces cerevisiae* variety diastaticus friend or foe?—spoilage potential and brewing ability of different *Saccharomyces cerevisiae* variety diastaticus yeast isolates by genetic, phenotypic and physiological characterization. *FEMS Yeast Res.* 18, foy023. doi: 10.1093/femsyr/foy023
- Meilgaard, M. C. (1982). Prediction of flavor differences between beers from their chemical composition. *J. Agric. Food Chem.* 30, 1009–1017. doi: 10.1021/jf00114a002
- Mertens, S., Steensels, J., Gallone, B., Souffriau, B., Malcorps, P., and Verstrepen, K. J. (2017). Rapid Screening Method for Phenolic Off-Flavor (POF) Production in Yeast. *J. Am. Soc. Brew. Chem.* 75, 318–323. doi: 10.1094/ASBCJ-2017-4142-01
- Mertens, S., Steensels, J., Saels, V., De Rouck, G., Aerts, G., and Verstrepen, K. J. (2015). A large set of newly created interspecific *Saccharomyces* hybrids increases aromatic diversity in lager beers. *Appl. Environ. Microbiol.* 81, 8202–8214. doi: 10.1128/AEM.02464-15
- Minh, B. Q., Nguyen, M. A. T., and Von Haeseler, A. (2013). Ultrafast approximation for phylogenetic bootstrap. *Mol. Biol. Evol.* 30, 1188–1195. doi: 10.1093/molbev/mst024
- Monerawela, C., and Bond, U. (2018). The hybrid genomes of *Saccharomyces pastorianus*: A current perspective. *Yeast* 35, 39–50. doi: 10.1002/yea.3250
- Muir, A., Harrison, E., and Wheals, A. (2011). A multiplex set of species-specific primers for rapid identification of members of the genus *Saccharomyces*. *FEMS Yeast Res.* 11, 552–563. doi: 10.1111/j.1567-1364.2011.00745.x
- Nguyen, H. V., Legras, J. L., Neuvéglise, C., and Gaillardin, C. (2011). Deciphering the hybridisation history leading to the lager lineage based on the mosaic genomes of *Saccharomyces bayanus* strains NBRC1948 and CBS380 T. *PLoS ONE* 6:e25821. doi: 10.1371/journal.pone.0025821
- Nguyen, L. T., Schmidt, H. A., Von Haeseler, A., and Minh, B. Q. (2015). IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32, 268–274. doi: 10.1093/molbev/msu300



- Nikulin, J., Krogerus, K., and Gibson, B. (2018). Alternative *Saccharomyces* interspecies hybrid combinations and their potential for low-temperature wort fermentation. *Yeast* 35, 113–127. doi: 10.1002/yea.3246
- Okuno, M., Kajitani, R., Ryusui, R., Morimoto, H., Kodama, Y., and Itoh, T. (2015). Next-generation sequencing analysis of lager brewing yeast strains reveals the evolutionary history of interspecies hybridization. *DNA Res* 23, 67–80. doi: 10.1093/dnares/dsv037
- Origone, A. C., Rodríguez, M. E., Oteiza, J. M., Querol, A., and Lopes, C. A. (2018). *Saccharomyces cerevisiae* × *Saccharomyces uvarum* hybrids generated under different conditions share similar winemaking features. *Yeast* 35, 157–171. doi: 10.1002/yea.3295
- Pengelly, R. J., and Wheals, A. E. (2013). Rapid identification of *Saccharomyces eubayanus* and its hybrids. *FEMS Yeast Res.* 13, 156–161. doi: 10.1111/1567-1364.12018
- Peris, D., Moriarty, R. V., Alexander, W. G., Baker, E., Sylvester, K., Sardi, M., et al. (2017). Hybridization and adaptive evolution of diverse *Saccharomyces* species for cellulosic biofuel production. *Biotechnol. Biofuels* 10:78. doi: 10.1186/s13068-017-0763-7
- Peter, J., De Chiara, M., Friedrich, A., Yue, J.-X., Pflieger, D., Bergström, A., et al. (2018). Genome evolution across 1,011 *Saccharomyces cerevisiae* isolates. *Nature* 556, 339–344. doi: 10.1038/s41586-018-0030-5
- Pham, T., Wimalasena, T., Box, W. G., Koivuranta, K., Storgårds, E., Smart, K. A., et al. (2011). Evaluation of ITS PCR and RFLP for differentiation and identification of brewing yeast and brewery “wild” yeast contaminants. *J. Inst. Brew.* 117, 556–568. doi: 10.1002/j.2050-0416.2011.tb00504.x
- Piotrowski, J. S., Nagarajan, S., Kroll, E., Stanbery, A., Chiotti, K. E., Kruckeberg, A. L., et al. (2012). Different selective pressures lead to different genomic outcomes as newly-formed hybrid yeasts evolve. *BMC Evol. Biol.* 12:46. doi: 10.1186/1471-2148-12-46
- Plech, M., de Visser, J. A. G. M., and Korona, R. (2014). Heterosis is prevalent among domesticated but not wild strains of *Saccharomyces cerevisiae*. *G* 3, 315–323. doi: 10.1534/g3.113.009381
- Preiss, R., Tyrawa, C., Krogerus, K., Garshol, L. M., and van der Merwe, G. (2018). Traditional Norwegian kveik are a genetically distinct group of domesticated *Saccharomyces cerevisiae* brewing yeasts. *Front. Microbiol.* 9:2137. doi: 10.3389/fmicb.2018.02137
- Quinlan, A. R., and Hall, I. M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842. doi: 10.1093/bioinformatics/btq033
- Rainieri, S., Kodama, Y., Kaneko, Y., Mikata, K., Nakao, Y., and Ashikari, T. (2006). Pure and mixed genetic lines of *Saccharomyces bayanus* and *Saccharomyces pastorianus* and their contribution to the lager brewing strain genome. *Appl. Environ. Microbiol.* 72, 3968–3974. doi: 10.1128/AEM.02769-05
- Roe, C., Smith, D. E., Williamson, C. H. D., Aziz, M., Keim, P., Hepp, C. M., et al. (2016). NASP: an accurate, rapid method for the identification of SNPs in WGS datasets that supports flexible input and output formats. *Microb. Genomics* 2:e000074. doi: 10.1099/mgen.0.000074
- Sato, M., Kishimoto, M., Watari, J., and Takashio, M. (2002). Breeding of brewer's yeast by hybridization between a top-fermenting yeast *Saccharomyces cerevisiae* and a cryophilic yeast *Saccharomyces bayanus*. *J. Biosci. Bioeng.* 93, 509–511. doi: 10.1016/S1389-1723(02)80101-r3
- Scannell, D. R., Zill, O. A., Rokas, A., Payen, C., Dunham, M. J., Eisen, M. B., et al. (2011). The awesome power of yeast evolutionary genetics: new genome sequences and strain resources for the *Saccharomyces sensu stricto* Genus. *G3* 1, 11–25. doi: 10.1534/g3.111.000273
- Smukowski Heil, C. S., DeSevo, C. G., Pai, D. A., Tucker, C. M., Hoang, M. L., and Dunham, M. J. (2017). Loss of heterozygosity drives adaptation in hybrid yeast. *Mol. Biol. Evol.* 34, 1596–1612. doi: 10.1093/molbev/msx098
- Snoek, T., Picca Nicolino, M., Van den Bremt, S., Mertens, S., Saels, V., Verplaetse, A., et al. (2015). Large-scale robot-assisted genome shuffling yields industrial *Saccharomyces cerevisiae* yeasts with increased ethanol tolerance. *Biotechnol. Biofuels* 8:32. doi: 10.1186/s13068-015-0216-0
- Steensels, J., Meersman, E., Snoek, T., Saels, V., and Verstrepen, K. J. (2014a). Large-scale selection and breeding to generate industrial yeasts with superior aroma production. *Appl. Environ. Microbiol.* 80, 6965–6975. doi: 10.1128/AEM.02235-14
- Steensels, J., Snoek, T., Meersman, E., Nicolino, M. P., Voordeckers, K., and Verstrepen, K. J. (2014b). Improving industrial yeast strains: Exploiting natural and artificial diversity. *FEMS Microbiol. Rev.* 38, 947–995. doi: 10.1111/1574-6976.12073
- van den Broek, M., Bolat, I., Nijkamp, J. F., Ramos, E., Luttik, M. A. H., Koopman, F., et al. (2015). Chromosomal copy number variation in *Saccharomyces pastorianus* is evidence for extensive genome dynamics in industrial lager brewing strains. *Appl. Environ. Microbiol.* 81, 6253–6267. doi: 10.1128/AEM.01263-15
- Verspohl, A., Pignedoli, S., and Giudici, P. (2018). The inheritance of mitochondrial DNA in interspecific *Saccharomyces* hybrids and their properties in winemaking. *Yeast* 35, 173–187. doi: 10.1002/yea.3288
- Vidgren, V., Multanen, J. P., Ruohonen, L., and Londesborough, J. (2010). The temperature dependence of maltose transport in ale and lager strains of brewer's yeast. *FEMS Yeast Res.* 10, 402–411. doi: 10.1111/j.1567-1364.2010.00627.x
- Wolters, J. F., Charron, G., Gaspary, A., Landry, C. R., Fiumera, A. C., and Fiumera, H. L. (2018). Mitochondrial recombination reveals mito-mito epistasis in yeast. *Genetics* 209, 307–319. doi: 10.1534/genetics.117.300660
- Yamashita, I., Maemura, T., Hatano, T., and Fukui, S. (1985). Polymorphic extracellular glucoamylase genes and their evolutionary origin in the yeast *Saccharomyces diastaticus*. *J. Bacteriol.* 161, 574–582.
- Yamauchi, H., Yamamoto, H., Shibano, Y., Amaya, N., and Saeki, T. (1998). Rapid methods for detecting *Saccharomyces diastaticus*, a beer spoilage yeast, using the polymerase chain reaction. *J. Am. Soc. Brew. Chem.* 56, 58–63. doi: 10.1094/ASBCJ-56-0058
- Yue, J. X., Li, J., Aigrain, L., Hallin, J., Persson, K., Oliver, K., et al. (2017). Contrasting evolutionary genome dynamics between domesticated and wild yeasts. *Nat. Genet.* 49, 913–924. doi: 10.1038/ng.3847

**Conflict of Interest Statement:** RP was employed by Escarpment Laboratories Inc., BG was employed by VTT Technical Research Centre of Finland Ltd., and KK is affiliated with VTT Technical Research Centre of Finland Ltd. All authors declare no competing interests. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright © 2018 Krogerus, Preiss and Gibson. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# The Paralogous Genes *PDR18* and *SNQ2*, Encoding Multidrug Resistance ABC Transporters, Derive From a Recent Duplication Event, *PDR18* Being Specific to the *Saccharomyces* Genus

Cláudia P. Godinho<sup>†</sup>, Paulo J. Dias<sup>†</sup>, Elise Ponçot and Isabel Sá-Correia\*

<sup>†</sup> iBB-Institute for Bioengineering and Biosciences, Department of Bioengineering, Instituto Superior Técnico, Universidade de Lisboa, Lisbon, Portugal

## OPEN ACCESS

### Edited by:

Feng Gao,  
Tianjin University, China

### Reviewed by:

Adam Michael Reitzel,  
University of North Carolina  
at Charlotte, United States  
Kenneth Wolfe,  
University College Dublin, Ireland

### \*Correspondence:

Isabel Sá-Correia  
isacorreia@tecnico.ulisboa.pt

<sup>†</sup>These authors have contributed  
equally to this work

### Specialty section:

This article was submitted to  
Evolutionary and Genomic  
Microbiology,  
a section of the journal  
Frontiers in Genetics

**Received:** 26 June 2018

**Accepted:** 26 September 2018

**Published:** 15 October 2018

### Citation:

Godinho CP, Dias PJ, Ponçot E  
and Sá-Correia I (2018) The  
Paralogous Genes *PDR18* and *SNQ2*,  
Encoding Multidrug Resistance ABC  
Transporters, Derive From a Recent  
Duplication Event, *PDR18* Being  
Specific to the *Saccharomyces*  
Genus. *Front. Genet.* 9:476.  
doi: 10.3389/fgene.2018.00476

Pleiotropic drug resistance (PDR) family of ATP-binding cassette (ABC) transporters play a key role in the simultaneous acquisition of resistance to a wide range of structurally and functionally unrelated cytotoxic compounds in yeasts. *Saccharomyces cerevisiae* Pdr18 was proposed to transport ergosterol at the plasma membrane, contributing to the maintenance of adequate ergosterol content and decreased levels of stress-induced membrane disorganization and permeabilization under multistress challenge leading to resistance to ethanol, acetic acid and the herbicide 2,4-D, among other compounds. *PDR18* is a paralog of *SNQ2*, first described as a determinant of resistance to the chemical mutagen 4-NQO. The phylogenetic and neighborhood analysis performed in this work to reconstruct the evolutionary history of *ScPDR18* gene in Saccharomycetaceae yeasts was focused on the 214 Pdr18/Snq2 homologs from the genomes of 117 strains belonging to 29 yeast species across that family. Results support the idea that a single duplication event occurring in the common ancestor of the *Saccharomyces* genus yeasts was at the origin of *PDR18* and *SNQ2*, and that by chromosome translocation *PDR18* gained a subtelomeric region location in chromosome XIV. The multidrug/multixenobiotic phenotypic profiles of *S. cerevisiae* *pdr18Δ* and *snq2Δ* deletion mutants were compared, as well as the susceptibility profile for *Candida glabrata* *snq2Δ* deletion mutant, given that this yeast species has diverged previously to the duplication event on the origin of *PDR18* and *SNQ2* genes and encode only one Pdr18/Snq2 homolog. Results show a significant overlap between ScSnq2 and CgSnq2 roles in multidrug/multixenobiotic resistance (MDR/MXR) as well as some overlap in azole resistance between ScPdr18 and CgSnq2. The fact that ScSnq2 and ScPdr18 confer resistance to different sets of chemical compounds with little overlapping is consistent with the subfunctionalization and neofunctionalization of

these gene copies. The elucidation of the real biological role of *ScSNQ2* will enlighten this issue. Remarkably, *PDR18* is only found in *Saccharomyces* genus genomes and is present in almost all the recently available 1,000 deep coverage genomes of natural *S. cerevisiae* isolates, consistent with the relevant encoded physiological function.

**Keywords:** ABC transporters, *PDR18* and *SNQ2*, phylogenetic and genomic neighborhood analyzes, comparative genomics and evolution, multidrug resistance

## INTRODUCTION

Several ATP-binding cassette (ABC) transporters that catalyze the ATP-dependent active solute transport across cell membranes in yeasts are associated with multidrug/multixenobiotic resistance (MDR/MXR) (Jungwirth and Kuchler, 2006; Monk and Goffeau, 2008; Piecuch and Oblak, 2014). Although these transporters are usually considered drug/xenobiotic pumps, evidence is arising supporting the idea that their involvement in MDR/MXR may result from their specific and, in general, not yet determined biological role in the active transport of physiological substrates (Prasad and Panwar, 2004; Cabrito et al., 2011; Prasad et al., 2016; Godinho et al., 2018). Moreover, the presence of a large number of ABC transporters involved in MDR/MXR in the genomes of yeasts and other organisms, from bacteria to man, also strongly suggests that these transporters may play important physiological roles even in the absence of the cytotoxic compounds to which they confer resistance. For example, the yeast ABC-MDR/MXR family of transporters that includes the Pleiotropic Drug Resistance (PDR) transporters, perform endogenous activities beyond their proposed role as drug exporters, in particular as lipid transporters (Balzi and Goffeau, 1995; Borst et al., 2000; Pomorski et al., 2004; Panwar et al., 2008; Van Meer et al., 2008; Prasad et al., 2016). *Saccharomyces cerevisiae* genome encodes 10 PDR proteins: *Adp1*, *Aus1*, *Pdr5*, *Pdr10*, *Pdr11*, *Pdr12*, *Pdr15*, *Pdr18*, *Snq2*, and *YOL075c* (Decottignies and Goffeau, 1997; Paulsen et al., 1998; Paumi et al., 2009). A combined phylogeny and neighborhood analysis of the evolution of these ABC transporters in nine yeast species belonging to the subphylum Saccharomycotina has shown that *Pdr18* is a paralog of *Snq2* and that *SNQ2* and *PDR18* genes reside in unshared chromosomal environments (Seret et al., 2009). However, the small number of yeast species genomes available when this study was performed did not allow a firm conclusion concerning the hypothesized gene duplication event at the origin of these two PDR gene sub-lineages. In fact, it was doubtful whether the duplication event remounted to the whole genome duplication (WGD) event or if it was an independent event that occurred post-WGD (Seret et al., 2009).

The *S. cerevisiae* plasma membrane transporter *Pdr18* was described as a MDR/MXR determinant required for ergosterol transport at the plasma membrane level (Cabrito et al., 2011; Teixeira et al., 2012; Godinho et al., 2018). *Pdr18* expression was found to lead to increased yeast tolerance to the herbicides 2,4-dichlorophenoxyacetic acid (2,4-D), 2-methyl-4-chlorophenoxyacetic acid (MCPA), and barban, the agricultural fungicide mancozeb, the metal cations  $Zn^{2+}$ ,  $Mn^{2+}$ ,  $Cu^{2+}$ , and  $Cd^{2+}$  (Cabrito et al., 2011) and to ethanol

(Teixeira et al., 2012) and acetic acid (Godinho et al., 2018). The involvement of *Pdr18* in the maintenance of yeast plasma membrane ergosterol content under 2,4-D or acetic acid stresses was related with its role as a determinant of resistance to multiple stresses in yeast (Cabrito et al., 2011; Teixeira et al., 2012; Godinho et al., 2018). A coordinated response involving the transcriptional activation of *PDR18* and several ergosterol biosynthetic genes was found to occur in response to acetic acid stress, strongly suggesting the involvement of *Pdr18* in ergosterol homeostasis in stressed yeast cells (Godinho et al., 2018). Moreover, the proposed role for *Pdr18* in ergosterol homeostasis was demonstrated to be important to counteract acetic acid-induced decrease of plasma membrane lipid order, increase of plasma membrane non-specific permeability and decrease of transmembrane electrochemical potential (Godinho et al., 2018).

The *PDR18* paralog gene *SNQ2* was first described based on its involvement in yeast resistance to the chemical mutagens 4-nitroquinoline 1-oxide (4-NQO) and triaziquone (Servos et al., 1993). Later, several other publications extended the range of compounds to which *SNQ2* expression confers increased tolerance in yeast (Servos et al., 1993; Hirata et al., 1994; Mahé et al., 1996a,b; Miyahara et al., 1996; Kolaczowski et al., 1998; Ververidis et al., 2001; Piper et al., 2003; Wehrschütz-Sigl et al., 2004; van Leeuwen et al., 2012; Ling et al., 2013; Nishida et al., 2013; Snider et al., 2013; Tsujimoto et al., 2015). Although no role in lipid homeostasis was demonstrated for *Snq2* transporter, it was shown that *Snq2* is involved in the alleviation of estradiol toxicity in *S. cerevisiae* (Mahé et al., 1996a). For this reason, it was hypothesized that *Snq2* could also have affinity for lipid transport, especially for the estradiol structurally related molecule ergosterol (Mahé et al., 1996a; Kuchler et al., 1997).

Gene duplication is considered to be one of the most important forces driving the evolution of genetic functional innovation and genes encoding membrane transporters are one of the functional gene categories that exhibit high number of duplication events (Ohno, 1970; Zhang, 2003; Taylor and Raes, 2004). In yeast, for example, this is the case of genes encoding proteins of the Major Facilitator Superfamily (MFS) of transporters (Dias et al., 2010; Dias and Sá-Correia, 2013, 2014). Also, transporters from the PDR family were involved in multiple gene duplications and gene losses occurring during their evolutionary history (Seret et al., 2009; Kovalchuk and Driessen, 2010). After gene duplication, it is possible the occurrence of the inactivation of one of the copies (pseudogenization), the maintenance of the two copies (dosage effect), the adoption of part of the function or of the expression pattern of their parental gene (subfunctionalization), or the acquisition of a related or

new function (neofunctionalization) (Zhang, 2003; Conant and Wolfe, 2008). The duplicate genes are maintained in the genome depending upon their function, mode of duplication, expression rate and the organism taxonomic lineage (Taylor and Raes, 2004).

To better understand the duplication event that gave rise to the *PDR18* and *SNQ2* paralog genes, the evolutionary history of *PDR18* was reconstructed in this work by combining phylogenetic tree building methodology with gene neighborhood analysis in 117 strains genomes belonging to 29 species across the Saccharomycetaceae family. A systematic multidrug/multixenobiotic phenotypic profiling of *S. cerevisiae* deletion mutants for *PDR18* or *SNQ2* genes was also performed. Given that the genomes of the post-WGD *Candida glabrata* species encode only one Pdr18/Snq2 homolog (*CgSNQ2*) the susceptibility profiling for the *Cgsnq2Δ* deletion mutant was also examined to get additional insights into the functional divergence of *S. cerevisiae* Pdr18 and Snq2 and the common ancestral gene on the origin of the post-WGD single duplication event.

## MATERIALS AND METHODS

### Identification of the Homologs of *S. cerevisiae* Pdr18 and Snq2 Proteins in Hemiascomycete Yeast Genomes

A total of 1,110,525 Open Reading Frames (ORFs) encoded in the genomes of 171 strains belonging to 68 different yeast species of the subphylum Saccharomycotina were retrieved and compiled in a local Genome DB (Dias and Sá-Correia, 2013, 2014). A second in-house database built in this work, BLASTp DB, comprises the output values of the blastp algorithm (Altschul et al., 1990) for each pairwise entry of all possible combinations between the translated ORFs compiled in the Genome DB, including length of the alignment, e-value, percentage of identity and similarity, and alignment score. The blastp algorithm used a gapped alignment with the following parameters: open gap (−1), extend gap (−1), threshold for extending hits (11), and word size (3). This approach generated a total of 328 million pairwise alignments. The ORFs encoding PDR proteins were identified through the adoption of a network traversal strategy considering the whole set of blastp pairwise relationships as a network that was subsequently traversed at a range of different e-value thresholds (Dias and Sá-Correia, 2013, 2014; Palma et al., 2017). The *S. cerevisiae* Pdr18 was selected to represent the PDR *sensu stricto* (Seret et al., 2009) transporters and was used as a starting node for network traversal. The *S. cerevisiae* Adp1 and Y075c were selected to represent the PDR *sensu lato* transporters (Seret et al., 2009) and also used as starting nodes in independent network traversals. The disjoint sets of translated ORFs obtained from these three traversals were merged. The amino acid sequences of these ORFs were analyzed for potential false positive members of the PDR protein family, protein fragments and/or frameshifts. The remaining amino acid sequences were aligned using MUSCLE software (Edgar, 2004). Subsequently, the protdist and neighbor algorithms made

available by the PHYLIP software suite (Felsenstein, 1989) were used to construct a preliminary phylogenetic tree. Using as reference the cluster of residence of the Snq2 and Pdr18 proteins, the branch comprising the homologs of these two *S. cerevisiae* proteins in this tree was identified. For species abbreviation a four letters code is used, composed by the first two letter of the genus and species. The number displayed after the first four letters is used to abbreviate the strain name when the genome of more than one strain from a given species was examined. To standardize the annotation used, translated ORFs are represented by small letters.

### Phylogenetic Analysis and Tree Construction

The MUSCLE software suite (Edgar, 2004) was used to build a multiple alignment of the amino acid sequences of the Snq2 and Pdr18 proteins encoded in Saccharomycetaceae yeasts that was analyzed using the Jalview 2.9 software suite (Clamp et al., 2004; Waterhouse et al., 2009). The processing of these sequences did not involve any step of masking or trimming. The “read.fasta” and “write.nexus.data” functions made available by the seqinr 3.4-5 (Charif et al., 2007) and by the ape 4.1 R packages (Paradis et al., 2004; Popescu et al., 2012), respectively, were used to convert the multiple alignment in fasta format into a nexus file that was subsequently fed into MrBayes 3.2.6 (Huelsenbeck and Ronquist, 2001; Ronquist and Huelsenbeck, 2003), a Bayesian Markov chain Monte Carlo (MCMC) package for phylogenetic analysis. The Message Passing Interface (MPI) version of MrBayes (Altekar et al., 2004) was used to speed up phylogeny computation using Metropolis coupling of MCMC sampling. The MCMC simulations used 100,000 generations, coupling one “cold” chain together with nine heated chains. Two independent runs of MCMC sampling (each started from two distinct random trees) confirmed parameter convergence of the posterior probability distribution. The option of estimating the fixed-rate amino acid prior model made available by MrBayes was used, allowing the MCMC sampler to explore all of the nine available models by regularly proposing new ones (upon parameter convergence, each model contributes to the results in proportion to its posterior probability) and rate variation over sites was assumed to follow a gamma distribution. The remaining MrBayes MPI parameters were set to default values. The PhyML 3.0 software suite (Guindon et al., 2010) was used to construct a Maximum Likelihood (ML) derived phylogenetic tree to confirm the gathered results. The PhyML parameters and models used as input were the default ones, with exception of the method for searching the optimal phylogenetic tree, where the Subtree Pruning and Regrafting (SPR) were used as algorithm instead of Nearest Neighbor Interchange (NNI). The phylogenetic trees obtained were analyzed using the visualization software Dendroscope 3.5.7 (Huson et al., 2007; Huson and Scornavacca, 2012). The MrBayes clade credibility score and PhyML bootstrapping values calculated for each internal node of the Bayesian and ML trees, respectively, were inspected using either the FigTree 1.4.3 software suite<sup>1</sup> or PhyloTree 0.1

<sup>1</sup><http://tree.bio.ed.ac.uk/software/figtree/>



package<sup>2</sup> installed in Cytoscape 2.8.3 (Shannon et al., 2003). The bootstrap score measures the confidence level of each clade in the consensus phylogenetic tree by repeated sub-sampling data from the original data set and determining the empirical frequency that each of these clades have in the whole set of phylogenetic trees originated from these sub-samples. The clade credibility score measures the posterior probability of each clade in the consensus phylogenetic tree by determining the frequency that each of these clades has in the whole set of phylogenetic tree sampled using the estimated parameter values.

## Gene Neighborhood Analysis

A chromosome block of 30 neighboring genes, 15 on each side of the pair of homologous genes under analysis, was selected to assess the conservation of the chromosome region where the members of the Snq2/Pdr18 subfamily reside (Seret et al., 2009; Dias et al., 2010). Scripting in the R language was used to retrieve 15 neighbor genes on each side of the query genes as well as the corresponding sequence clustering classification from Genome DB. The classification of each of the 30 genes neighboring each query gene was obtained using a conservative blastp *e*-value of E-50 to limit the number of false positive sequences gathered together with true cluster members. When dubious synteny connections between genes needed corroboration, the amino acid sequence clustering was performed at a less restrictive *e*-value threshold of E-40. The existence of synteny between query genes was verified through the analysis of network topology (number of shared neighbor pairs) and the biological information associated with the corresponding edges. Three sources of biological information were used as independent evidence confirming the strength of the synteny between members of the Snq2 and Pdr18 protein subfamily (Dias et al., 2010): (1) distance of the neighbors in relation to the query genes, (2) similarity of the amino acid sequences of the shared neighbors, and (3) total number of members comprised in the cluster of amino acid sequence to which the homologous neighbors belong to; sequence clusters comprising a small number of members are more reliable as synteny evidence since the probability that two homologous neighbors being in the vicinity of two query genes by chance is small.

## Susceptibility Phenotypes of *S. cerevisiae* *pdr18Δ* and *snq2Δ* and *C. glabrata* *snq2Δ* Deletion Mutants

### Strains, Media, and Growth Conditions

*S. cerevisiae* BY4741 (*MATa*, *his3Δ1*, *leu2Δ0*, *met15Δ0*, *ura3Δ0*) and the derived deletion mutants *pdr18Δ* and *snq2Δ* were obtained from EUROSCARF collection. *C. glabrata* BPY55 (clinical isolate) and the derived deletion mutant *snq2Δ* built using the SAT1 flipper system were kindly provided by Professor Dominique Sanglard, Institut de Microbiologie, Centre Hospitalier Universitaire Vaudois (CHUV), Lausanne, Switzerland.

Cultivation of *S. cerevisiae* strains was performed in MM4 medium, containing 1.7 g/L yeast nitrogen base without amino acids and ammonium sulfate (Difco, Detroit, MI, United States), 20 g/L glucose (Merck, Darmstadt, Germany), 2.65 g/L (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> (Panreac AppliChem, CT, United States), 20 mg/L L-methionine, 20 mg/L L-histidine (both from Merck, Darmstadt, Germany), 60 mg/L L-leucine and 20 mg/L L-uracil (both from Sigma, St. Louis, MO, United States). *C. glabrata* strains were cultivated in MM medium, with the composition of MM4 medium, without supplementation with amino acids and uracil. YPD medium contained 20 g/L glucose, 20 g/L Bacto<sup>TM</sup> Peptone and yeast extract (both from BD Biosciences, Franklin Lakes, NJ, United States). Solid media were prepared by the addition of 20 g/L agar (Iberagar, Barreiro, Portugal) to the different liquid media. Media pH were adjusted to 4.5 with HCl. Growth in liquid media was performed at 30°C with orbital agitation (250 rpm).

### Susceptibility Tests

Growth susceptibility tests of *S. cerevisiae* parental strain BY4741, the corresponding *pdr18Δ* and *snq2Δ* deletion mutants and of *C. glabrata* BPY55 and derived deletion mutant *snq2Δ* to a wide range of growth inhibitory compounds was evaluated by spot assays. Yeast cell suspensions used for the spot assays were prepared from mid-exponential cell cultures grown in liquid media MM4 (*S. cerevisiae*), MM (*C. glabrata*), or YPD (both strains) by harvesting (5,000 rpm, 5 min) and resuspending the cells in sterile ddH<sub>2</sub>O to an OD<sub>600 nm</sub> of 0.25 followed by four serial dilutions of 1:5 each. These cell suspensions were plated as 4 μL spots onto the surface of MM4 (*S. cerevisiae*) or MM (*C. glabrata*) or YPD (both strains) at pH 4.5 solid media, supplemented or not with the toxic compound to be tested. The plates were incubated at 30°C for 72 h and pictures were taken every 24 h. The toxic compounds tested and the selected concentrations for the screening of *S. cerevisiae* and *C. glabrata* strains in MM4 and MM are listed in **Table 1**, and those selected for experiments in YPD are indicated close to the respective pictures. The susceptibility of the deletion mutants was assessed by comparison of their growth performance with the growth performance of the corresponding parental strain in the presence of the different cytotoxic compounds tested. Moderately reduced growth of the deletion mutants compared with the parental strain was classified as minus (−), marked reduced growth as a double minus (− −), improved growth as a plus (+) and identical growth as zero (0). Results are representative from three independent experiments.

## RESULTS

### Identification of Snq2 and Pdr18 Proteins Encoded in the Saccharomycetaceae Yeast Genomes

To identify the PDR proteins encoded in the examined yeast genomes, the network representing the amino acid

<sup>2</sup><http://apps.cytoscape.org/apps/phyloree>

**TABLE 1 |** Susceptibility of the deletion mutants *pdr18Δ* and *snq2Δ* of *S. cerevisiae* BY4741 and *snq2Δ* of *C. glabrata* BPY55 to a wide range of chemical compounds supplemented in minimal media.

Class	Growth inhibitory compounds			<i>S. cerevisiae</i>		<i>C. glabrata</i>
	Compound	Selected concentrations ( <i>S. cerevisiae</i> / <i>C. glabrata</i> )	Manufacturer	<i>snq2Δ</i>	<i>pdr18Δ</i>	<i>snq2Δ</i>
Weak acids	Acetic acid	60 mM/60 mM	Fluka, United States	0	— Godinho et al., 2018	0
	Benzoic acid	1.2 mM/1 mM	Sigma, United States	0	— —	0
	Propionic acid	15 mM	Sigma, United States	0	— —	
	Formic acid	1 mM	Sigma, United States	0	— —	
	Sorbic acid	30 mM	Fluka, United States	0	—	
Herbicides	2,4-D	2.5 mM/2 mM	Sigma, United States	0	— Cabrito et al., 2011	0
	MCPA	1.5 mM/1 mM	Sigma, United States	0	—	0
					Cabrito et al., 2011	
	Barban	0.02 mM	Sigma, United States	—	— Cabrito et al., 2011	
	Alachlor	1.5 mM	Fluka, United States	—	—	
Alcohols	Metolachlor	1.1 mM	Fluka, United States	— —	—	
	Ethanol	1.37 M/1.37 M	Merck, United States	0	— Teixeira et al., 2012	0
	Propanol	1 M	Acros, United States	0	—	
	1-Butanol	0.21 M	Acros, United States	0	—	
	2-Butanol	0.5 M	Acros, United States	0	— —	
	1,3-Propanediol	0.98 M	Acros, United States	0	+	
	1,4-Butanediol	0.9 M/0.9 M	Acros, United States	0	+	0
Clinical and agricultural fungicides	Itraconazole	14.2 μM/71 μM	Sigma, United States	0	— —	+
	Ketoconazole	28.2 μM/226 μM	Sigma, United States	0	— —	0
	Fluconazole	98 μM/1.96 mM	Sigma, United States	0	— —	0
	Clotrimazole	8.7 μM/116 μM	Sigma, United States	0	— —	0
	Miconazole	1.2 μM/48 μM	Sigma, United States	0	— —	+
	Tioconazole	1.3 μM	Sigma, United States	0	— —	0
	Amphotericin B	0.33 μM/0.33 μM	Sigma, United States	0	— —	0
	Mancozeb	3.7 μM	Sigma, United States	0	— Cabrito et al., 2011	

(Continued)

TABLE 1 | Continued

Class	Growth inhibitory compounds			<i>S. cerevisiae</i>		<i>C. glabrata</i>
	Compound	Selected concentrations ( <i>S. cerevisiae</i> / <i>C. glabrata</i> )	Manufacturer	<i>snq2Δ</i>	<i>pdr18Δ</i>	<i>snq2Δ</i>
Cations	Cd <sup>2+</sup>	16 μM	Merck, United States	0	— Cabrito et al., 2011	
	Li <sup>+</sup>	9 mM/80 mM	Sigma, United States	— Miyahara et al., 1996	0	0
	Mn <sup>2+</sup>	1 mM	Merck, United States	— Miyahara et al., 1996	—	
	Mg <sup>2+</sup>	90 mM	Sigma, United States	0	0	
	Cu <sup>2+</sup>	0.13 mM/0.25 mM	Merck, United States	0	—	—
Mutagens	4-NQO	5.26 μM/5.26 μM	Sigma, United States	— — Servos et al., 1993	0	—
Anti-arrhythmic and -malarial compounds	Quinine	4 mM	Sigma, United States	—	— —	
	Quinidine	5 mM	Sigma, United States	0	—	
Polyamines	Spermine	2.5 mM	Sigma, United States	—	—	
	Spermidine	7 mM	Sigma, United States	0	0	
	Putrescine	43 μM/55 μM	Sigma, United States	0	—	0

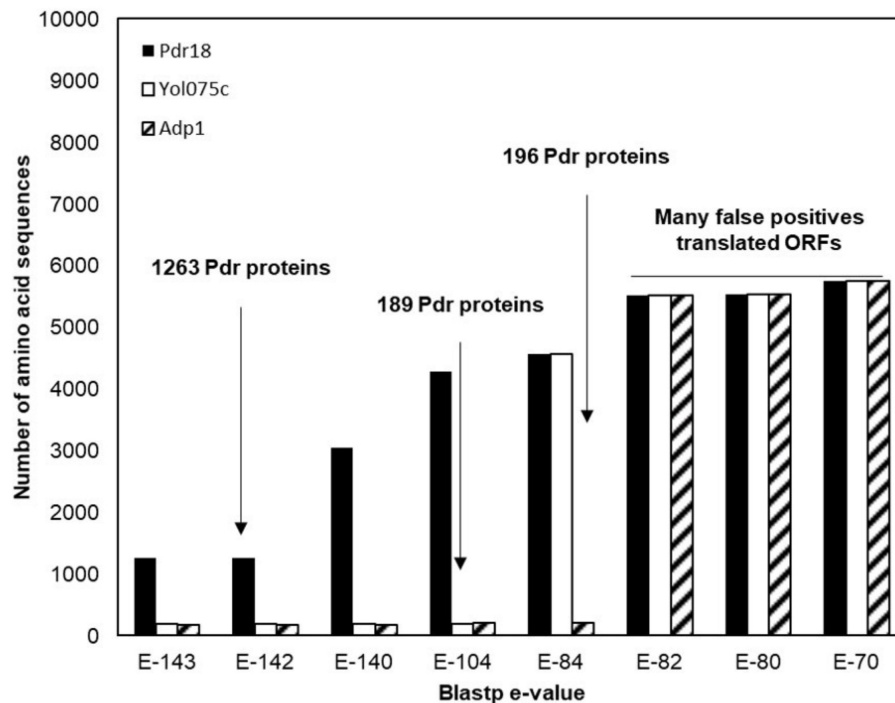
The phenotypes previously reported and confirmed in our study are identified with the corresponding bibliographic reference. 0: no growth phenotype was identified, — or — —: susceptibility phenotype, moderate or marked, respectively, +: resistance phenotype, a gray box indicates that the susceptibility was not tested. Results arise from three independent experiments.

sequence similarity of the translated ORFs comprised in the Genome DB was traversed at different blastp *e*-value thresholds using as starting nodes *Saccharomyces cerevisiae* proteins Pdr18 [representing the PDR sensu stricto proteins (Seret et al., 2009)] and Yol075c and Adp1 [representing the PDR sensu lato proteins (Seret et al., 2009)]. When Pdr18 was used as starting node the analysis of the protein sets obtained at different blastp *e*-values indicated that a threshold of E-142 was adequate, allowing the gathering of 1,263 translated ORFs (Figure 1). Using a similar approach, a threshold of E-104 and E-84 was found to be adequate for network traversal using Yol075c and Adp1 as starting nodes, allowing the gathering of 189 and 196 translated ORFs, respectively (Figure 1). The three disjoint protein sets were merged, giving a total of 1,648 translated ORFs homologous to the yeast PDR proteins in yeast strains belonging to the subphylum Saccharomycotina. This set included members of another family of transporters (hexose transporters) that were manually removed. The corresponding amino acid sequences were aligned using MUSCLE software and the protdist and neighbor algorithms made available by the PHYLIP software suite were used to construct a preliminary phylogenetic tree. Using as reference the cluster of residence of the Snq2 and Pdr18 proteins, the phylogenetic branch comprising the homologs of these two *S. cerevisiae* proteins was identified.

The comparison of the PDR protein set gathered in this study with those reported before encoded in the genomes of 10 yeast species (Seret et al., 2009) confirmed the co-clustering of the homologs of *S. cerevisiae* Snq2/Pdr18 proteins in a single branch of the phylogenetic tree. However, since it is well known that translated ORFs residing in long phylogenetic branches may correspond to protein fragments or sequence frameshifts, the blastp algorithm was used to test ORFs with dubious sequence similarity. After determining the homology of each of these proteins selected for blastp testing, a total of 214 translated ORFs, encoded in the Saccharomycetaceae yeast genomes examined, showing strong amino acid sequence similarity to the *S. cerevisiae* Snq2/Pdr18 transporters, were retained for further analysis (Table 2 and Supplementary Table S1).

## Phylogenetic Analysis of Snq2 and Pdr18 Proteins

For the phylogenetic analysis of Snq2 and Pdr18 proteins, more than one strain from a number of yeast species were used (60 strains for *S. cerevisiae*, 25 for *Saccharomyces paradoxus*, 2 for *Saccharomyces eubayanus*, 2 for *Saccharomyces bayanus*, 2 for *Naumovozyma castellii*, 2 for *Candida glabrata*, and 2 for *Zygosaccharomyces bailii*). All the repeated sequences were



**FIGURE 1 |** Identification of the PDR proteins encoded in the genomes of 117 strains, belonging to 29 different yeast species belonging to the Saccharomycetaceae family. Plot representing the number of amino acid sequences retrieved after constraining and traversing the pairwise similarity network at different e-values taking Pdr18, Yol075c, and Adp1 proteins as starting nodes. The blastp e-value used as threshold for network constraining is marked by an arrow.

removed from the protein dataset, leaving just a representative member of each species. This resulted in 146 unique translated ORFs encoded in the Saccharomycetaceae yeast genomes that were used to construct a phylogenetic tree for Pdr18 and Snq2 proteins. For this, the 146 unique translated ORFs were aligned and the corresponding phylogenetic tree constructed using the MrBayes software suite (Figure 2). After assuring that all model parameters had converged, the corresponding consensus Bayesian phylogenetic tree was retained for further analysis. Because MrBayes software suite allowed model jumping between nine different fixed-rate amino acid models, after parameter convergence the analysis of results showed that the model Wag (Whelan and Goldman, 2001) has a contribution of 100% to the posterior probability, meaning that this model became dominant over the remaining ones during the MCMC simulation. The translated ORF *lakl\_1\_h21010g* was selected as root of the phylogenetic tree because it comprised the most divergent amino acid sequence present in the protein dataset. Complementing the Bayesian approach adopted for the construction of the phylogenetic tree, the PhyML software suite was also used to obtain a ML derived tree (not shown). The analysis of the clade credibility score and bootstrap values obtained for each internal node of the Bayesian and ML trees, respectively, indicates that the two distinct statistical approaches generated similar trees. Due to their similarity, the analysis presented in the manuscript is solely based on the Bayesian tree (Figure 2).

The analysis of (i) the tree topology, (ii) the clade credibility score of the branches, and (iii) the phylogenetic distances separating the 146 unique translated ORFs led to the proposal of dividing the tree into thirteen clusters, labeled from A1 to A13 (Figures 2A,B). The Snq2 and Pdr18 homologs encoded in *Lachancea*, *Kluyveromyces*, and *Eremothecium* species occupy a basal position in this phylogenetic tree (Figure 2B – clusters A1, A11, A12, and A13). On the other hand, the early divergence of the Snq2 and Pdr18 homologs encoded in the genomes of the pre-WGD *Zygosaccharomyces* and *Torulaspora* species is not observed (Figure 2B – clusters A8 and A9). All translated ORFs showing strong amino acid sequence similarity to the *S. cerevisiae* Snq2 or Pdr18 proteins are only encoded in the genomes of yeast species classified in the *Saccharomyces* genus. The translated ORFs showing strong amino acid sequence similarity to either *S. cerevisiae* Snq2 or Pdr18 proteins are divided in two distinct branches of this phylogenetic tree, residing in the clusters A5 and A10, respectively (Figure 2B). The analysis of the phylogenetic tree also indicates that the Snq2 homologs *cagl\_1\_i04862g* and *kana\_1\_k01350* also share strong sequence similarity (cluster 6, 74.5% and 85.9% of identity and similarity, respectively). The percentage of identity and similarity shown between these Snq2/Pdr18 homologs is unexpectedly high suggesting that a lateral gene transfer event to an ancestral strain of *K. naganishii* species might have mediated the acquisition of the Snq2/Pdr18 homolog encoded in this species. The homologs of the Snq2 and Pdr18 proteins encoded in yeast species of the remaining



**TABLE 2 |** Saccharomycetaceae yeast strains examined in this work, and the number of Snq2/Pdr18 homologs, Snq2 orthologs, and Pdr18 orthologs identified.

Species	Nr of strains	Nr of Snq2 and Pdr18 homologs	Nr of Snq2 orthologs	Nr of Pdr18 orthologs
<i>Saccharomyces cerevisiae</i>	60	115	59	56
<i>Saccharomyces paradoxus</i>	25	50	25	25
<i>Saccharomyces mikatae</i>	1	3	1	1
<i>Saccharomyces kudriavzevii</i>	1	2	1	1
<i>Saccharomyces arboricola</i>	1	2	1	1
<i>Saccharomyces eubayanus</i>	2	4	2	2
<i>Saccharomyces bayanus</i> (HYBRID)	2	1	1	0
<i>Saccharomyces uvarum</i>	1	1	0	1
<i>Kazachstania africana</i>	1	2	2	0
<i>Kazachstania naganishii</i>	1	1	1	0
<i>Naumovozyma castellii</i>	2	4	4	0
<i>Naumovozyma dairenensis</i>	1	2	2	0
<i>Candida glabrata</i>	2	2	2	0
<i>Tetrapispora blattae</i>	1	2	2	0
<i>Tetrapispora phaffii</i>	1	2	2	0
<i>Vanderwaltozyma polyspora</i>	1	2	2	0
<i>Zygosaccharomyces bailii</i>	2	6	6	0
<i>Zygosaccharomyces rouxii</i>	1	2	2	0
<i>Torulaspora delbrueckii</i>	1	1	1	0
<i>Lachancea kluyveri</i>	1	2	2	0
<i>Lachancea thermotolerans</i>	1	1	1	0
<i>Lachancea waltii</i>	1	1	1	0
<i>Kluyveromyces aestuarii</i>	1	1	1	0
<i>Kluyveromyces lactis</i>	1	1	1	0
<i>Kluyveromyces marxianus</i> var. <i>marxianus</i>	1	1	1	0
<i>Kluyveromyces wickerhamii</i>	1	1	1	0
<i>Ashbya aceri</i>	1	1	1	0
<i>Eremothecium cymbalariae</i>	1	1	1	0
<i>Eremothecium gossypii</i>	1	1	1	0

Pre-whole genome duplication (WGD) species are highlighted in gray. A detailed list of the strains analyzed, number of Pdr18 and Snq2 proteins and sources of genome information and annotation tools is provided as **Supplementary Table S1**.

post-WGD taxonomic genera (*Kazachstania*, *Naumovozyma*, *Tetrapispora*, and *Vanderwaltozyma*) reside in the phylogenetic clusters A2, A3, A4, and A7 (**Figure 2B**).

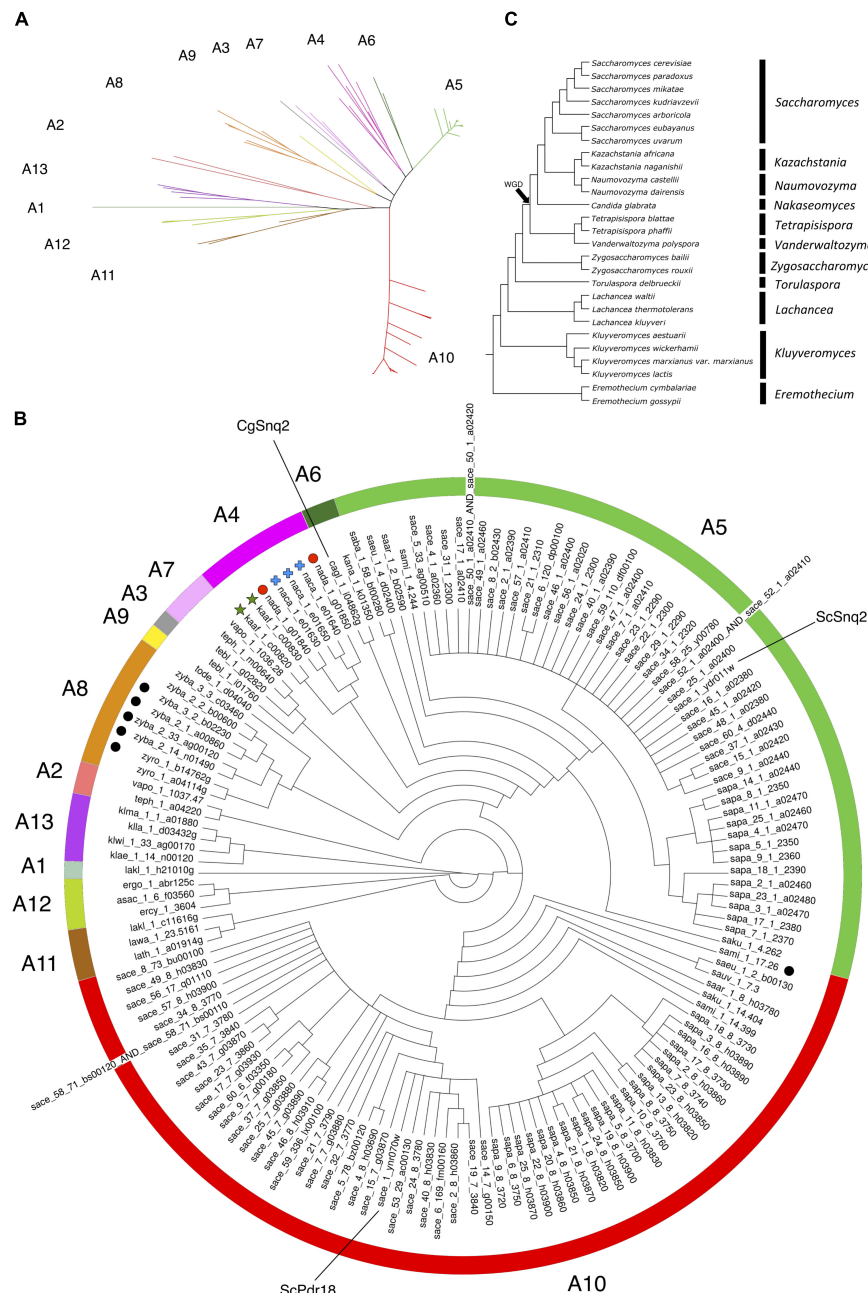
## Gene Neighborhood Analysis of Pdr18 and Snq2 Orthologs in Saccharomycetaceae Yeasts

Gene neighborhood analysis of *S. cerevisiae* SNQ2 and PDR18 homolog genes encoded in the examined Saccharomycetaceae yeast strains genomes was performed. This study, together with the phylogenetic analysis, is useful to contribute to the elucidation of their ortholog/paralog status.

Results show that for the pre-WGD yeasts species of *Torulaspora*, *Lachancea*, *Kluyveromyces*, and *Eremothecium* genera there is a single lineage, with genes sharing strong synteny (**Figures 3, 4**). The main exception to this rule is the *Lachancea kluyveri* CBS 3082 strain because, in addition to the gene *lakl\_1\_c11616g* residing in the above mentioned conserved chromosome environment, this yeast strain also encodes one singleton gene, *lakl\_1\_h21010g*, sharing very weak synteny with *Kluyveromyces lactis* and *Kluyveromyces marxianus* var. *marxianus* genes (**Figure 3**). The pre-WGD yeast strains of the *Zygosaccharomyces* genus, *Z. rouxii* CBS 732 and *Z. bailii* CLIB213, both encode two SNQ2/PDR18 homolog genes (**Figure 3**). However, *Z. bailii* IST302 encodes two additional SNQ2/PDR18 homolog genes, *zyba\_2\_14\_n01490* and *zyba\_2\_33\_ag00120*, lacking common neighbors with the remaining SNQ2/PDR18 homolog genes in the Saccharomycetaceae strains genomes examined (**Figure 3** and **Supplementary Figure S1**). For this reason, these two genes were considered singletons. The high amino acid sequence identity shared by these two genes suggests that they are a paralog pair originated in a duplication event that occurred recently in the evolution of *Zygosaccharomyces* species.

After the WGD event, the above described single gene lineage gave rise to two gene sub-lineages and the chromosome regions where these genes reside in pre- and early-divergent post-WGD yeast species are conserved (**Figure 3**). For instance, the analysis of the neighborhood of the pre-WGD *Zygosaccharomyces rouxii* gene *zyro\_1\_a04114g* and of the post-WGD species *Vanderwaltozyma polyspora* gene *vapo\_1\_1037.47* shows the existence of 9 common neighboring genes, some of them absent from **Figure 3** but that can be seen in **Supplementary Figure S1**. *Z. rouxii* gene *zyro\_1\_a04114g* also shares 9 common neighboring genes with the second gene encoded in *V. polyspora* (*vapo\_1\_1036.28*), belonging to the second sub-lineage of ohnolog genes originating from WGD (**Figure 3** and **Supplementary Figure S1**).

The *Candida glabrata* strains' genomes analyzed only encode one Pdr18/Snq2 homolog gene sharing strong synteny with all the homologs encoded in post-WGD species, from *Naumovozyma dairenensis* to *Kazachstania africana* (**Figure 3** and **Supplementary Figure S1**). This sub-lineage also shows strong synteny with one of the gene sub-lineages encoded in *Saccharomyces* genus species, corresponding to the ScSNQ2 orthologs in *Saccharomyces* yeasts, supporting the ScSNQ2-ortholog status for the SNQ2/PDR18 homologs encoded in the post-WGD species, from *C. glabrata* to *S. cerevisiae* (**Figures 3, 4**). We also hypothesize the loss of the ohnolog

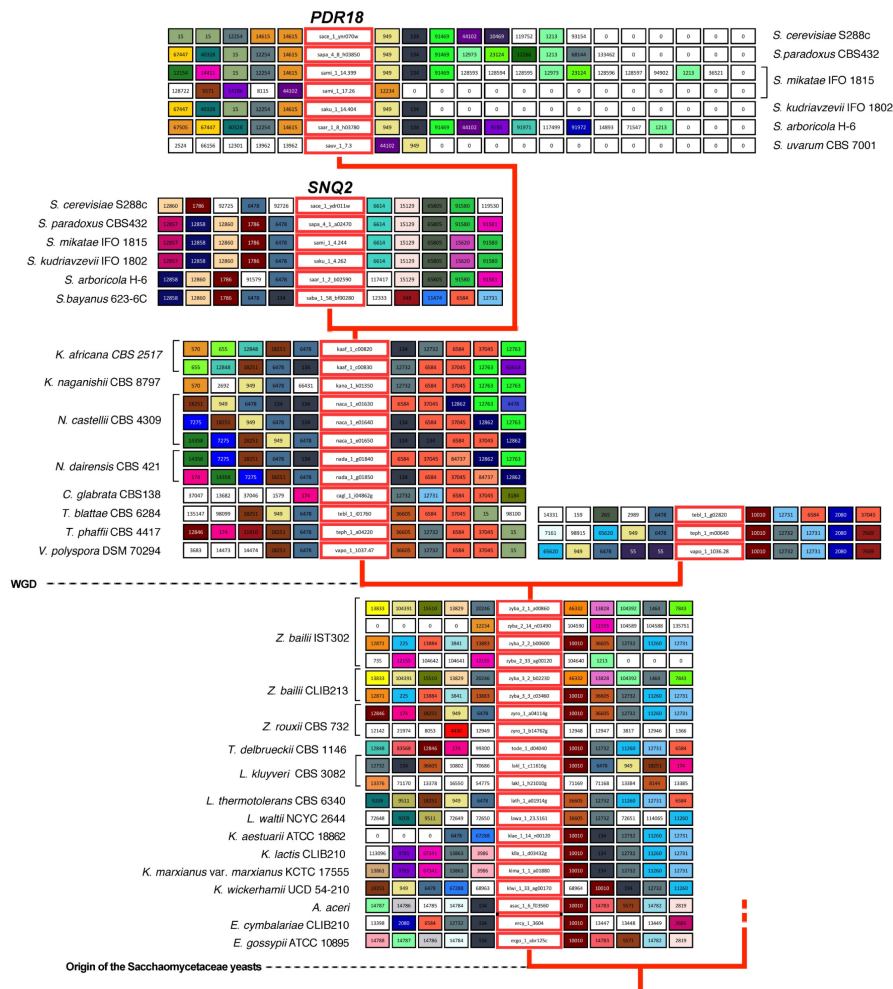


**FIGURE 2 |** Phylogenetic tree of the *Pdr18* and *Snq2* homologs encoded in the genome sequences of 117 strains of 29 yeast species belonging to the *Saccharomycetaceae* family. **(A)** Radial phylogram showing the amino acid sequence similarity distances between the protein homologs of *Pdr18* and *Snq2*. **(B)** Circular cladogram showing the tree topology. Singleton genes were identified with graphical labeling of the translated ORFs using black circles, and tandemly duplicated genes were identified with graphical labeling of the translated ORFs using green stars, red circles and blue crosses. The translated ORF *lak1\_1\_h21010g* was chosen as outgroup due to the strong dissimilarity of its amino acid sequence. **(C)** Schematic phylogenetic tree of the species (adapted from Kurtzman and Robnett, 2003).

gene from the second sub-lineage in the common ancestor of *Saccharomyces*, *Nakaseomyces*, *Kazachstania*, and *Naumovozyma* genera, given that all the encoded *PDR18/SNQ2* homologs from *C. glabrata* to *K. africana* appear to be *ScSNQ2* orthologs.

The gene neighborhood analysis of the *S. cerevisiae* *SNQ2/PDR18* homologs encoded in the genomes of the

*Saccharomyces* genus yeast species *S. paradoxus*, *S. mikatae*, *S. kudriavzevii*, and *S. arboricola*, shows the existence of two gene sub-lineages, one comprising the above described *SNQ2* orthologs and the other comprising *PDR18* orthologs. *PDR18* orthologs occur exclusively in the *Saccharomyces* genus yeast species and the sub-lineage constituted by these orthologs shows

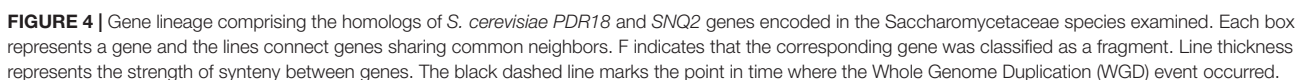


**FIGURE 3 |** Gene neighborhood of *SNQ2* and *PDR18* genes and corresponding homologs encoded in the genomes examined. Genes in the central boxes represent *SNQ2*, *PDR18* and their orthologs. Adjacent boxes represent gene neighbors and homologous neighbors are highlighted in the same color. A white box represents genes with no homologous neighbors in the represented chromosome region and white boxes with a zero represent the end of the contig/chromosome. The synteny was assessed with 15 neighbors on each side but this representation was truncated to five neighbors. In the case of *PDR18* orthologs, the gene neighborhood data was not truncated in one of the sides to highlight the sub-telomeric position occupied by these genes.

a very strong synteny (Figure 3). However, the analysis of the gene neighborhood of the *SNQ2* and *PDR18* orthologs in *Saccharomyces* species residing in a basal phylogenetic position could not be performed due to lack of data: the translated ORF encoding the *S. uvarum* *SNQ2* ortholog, and the *PDR18* ortholog encoded in the genome of *S. bayanus* 623-6C were fragmented into two different contigs of small dimension. Therefore the closely related species *S. eubayanus* (Baker et al., 2015), whose genome had not been at first included in Genome DB, was used instead. The chromosome environment where the *SNQ2* and *PDR18* orthologs reside in the *S. eubayanus* genome was inspected by performing manual blastp pairwise comparisons of the amino acid sequence of the neighboring genes against the full protein set of *S. cerevisiae*. This analysis also showed that *S. eubayanus* CBS12357 and FM1318 strains encode a single *SNQ2* ortholog and a single *PDR18* ortholog sharing strong

synteny with the remaining genes within each sub-lineage. These results support the notion that the chromosome environment where *SNQ2* and *PDR18* paralogs reside in the genome of *Saccharomyces* species has been conserved since their appearance in the last common ancestor of the species comprised in this taxonomic genus.

The *S. cerevisiae* *PDR18* gene resides in the subtelomeric region of chromosome XIV (Figure 3), a region poorly conserved throughout Saccharomycetaceae yeasts' evolution, sharing little synteny with the homologs of the *SNQ2/PDR18* genes encoded in the genomes of yeast species belonging to the other post-WGD taxonomic genera. In fact, the only neighbor in the vicinity of the *PDR18* gene that did not belong to large gene families and is also shared with the *K. naganishii*, *N. castellii* and *N. dairenensis* species is ORF YNR071C, a non-biochemically characterized member of a gene family of aldose 1-epimerases (Li et al., 2013). This



acid sequence similarity comprises a total of 366 members in the 170 hemiascomycetous strains of the Genome DB (gathered at an *e*-value threshold of E-50). The gene neighborhood analysis



showed that the members of this cluster are present in the chromosome environment where *Snq2* orthologs from pre- and post-WGD species and *Pdr18* orthologs from *Saccharomyces* genus species reside (Figure 3).

Although the gene neighborhood results suggest that the gene duplication event giving rise to the *S. cerevisiae* *SNQ2* and *PDR18* genes occurred in a recent ancestor of the genus *Saccharomyces*, the phylogenetic relationships between the A5 and A10 clades (Figure 2) is apparently not consistent with the proposed evolutionary scenario. To clarify this seemingly inconsistency, the amino acid sequences of the orthologs of the *SNQ2/PDR18* genes encoded in the post-WGD species (left-sublineage in Figure 4) were used to construct a phylogenetic tree using the same initial parameters used in construction of the tree shown in Figure 2 (Supplementary Figure S2). The decision of constructing a new tree was based on the plausible notion that the inclusion of non-orthologous sequences might be introducing “noise” in the multiple alignment of the sequences and, consequently, distorting tree topology and introducing error in the true phylogenetic relationships established between the genes comprised in the WGD left sub-lineage. After parameter convergence, the analysis of this phylogenetic results indicated that the model Jones (Jones et al., 1992) has a contribution of 100% to the posterior probability. In fact, the analysis of this new tree clearly shows that the node joining clades A5 and A10 defines a branch comprising only genes encoded in the *Saccharomyces* species and that the last common ancestor of the homologs of the *SNQ2* and *PDR18* genes coincides with the origin of the *Saccharomyces* genus. The obtained clade credibility score for the orthologs of these two *S. cerevisiae* genes residing together in a single phylogenetic branch of this Bayesian consensus tree was calculated as being 0.6724. These results are consistent with the evolutionary scenario proposed based on the gene neighborhood analysis and confirm the previously proposed existence of a paralogous relationship between the *S. cerevisiae* *SNQ2* and *PDR18* genes but also establishing that the gene duplication event in its origin occurred more recently than formerly proposed (Seret et al., 2009), presumably coinciding with the first common ancestor of the yeast species comprised in the *Saccharomyces* genus, before the radiation of these species.

### Susceptibility Profiling of *S. cerevisiae* *snq2Δ* and *pdr18Δ* and *C. glabrata* *snq2Δ* Deletion Mutants

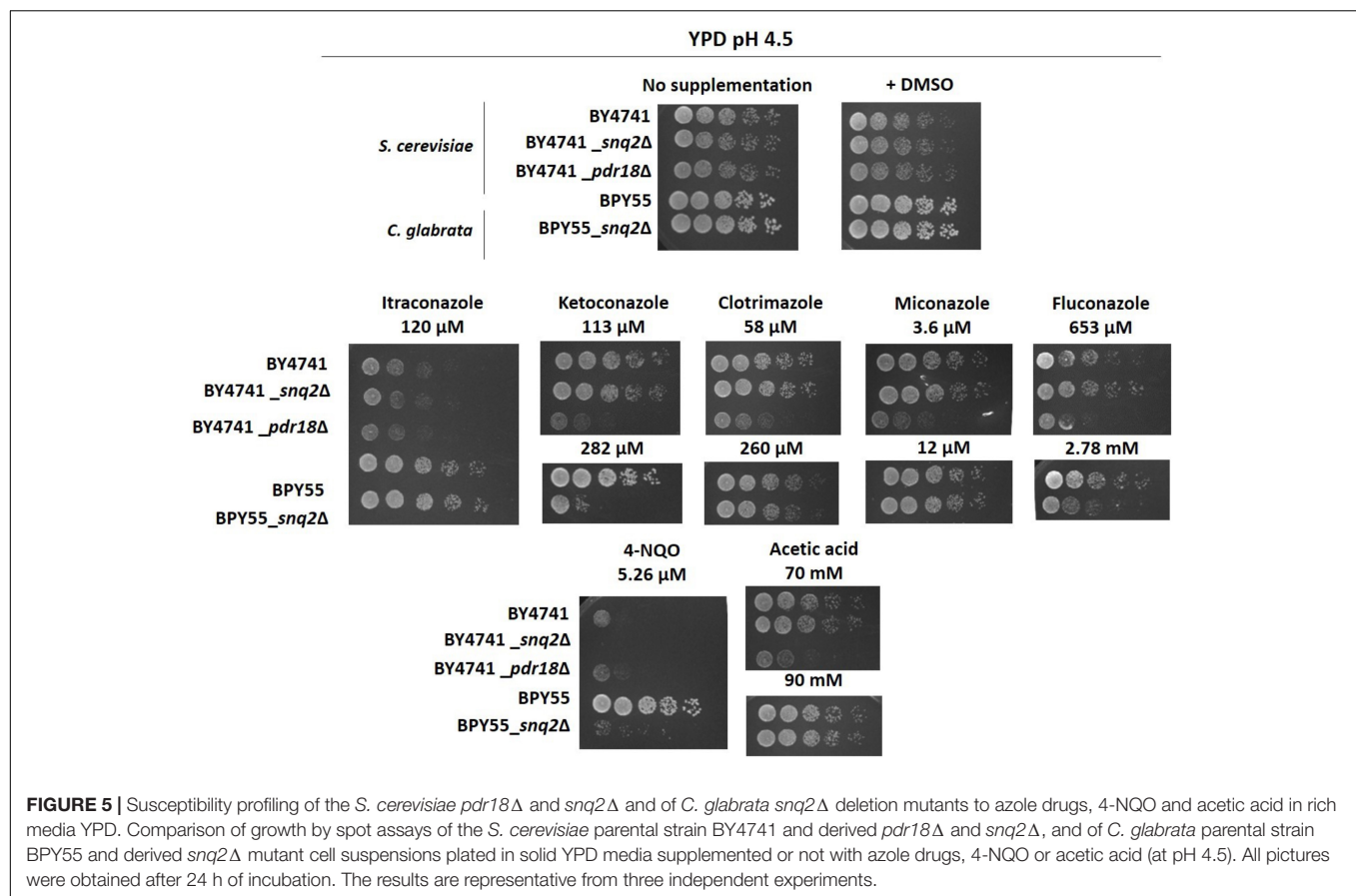
The post-WGD *C. glabrata* species genome encodes a sole *ScSnq2* ortholog, the *CgSnq2* (Sanglard et al., 2001). The encoded gene has diverged before the hypothesized duplication event that originated *S. cerevisiae* *PDR18*. For this reason, *C. glabrata* was the selected species to get insights into the functional divergence of the ancestral gene and *S. cerevisiae* *PDR18* and *SNQ2* genes. Since *ScPdr18*, *ScSnq2*, and *CgSnq2* were reportedly involved in MDR/MXR, the functional divergence of these proteins was examined by profiling the growth susceptibility of the corresponding deletion mutants against a wide range of cytotoxic compounds, some of them already described as potential substrates for these drug/xenobiotic pumps. The susceptibility

of *S. cerevisiae* BY4741-derived *snq2Δ* and *pdr18Δ* deletion mutant strains to a wide range of chemical compounds was screened under identical conditions in minimal medium MM4 (*S. cerevisiae*), ranging from weak acids, alcohols, polyamines, metal cations to herbicides, fungicides and anti-arrhythmic and -malarial compounds (Table 1). The deletion of the *SNQ2* gene in *S. cerevisiae* was found to lead to increased susceptibility to 4-NQO,  $\text{Li}^+$  and  $\text{Mn}^{2+}$ , in agreement with previous studies (Servos et al., 1993; Miyahara et al., 1996). Moreover, this systematic screening contributed to extend to the herbicides barban, alachlor, and metolachlor, the anti-malarial anti-arrhythmic quinine and the polyamine spermine the list of toxic compounds to which *Snq2* confers protection in *S. cerevisiae* (Table 1).

The deletion of the *PDR18* gene was found to render *S. cerevisiae* cells more susceptible toward almost all the compounds tested, contrasting with *SNQ2* whose MDR/MXR spectrum is apparently more limited, the phenotypes only coinciding for 6 of the 35 compounds tested (Table 1). In particular, the higher toxic effect of weak acids and of fungicides that target either ergosterol biosynthesis (azoles) or the ergosterol molecule itself (amphotericin B) toward the *pdr18Δ* strain is evident (Table 1), consistent with the role described for this ABC transporter in acetic acid resistance and in ergosterol transport at plasma membrane (Cabrito et al., 2011; Teixeira et al., 2012; Godinho et al., 2018).

The expression of *CgSNQ2* gene in *C. glabrata* was found not to confer either protection or increased susceptibility to stress induced by acetic and benzoic acids, the herbicides 2,4-D and MCPA, the alcohols ethanol and 1,4-butanediol, and the polyamine putrescine in minimal medium, similarly to former observations for the expression of *ScSNQ2* in *S. cerevisiae*. This profile is considerably different from that exhibited by *S. cerevisiae* cells devoid of *PDR18*, which were more susceptible to all these compounds, except for 1,4-butanediol, toward which this deletion mutant was more tolerant (Table 1). Moreover, susceptibility phenotypes for *Cgsnq2Δ* deletion mutant were detected with toxic concentrations of the mutagen 4-NQO [which is a described putative substrate for *S. cerevisiae* efflux pump *Snq2* (Servos et al., 1993)]. However, the susceptibility phenotype for the lithium cation exhibited by *Scsnq2Δ* mutant, was not detected for the *Cgsnq2Δ* mutant (Table 1). Moreover, *CgSnq2* is apparently a determinant of resistance to  $\text{Cu}^{2+}$ , while no phenotype was found for *Scsnq2Δ* (Table 1). The toxic effect of the azole drugs clotrimazole, ketoconazole, and fluconazole, and of amphotericin B was not alleviated by the expression of *CgSNQ2* when tested in minimal medium, consistent with what was observed for *S. cerevisiae* *ScSNQ2* expression. Surprisingly, in the presence of itraconazole and miconazole, the deletion of *CgSNQ2* was apparently advantageous (Table 1).

Although the results in Table 1 suggest that *CgSnq2* has no positive effect in *C. glabrata* BPY55 resistance to azoles, a susceptibility phenotype for the BPY55\_ *snq2Δ* mutant in the presence of the azole drugs fluconazole and ketoconazole was previously reported (Torelli et al., 2008). However, the spot assays performed in the referred study were performed in YPD



medium while all the phenotypes described in **Table 1** for *C. glabrata* were performed in minimal medium MM. Therefore, we investigated the susceptibility to azole drugs in rich medium YPD for *C. glabrata* strains, as well as for *S. cerevisiae* strains to confirm results obtained in minimal media MM4 (**Figure 5**). We also included the well-characterized phenotypes for *Scpdr18Δ* and *Scsnq2Δ* in the presence of acetic acid and 4-NQO, respectively, and confirmed that there is no interference of the growth medium used in the phenotypes obtained (**Figure 5**). The higher susceptibility of *Scpdr18Δ* deletion mutant, when compared to the corresponding parental strain, to the azole drugs ketoconazole, clotrimazole, miconazole, fluconazole, and to a lower extent, to itraconazole was confirmed in YPD (**Figure 5**). Also, as found in minimal medium MM4 (**Table 1**), *Scsnq2Δ* shows no susceptibility phenotype in the presence of the azole drugs in YPD (**Figure 5**). In summary, even in YPD media, *CgSnq2* is not a determinant of *C. glabrata* tolerance to itraconazole, clotrimazole or miconazole (**Figure 5**), but the phenotypes previously reported for fluconazole and ketoconazole (Torelli et al., 2008) were here confirmed.

## DISCUSSION

The evolutionary history of *Saccharomyces cerevisiae* *PDR18* gene in Saccharomycetaceae yeasts was reconstructed in this

study. Compared with a former analysis based on the genome sequences of only nine yeast species belonging to the *Hemiascomycetes* phylum (Seret et al., 2009), our study took advantage of the increasing number of yeast genomes currently available, and has examined sixteen post-WGD yeast species instead of only two, spanning six different taxonomic genera, instead of only two. This fact allowed the clarification of the evolution of *S. cerevisiae* *PDR18* and *SNQ2* genes homologs after the WGD event and led us to propose that a single gene loss event has occurred in the last common ancestor of *Nakaseomyces*, *Naumovozyma*, *Kazachstania*, and *Saccharomyces* yeasts. This event explains the “interruption” of one of the post-WGD sub-lineages given that yeast strains of the *Nakaseomyces*, *Naumovozyma*, and *Kazachstania* genera encode only one *PDR18/SNQ2* homolog gene with high synteny between them. The probability that the duplication event has occurred during the WGD event is negligible, as it would imply the gene loss in the five yeast species belonging to the *Nakaseomyces*, *Naumovozyma*, and *Kazachstania* genera. Also, the gene neighborhood analysis does not provide support to the second evolutionary scenario, as only one common neighboring gene in these sub-lineages belongs to the similarity cluster 949, which is also present in the chromosome environment of many genes comprised in the other post-WGD sub-lineage proposed to give rise to the duplication event. Altogether, results do not support the ohnolog status for *S. cerevisiae*

*SNQ2* and *PDR18* genes, but instead support the first proposed scenario.

Concerning the point in time where the duplication event originating the *SNQ2* and *PDR18* sub-lineages might have occurred in the evolution of the post-WGD yeast species, it is possible that the *PDR18* gene ancestor was nurtured in one of the tandem repeats encoding *PDR* genes that, with exception of the *K. naganishii* genome, are found in the genomes of all yeast species belonging to the *Kazachstania* and *Naumovozyma* genera. Under this scenario, the amino acid sequence of the ancient gene that gave rise to *PDR18* did not diverge until the last common ancestor in the origin of the *Saccharomyces* genus split apart from the ancestral yeast population giving rise to the *Kazachstania* genus. In the second scenario, the *PDR18* gene ancestral was originated in an independent duplication event not related with the events on the origin of the tandem repeats observed in the genomes of the *Kazachstania* and *Naumovozyma* species. Independently of which of these evolutionary scenarios is true, subsequent genome shuffling and/or other mechanisms of genome evolution should have been responsible for the transposition of the two ancestral genes encoding *Pdr18* and the neighboring gene, encoding the aldose 1-epimerase, into a new chromosome environment.

Beside the gene duplication events in the origin of the *SNQ2* and *PDR18* paralog genes, other gene duplication events were identified in the genomes of the protoploid *Saccharomycetaceae* (pre-WGD). The yeast species belonging the *Zygosaccharomyces* genus and the *Lachancea kluyveri* species encoded more than one *Snq2/Pdr18* homolog in their genomes, escaping the typical pattern observed in the majority of the pre-WGD species analyzed in this study, where a sole member of the *Snq2/Pdr18* protein subfamily were found encoded in the corresponding genome sequences. Interestingly, the amino acid sequences of the two singletons encoded in the genome of the *Z. bailii* IST302 strain showed the existence of a strong divergence in respect to the amino acid sequences of the *Snq2/Pdr18* homologs encoded in the other *Zygosaccharomyces* strains analyzed in this study. This fact and the high amino acid sequence identity shared by these two singletons suggests that these are a paralog pair originated in a duplication event that occurred recently in the evolution of the *Zygosaccharomyces* yeast species. The consultation of the MIPS website comprising the aligning of the presently available genome sequences of yeast strains belonging to the *Zygosaccharomyces* genus showed that these two singleton genes are also absent in the interspecies hybrid *Z. bailii* ISA1307<sup>3</sup>. This result further strengthens the hypothesis of a recent origin of these two singletons occurring at the intraspecific level.

Since it was found that *PDR18* is specific for the *Saccharomyces* genus, with very high conservation among the 93 *Saccharomyces* yeast genomes examined, the MDR/MXR profiling of this transporter in *S. cerevisiae* was examined for a wide number of relevant toxic compounds and the possible overlapping of the susceptibility phenotypes exhibited by the *Scpdr18Δ* mutant and the *Scsnq2Δ* mutant with its paralog gene *ScSNQ2* deleted was

systematically examined. *ScSnq2* was found to confer resistance to a more restricted range of the toxic compounds tested, compared with *ScPdr18* with a demonstrated physiological function as plasma membrane transporter in the maintenance of plasma membrane ergosterol content specially under chemical stress. This biological role was related with the decreased levels of stress-induced membrane disorganization and permeabilization and counteracting transmembrane electrochemical potential dissipation (Cabrito et al., 2011; Godinho et al., 2018) and thus with the maintenance under stress of a functional plasma membrane as a selective barrier and a suitable lipid environment for the physiological activity of the embedded proteins (del Castillo Agudo, 1992; Parks and Casey, 1995; Eisenkolb, 2002; Aguilera et al., 2006; Abe and Hiraki, 2009; Caspeta et al., 2014; Kodedová and Sychrová, 2015).

Since the anticipated functional divergence between *ScPdr18*, *ScSnq2* and the ancestral gene on the origin of the duplication event is of relevance to understand the evolutionary process acting on the two duplicate genes, the post-WGD pathogenic species *Candida glabrata* was selected for a systematic analysis of the susceptibility phenotype of the corresponding deletion mutant *Cgsnq2Δ*. Based on the susceptibility assays performed, the sole *SNQ2/PDR18* homolog in *C. glabrata* encoding *CgSnq2* appeared to be functionally closer to *ScSnq2* than to *ScPdr18*, playing a role in 4-NQO resistance in both species and having no impact in *C. glabrata* tolerance to the weak acids acetic and benzoic, the herbicides 2,4-D and MCPA, the alcohols ethanol and 1,4-butanediol, and the polyamine putrescine, to which *ScPDR18* expression confers tolerance. It is noteworthy that *CgSnq2* was previously found to play a role in *C. glabrata* BPY55 tolerance to fluconazole and ketoconazole by susceptibility assays in solid YPD medium (Torelli et al., 2008). This phenotype was confirmed in this study under similar conditions, a fact that could indicate some overlapping of the function associated to *CgSnq2* and *ScPdr18*. However, these phenotypes were not reproduced in minimal media MM, the growth conditions used for the systematic analysis of the phenotypic profiling performed.

The apparent overlapping between *ScPdr18* and *CgSnq2* role observed in azole resistance in rich media can be related with species-specific adaptation of *C. glabrata* to these fungicides, given that this yeast species is one of the most common in nosocomial fungal infections and that azoles are one of the main families of drugs that are currently being used to treat or prevent fungal infections (Perlroth et al., 2007; Jandric and Schüller, 2011; Roetzer et al., 2011) and that the *C. glabrata* strain used in this work is a highly azole resistant clinical isolate (Torelli et al., 2008). The sensitivity phenotype exhibited by *Cgsnq2Δ* toward  $\text{Cu}^{2+}$  toxicity might also be related to the fact that BPY55 is a clinical isolate and that adaptation to high  $\text{Cu}^{2+}$  environmental concentrations is a described determinant for survival of human pathogens (Festa and Thiele, 2012; Samanovic et al., 2012; Chaturvedi and Henderson, 2014; Fu et al., 2014; García-Santamarina and Thiele, 2015).

The fact that *S. cerevisiae* *Snq2* and *Pdr18* confer resistance to a very different set of chemical compounds with little overlapping,

<sup>3</sup><https://www.helmholtz-muenchen.de/ibis/index.html>



appears to exclude the evolutionary scenario where these highly similar MDR/MXR transporters have been retained in the *S. cerevisiae* genome due to functional redundancy or dosage effect, suggesting as the most consistent scenarios the subfunctionalization and the neofunctionalization of the gene copies. Although subfunctionalization is commonly associated with the mere division of functions of the ancestral protein by the two duplicates, another possible model is that one of the duplicate proteins becomes more efficient at performing one of the original functions of the progenitor gene (Zhang, 2003). Considering the neofunctionalization process, in most of the cases the function adopted by one of the duplicate proteins is a related function rather than an entirely new function (Zhang, 2003; Conant and Wolfe, 2008). Although there are no detailed studies focusing on *SNQ2* expression impact on yeast plasma membrane lipid content and properties, *Snq2* was previously found to contribute to alleviate estradiol toxicity in *S. cerevisiae*, a molecule highly similar to ergosterol (Mahé et al., 1996a) and *Pdr18* is described to be an ergosterol transporter at the plasma membrane (Godinho et al., 2018). The clarification of whether *ScSnq2* and/or *CgSnq2* may play some role on lipid homeostasis and of whether *ScPdr18* is the result from subfunctionalization by function specialization or whether the biological function acquired by *ScPdr18* is totally different from *ScSnq2* and *CgSnq2*, require further work and the elucidation of *SNQ2* biological function.

*Pdr18* is encoded in at least 87 out of the 93 genomes from the *Saccharomyces* genus yeasts examined in this study: in 56 out of 60 *S. cerevisiae* genomes, in the 25 *S. paradoxus* genomes, in the genomes of *S. mikatae*, *S. kudriavzevii*, *S. arboricola*, and *S. uvarum*, and in the two *S. eubayanus* genomes. DNA degradation during preparation for genome sequencing is the most plausible explanation for the lack of the *PDR18* gene in some of the strains. For example, *PDR18* gene was not found in the genome sequence of *S. cerevisiae* cen.pk113-7d strain since the genome sequencing shows no coverage for the right arm of chromosome XIV, where *PDR18* gene resides (Nijkamp et al., 2012). Moreover, *PDR18* gene was found to be present in the genomes of 964 isolates of more than 1,000 natural *S. cerevisiae* isolates that were a recently examined using deep coverage genome sequencing (Peter et al., 2018). This fact is consistent with the relevant physiological function encoded by *PDR18* in this yeast species (Godinho et al., 2018).

## REFERENCES

- Abe, F., and Hiraki, T. (2009). Mechanistic role of ergosterol in membrane rigidity and cycloheximide resistance in *Saccharomyces cerevisiae*. *Biochim. Biophys. Acta* 1788, 743–752. doi: 10.1016/j.bbmem.2008.12.002
- Aguilera, F., Peinado, R. A., Millan, C., Ortega, J. M., and Mauricio, J. C. (2006). Relationship between ethanol tolerance,  $H^+$ -ATPase activity and the lipid composition of the plasma membrane in different wine yeast strains. *Int. J. Food Microbiol.* 110, 34–42. doi: 10.1016/j.ijfoodmicro.2006.02.002
- Altekar, G., Dwarkadas, S., Huelsenbeck, J. P., and Ronquist, F. (2004). Parallel metropolis coupled markov chain monte carlo for Bayesian phylogenetic inference. *Bioinformatics* 20, 407–415. doi: 10.1093/bioinformatics/btg427
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410. doi: 10.1016/S0022-2836(05)80360-2

## DATA AVAILABILITY

The dataset generated for this study can be found in TreeBase (<http://purl.org/phylo/treebase/phyloids/study/TB2:S23282>).

## AUTHOR CONTRIBUTIONS

PD and CG carried out the phylogenetic and gene neighborhood analysis. CG and EP carried out the susceptibility tests. IS-C guided and coordinated this study and together with CG and PD prepared the manuscript. All authors read and approved the final manuscript.

## FUNDING

This work was supported by ‘Fundação para a Ciência e a Tecnologia’ (FCT) (YEASTPEC project contract ERA-IB-2/0003/2015) and Ph.D. and postdoctoral fellowships to CG (SFRH/BD/ 92252/2013) and PD (SFRH/BPD/74618/2010). Funding received by iBB from FCT (UID/BIO/04565/2013) and from Programa Operacional Regional de Lisboa 2020 (Project No. 007317) is also acknowledged.

## ACKNOWLEDGMENTS

We thank D. Sanglard, CHUV, Lausanne, Switzerland for kindly providing the *C. glabrata* strains. This work was dedicated to the memory of Professor André Goffeau who coordinated the international collaborative effort that lead to the first completely sequenced yeast genome, opening the door to the field of functional genomics and genome evolution of yeasts, in particular of multidrug resistance transporter genes.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2018.00476/full#supplementary-material>

- Baker, E., Wang, B., Bellora, N., Peris, D., Hulfachor, A. B., Koshalek, J. A., et al. (2015). The genome sequence of *Saccharomyces eubayanus* and the domestication of lager-brewing yeasts. *Mol. Biol. Evol.* 32, 2818–2831. doi: 10.1093/molbev/msv168
- Balzi, E., and Goffeau, A. (1995). Yeast multidrug resistance: the PDR network. *J. Bioenerg. Biomembr.* 27, 71–76. doi: 10.1007/BF02110333
- Borst, P., Zelcer, N., and van Helvoort, A. (2000). ABC transporters in lipid transport. *Biochim. Biophys. Acta* 1486, 128–144. doi: 10.1016/S1388-1981(00)00053-6
- Cabrito, T. R., Teixeira, M. C., Singh, A., Prasad, R., and Sá-Correia, I. (2011). The yeast ABC transporter *Pdr18* (ORF YNR070w) controls plasma membrane sterol composition, playing a role in multidrug resistance. *Biochem. J.* 440, 195–202. doi: 10.1042/BJ20110876



- Caspeta, L., Chen, Y., Ghiaci, P., Feizi, A., Buskov, S., Hallström, B. M., et al. (2014). Altered sterol composition renders yeast thermotolerant. *Science* 346, 75–78. doi: 10.1126/science.1258137
- Charif, D., Lobry, J., Necsulea, A., Palmeira, L., and Penel, M. S. (2007). *The Seqinr Package. R Package*. Available at: <https://cran.r-project.org/web/packages/seqinr/seqinr.pdf>
- Chaturvedi, K. S., and Henderson, J. P. (2014). Pathogenic adaptations to host-derived antibacterial copper. *Front. Cell. Infect. Microbiol.* 4:3. doi: 10.3389/fcimb.2014.00003
- Clamp, M., Cuff, J., Searle, S. M., and Barton, G. J. (2004). The Jalview Java alignment editor. *Bioinformatics* 20, 426–427. doi: 10.1093/bioinformatics/btg430
- Conant, G. C., and Wolfe, K. H. (2008). Turning a hobby into a job: how duplicated genes find new functions. *Nat. Rev. Genet.* 9, 938–950. doi: 10.1038/nrg2482
- Decottignies, A., and Goffeau, A. (1997). Complete inventory of the yeast ABC proteins. *Nat. Genet.* 15, 137–145. doi: 10.1038/ng0297-137
- del Castillo Agudo, L. (1992). Lipid content of *Saccharomyces cerevisiae* strains with different degrees of ethanol tolerance. *Appl. Microbiol. Biotechnol.* 37, 647–651. doi: 10.1007/BF00240742
- Dias, P. J., and Sá-Correia, I. (2013). The drug: H + antiporters of family 2 (DHA2), siderophore transporters (ARN) and glutathione: H + antiporters (GEX) have a common evolutionary origin in hemiascomycete yeasts. *BMC Genomics* 14:901. doi: 10.1186/1471-2164-14-901
- Dias, P. J., and Sá-Correia, I. (2014). Phylogenetic and syntenic analyses of the 12-spanner drug: H + antiporter family 1 (DHA1) in pathogenic *Candida* species: evolution of MDR1 and FLU1 genes. *Genomics* 104, 45–57. doi: 10.1016/j.ygeno.2014.05.005
- Dias, P. J., Seret, M.-L., Goffeau, A., Correia, I. S., and Baret, P. V. (2010). Evolution of the 12-spanner drug: H + antiporter DHA1 family in Hemiascomycetous yeasts. *OMICS* 14, 701–710. doi: 10.1089/omi.2010.0104
- Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797. doi: 10.1093/nar/gkh340
- Eisenkolb, M., Zenzmaier, C., Leitner, E., and Schneider, R. (2002). A specific structural requirement for ergosterol in long-chain fatty acid synthesis mutants important for maintaining raft domains in yeast. *Mol. Biol. Cell* 13, 4414–4428. doi: 10.1091/mbc.E02-02-0116
- Felsenstein, J. (1989). Phylip: phylogeny inference package (version 3.2). *Cladistics* 5, 164–166. doi: 10.1111/j.1096-0031.1989.tb00562.x
- Festa, R. A., and Thiele, D. J. (2012). Copper at the front line of the host-pathogen battle. *PLoS Pathog.* 8:e1002887. doi: 10.1371/journal.ppat.1002887
- Fu, Y., Chang, F. M. J., and Giedroc, D. P. (2014). Copper transport and trafficking at the host-bacterial pathogen interface. *Acc. Chem. Res.* 47, 3605–3613. doi: 10.1021/ar500300n
- García-Santamarina, S., and Thiele, D. J. (2015). Copper at the fungal pathogen-host axis. *J. Biol. Chem.* 290, 18945–18953. doi: 10.1074/jbc.R115.649129
- Godinho, C. P., Prata, C. S., Pinto, S. N., Cardoso, C., Bandarra, N. M., Fernandes, F., et al. (2018). Pdr18 is involved in yeast response to acetic acid stress counteracting the decrease of plasma membrane ergosterol content and order. *Sci. Rep.* 8:7860. doi: 10.1038/s41598-018-26128-7
- Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* 59, 307–321. doi: 10.1093/sysbio/syq010
- Hirata, D., Yano, K., Miyahara, K., and Miyakawa, T. (1994). *Saccharomyces cerevisiae* YDR1, which encodes a member of the ATP-binding cassette (ABC) superfamily, is required for multidrug resistance. *Curr. Genet.* 26, 285–294. doi: 10.1007/BF00310491
- Huelsenbeck, J. P., and Ronquist, F. (2001). MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17, 754–755. doi: 10.1093/bioinformatics/17.8.754
- Huson, D. H., Richter, D. C., Rausch, C., DeZulian, T., Franz, M., and Rupp, R. (2007). Dendroscope: an interactive viewer for large phylogenetic trees. *BMC Bioinformatics* 8:460. doi: 10.1186/1471-2105-8-460
- Huson, D. H., and Scornavacca, C. (2012). Dendroscope 3: an interactive tool for rooted phylogenetic trees and networks. *Syst. Biol.* 61, 1061–1067. doi: 10.1093/sysbio/sys062
- Jandric, Z., and Schüller, C. (2011). Stress response in *Candida glabrata*: pieces of a fragmented picture. *Future Microbiol.* 6, 1475–1484. doi: 10.2217/fmb.11.131
- Jones, D. T., Taylor, W. R., and Thornton, J. M. (1992). The rapid generation of mutation data matrices from protein sequences. *Bioinformatics* 8, 275–282. doi: 10.1093/bioinformatics/8.3.275
- Jungwirth, H., and Kuchler, K. (2006). Yeast ABC transporters - A tale of sex, stress, drugs and aging. *FEBS Lett.* 580, 1131–1138. doi: 10.1016/j.febslet.2005.12.050
- Kodedová, M., and Sychrová, H. (2015). Changes in the sterol composition of the plasma membrane affect membrane potential, salt tolerance and the activity of multidrug resistance pumps in *Saccharomyces cerevisiae*. *PLoS One* 10:e0139306. doi: 10.1371/journal.pone.0139306
- Kolaczekowski, M., Kolaczowska, A., Luczynski, J., Witek, S., and Goffeau, A. (1998). In vivo characterization of the drug resistance profile of the major ABC transporters and other components of the yeast pleiotropic drug resistance network. *Microb. Drug Resist.* 4, 143–158. doi: 10.1089/mdr.1998.4.143
- Kovalchuk, A., and Driessen, A. J. M. (2010). Phylogenetic analysis of fungal ABC transporters. *BMC Genomics* 11:177. doi: 10.1186/1471-2164-11-177
- Kuchler, K., Egner, R., Rosenthal, F., and Mahé, Y. (1997). “The molecular basis for pleiotropic drug resistance in the yeast *Saccharomyces cerevisiae*: regulation of expression, intracellular trafficking and proteolytic turnover of atp binding cassette (ABC) multidrug resistance transporters,” in *Molecular Mechanisms of Signalling and Membrane Transport*, ed. K. Wirtz (Berlin: Springer-Verlag), 305–318.
- Kurtzman, C. P., and Robnett, C. J. (2003). Phylogenetic relationships among yeasts of the “*Saccharomyces* complex” determined from multigene sequence analyses. *FEMS Yeast Res.* 3, 417–432. doi: 10.1016/S1567-1356(03)00012-6
- Li, S., Ha, S. J., Kim, H. J., Galazka, J. M., Cate, J. H., Jin, Y. S., et al. (2013). Investigation of the functional role of aldose 1-pyrimidase in engineered cellobiose utilization. *J. Biotechnol.* 168, 1–6. doi: 10.1016/j.jbiotec.2013.08.003
- Ling, H., Chen, B., Kang, A., Lee, J. M., and Chang, M. W. (2013). Transcriptome response to alkane biofuels in *Saccharomyces cerevisiae*: identification of efflux pumps involved in alkane tolerance. *Biotechnol. Biofuels* 6:95. doi: 10.1186/1754-6834-6-95
- Mahé, Y., Lemoine, Y., and Kuchler, K. (1996a). The ATP binding cassette transporters Pdr5 and Snq2 of *Saccharomyces cerevisiae* can mediate transport of steroids in vivo. *J. Biol. Chem.* 271, 25167–25172. doi: 10.1074/jbc.271.41.25167
- Mahé, Y., Parle-mcdermott, A., Nourani, A., Delahodde, A., Lamprecht, A., and Kuchler, K. (1996b). The ATP-binding cassette multidrug transporter Snq2 of *Saccharomyces cerevisiae*: a novel target for the transcription factors Pdr1 and Pdr3. *Mol. Microbiol.* 20, 109–117.
- Miyahara, K., Mizunuma, M., Hirata, D., Tsuchiya, E., and Miyakawa, T. (1996). The involvement of the *Saccharomyces cerevisiae* multidrug resistance transporters Pdr5p and Snq2p in cation resistance. *FEBS Lett.* 399, 317–320. doi: 10.1016/S0014-5793(96)01353-1
- Monk, B. C., and Goffeau, A. (2008). Outwitting multidrug resistance to antifungals. *Science* 321, 367–369. doi: 10.1126/science.1159746
- Nijkamp, J. F., van den Broek, M., Datema, E., de Kok, S., Bosman, L., Luttik, M. A., et al. (2012). De novo sequencing, assembly and analysis of the genome of the laboratory strain *Saccharomyces cerevisiae* CEN.PK113-7D, a model for modern industrial biotechnology. *Microb. Cell Fact.* 11:36. doi: 10.1186/1475-2859-11-36
- Nishida, N., Ozato, N., Matsui, K., Kuroda, K., and Ueda, M. (2013). ABC transporters and cell wall proteins involved in organic solvent tolerance in *Saccharomyces cerevisiae*. *J. Biotechnol.* 165, 145–152. doi: 10.1016/j.jbiotec.2013.03.003
- Ohno, S. (1970). *Evolution by Gene Duplication*. Berlin: Springer. doi: 10.1007/978-3-642-86659-3
- Palma, M., Dias, P. J., de Canaveira Roque, F., Luzia, L., Guerreiro, J. F., and Sá-Correia, I. (2017). The *Zygosaccharomyces bailii* transcription factor Haa1 is required for acetic acid and copper stress responses suggesting subfunctionalization of the ancestral bifunctional protein Haa1/Cup2. *BMC Genomics* 18:75. doi: 10.1186/s12864-016-3443-2
- Panwar, S. L., Pasrija, R., and Prasad, R. (2008). Membrane homeostasis and multidrug resistance in yeast. *Biosci. Rep.* 28, 217–228. doi: 10.1042/BSR20080071
- Parks, L. W., and Casey, W. M. (1995). Physiological implications of sterol biosynthesis in yeast. *Annu. Rev. Microbiol.* 49, 95–116. doi: 10.1146/annurev.mi.49.100195.000523

- Paradis, E., Claude, J., and Strimmer, K. (2004). APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* 20, 289–290. doi: 10.1093/bioinformatics/btg412
- Paulsen, I. T., Sliwinski, M. K., Nelissen, B., Goffeau, A., and Saier, M. H. (1998). Unified inventory of established and putative transporters encoded within the complete genome of *Saccharomyces cerevisiae*. *FEBS Lett.* 430, 116–125. doi: 10.1016/S0014-5793(98)00629-2
- Paumi, C. M., Chuk, M., Snider, J., Stagljar, I., and Michaelis, S. (2009). ABC transporters in *saccharomyces cerevisiae* and their interactors: new technology advances the biology of the ABCC (MRP) Subfamily. *Microbiol. Mol. Biol. Rev.* 73, 577–593. doi: 10.1128/MMBR.00020-09
- Perlroth, J., Choi, B., and Spellberg, B. (2007). Nosocomial fungal infections: epidemiology, diagnosis, and treatment. *Med. Mycol.* 45, 321–346. doi: 10.1080/13693780701218689
- Peter, J., Chiara, M., De Friedrich, A., Yue, J., Pflieger, D., Bergström, A., et al. (2018). Genome evolution across 1,011 *Saccharomyces cerevisiae* isolates. *Nature* 556, 339–347. doi: 10.1038/s41586-018-0030-5
- Piecuch, A., and Oblak, E. (2014). Yeast ABC proteins involved in multidrug resistance. *Cell. Mol. Biol. Lett.* 19, 1–22. doi: 10.2478/s11658-013-0111-2
- Piper, P. W., Millson, S. H., Mollapour, M., Panaretou, B., Siligardi, G., Pearl, L. H., et al. (2003). Sensitivity to Hsp90-targeting drugs can arise with mutation to the Hsp90 chaperone, cochaperones and plasma membrane ATP binding cassette transporters of yeast. *Eur. J. Biochem.* 270, 4689–4695. doi: 10.1046/j.1432-1033.2003.03866.x
- Pomorski, T., Pomorski, T., Holthuis, J. C., Holthuis, J. C., Herrmann, A., Herrmann, A., et al. (2004). Tracking down lipid flippases and their biological functions. *J. Cell Sci.* 117, 805–813. doi: 10.1242/jcs.01055
- Popescu, A. A., Huber, K. T., and Paradis, E. (2012). Ape 3.0: new tools for distance-based phylogenetics and evolutionary analysis in R. *Bioinformatics* 28, 1536–1537. doi: 10.1093/bioinformatics/bts184
- Prasad, R., Khandelwal, N. K., and Banerjee, A. (2016). Yeast ABC transporters in lipid trafficking. *Fungal Genet. Biol.* 93, 25–34. doi: 10.1016/j.fgb.2016.05.008
- Prasad, R., and Panwar, S. (2004). Physiological functions of multidrug transporters in yeast. *Curr. Sci.* 86, 62–73.
- Roetzer, A., Gabaldón, T., and Schüller, C. (2011). From *Saccharomyces cerevisiae* to *Candida glabrata* in a few easy steps: important adaptations for an opportunistic pathogen. *FEMS Microbiol. Lett.* 314, 1–9. doi: 10.1111/j.1574-6968.2010.02102.x
- Ronquist, F., and Huelsenbeck, J. P. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19, 1572–1574. doi: 10.1093/bioinformatics/btg180
- Samanovic, M. I., Ding, C., Thiele, D. J., and Darwin, K. H. (2012). Copper in microbial pathogenesis: meddling with the metal. *Cell Host Microbe* 11, 106–115. doi: 10.1016/j.chom.2012.01.009
- Sanglard, D., Ischer, F., and Bille, J. (2001). Role of ATP-binding-cassette transporter genes in high-frequency acquisition of resistance to azole antifungals in *Candida glabrata*. *Antimicrob. Agents Chemother.* 45, 1174–1183. doi: 10.1128/AAC.45.4.1174-1183.2001
- Seret, M. L., Diffels, J. F., Goffeau, A., and Baret, P. V. (2009). Combined phylogeny and neighborhood analysis of the evolution of the ABC transporters conferring multiple drug resistance in hemiascomycete yeasts. *BMC Genomics* 10:459. doi: 10.1186/1471-2164-10-459
- Servos, J., Haase, E., and Brendel, M. (1993). Gene SNQ2 of *Saccharomyces cerevisiae*, which confers resistance to 4-nitroquinoline-N-oxide and other chemicals, encodes a 169 kDa protein homologous to ATP-dependent permeases. *Mol. Gen. Genet.* 236, 214–218. doi: 10.1007/BF00277115
- Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504. doi: 10.1101/gr.1239303
- Snider, J., Hanif, A., Lee, M. E., Jin, K., Yu, A. R., Graham, C., et al. (2013). Mapping the functional yeast ABC transporter interactome. *Nat. Chem. Biol.* 9, 565–574. doi: 10.1038/nchembio.1293
- Taylor, J. S., and Raes, J. (2004). Duplication and divergence: the evolution of new genes and old ideas. *Annu. Rev. Genet.* 38, 615–643. doi: 10.1146/annurev.genet.38.072902.092831
- Teixeira, M. C., Godinho, C. P., Cabrito, T. R., Mira, N. P., and Sá-Correia, I. (2012). Increased expression of the yeast multidrug resistance ABC transporter Pdr18 leads to increased ethanol tolerance and ethanol production in high gravity alcoholic fermentation. *Microb. Cell Fact.* 11:98. doi: 10.1186/1475-2859-11-98
- Torelli, R., Posteraro, B., Ferrari, S., La Sorda, M., Fadda, G., Sanglard, D., et al. (2008). The ATP-binding cassette transporter-encoding gene CgSNQ2 is contributing to the CgPDR1-dependent azole resistance of *Candida glabrata*. *Mol. Microbiol.* 68, 186–201. doi: 10.1111/j.1365-2958.2008.06143.x
- Tsujimoto, Y., Shimizu, Y., Otake, K., Nakamura, T., Okada, R., Miyazaki, T., et al. (2015). Multidrug resistance transporters Snq2p and Pdr5p mediate caffeine efflux in *Saccharomyces cerevisiae*. *Biosci. Biotechnol. Biochem.* 79, 1103–1110. doi: 10.1080/09168451.2015.1010476
- van Leeuwen, J. S., Ünü, B., Vermeulen, N. P. E., and Vos, J. C. (2012). Differential involvement of mitochondrial dysfunction, cytochrome P450 activity, and active transport in the toxicity of structurally related NSAIDs. *Toxicol. In Vitro* 26, 197–205. doi: 10.1016/j.tiv.2011.11.013
- Van Meer, G., Voelker, D. R., and Feigenson, G. W. (2008). Membrane lipids: where they are and how they behave. *Nat. Rev. Mol. Cell Biol.* 9, 112–124. doi: 10.1038/nrm2330
- Ververidis, P., Davrazou, F., Dailianas, G., Georgakopoulos, D., Kanellis, A. K., and Panopoulos, N. (2001). A novel putative reductase (Cpd1p) and the multidrug exporter Snq2p are involved in resistance to cercosporin and other singlet oxygen-generating photosensitizers in *Saccharomyces cerevisiae*. *Curr. Genet.* 39, 127–136. doi: 10.1007/s002940100189
- Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M., and Barton, G. J. (2009). Jalview Version 2-A multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25, 1189–1191. doi: 10.1093/bioinformatics/btp033
- Wehrschütz-Sigl, E., Jungwirth, H., Bergler, H., and Högenauer, G. (2004). The transporters Pdr5p and Snq2p mediate diazaborine resistance and are under the control of the gain-of-function allele PDR1-12. *Eur. J. Biochem.* 271, 1145–1152. doi: 10.1111/j.1432-1033.2004.04018.x
- Whelan, S., and Goldman, N. (2001). A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol. Biol. Evol.* 18, 691–699. doi: 10.1093/oxfordjournals.molbev.a003851
- Zhang, J. (2003). Evolution by gene duplication: an update. *Trends Ecol. Evol.* 18, 292–298. doi: 10.1016/S0169-5347(03)00033-8

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Godinho, Dias, Ponçot and Sá-Correia. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# The Genetic Makeup and Expression of the Glycolytic and Fermentative Pathways Are Highly Conserved Within the *Saccharomyces* Genus

Francine J. Boonekamp<sup>1</sup>, Sofia Dashko<sup>1</sup>, Marcel van den Broek<sup>1</sup>, Thies Gehrman<sup>2</sup>, Jean-Marc Daran<sup>1</sup> and Pascale Daran-Lapujade<sup>1\*</sup>

<sup>1</sup> Department of Biotechnology, Delft University of Technology, Delft, Netherlands, <sup>2</sup> Westerdijk Institute, Utrecht, Netherlands

## OPEN ACCESS

### Edited by:

Ed Louis,  
University of Leicester,  
United Kingdom

### Reviewed by:

Catherine Tesnière,  
Institut National de la Recherche  
Agronomique Centre Montpellier,  
France

Jing Hua Zhao,  
University of Cambridge,  
United Kingdom

### \*Correspondence:

Pascale Daran-Lapujade  
p.a.s.daran-lapujade@tudelft.nl

### Specialty section:

This article was submitted to  
Evolutionary and Genomic  
Microbiology,  
a section of the journal  
Frontiers in Genetics

Received: 27 July 2018

Accepted: 08 October 2018

Published: 16 November 2018

### Citation:

Boonekamp FJ, Dashko S,  
van den Broek M, Gehrman T,  
Daran J-M and Daran-Lapujade P  
(2018) The Genetic Makeup  
and Expression of the Glycolytic  
and Fermentative Pathways Are  
Highly Conserved Within  
the *Saccharomyces* Genus.  
Front. Genet. 9:504.  
doi: 10.3389/fgene.2018.00504

The ability of the yeast *Saccharomyces cerevisiae* to convert glucose, even in the presence of oxygen, via glycolysis and the fermentative pathway to ethanol has played an important role in its domestication. Despite the extensive knowledge on these pathways in *S. cerevisiae*, relatively little is known about their genetic makeup in other industrially relevant *Saccharomyces* yeast species. In this study we explore the diversity of the glycolytic and fermentative pathways within the *Saccharomyces* genus using *S. cerevisiae*, *Saccharomyces kudriavzevii*, and *Saccharomyces eubayanus* as paradigms. Sequencing data revealed a highly conserved genetic makeup of the glycolytic and fermentative pathways in the three species in terms of number of paralogous genes. Although promoter regions were less conserved between the three species as compared to coding sequences, binding sites for Rap1, Gcr1 and Abf1, main transcriptional regulators of glycolytic and fermentative genes, were highly conserved. Transcriptome profiling of these three strains grown in aerobic batch cultivation in chemically defined medium with glucose as carbon source, revealed a remarkably similar expression of the glycolytic and fermentative genes across species, and the conserved classification of genes into major and minor paralogs. Furthermore, transplantation of the promoters of major paralogs of *S. kudriavzevii* and *S. eubayanus* into *S. cerevisiae* demonstrated not only the transferability of these promoters, but also the similarity of their strength and response to various environmental stimuli. The relatively low homology of *S. kudriavzevii* and *S. eubayanus* promoters to their *S. cerevisiae* relatives makes them very attractive alternatives for strain construction in *S. cerevisiae*, thereby expanding the *S. cerevisiae* molecular toolbox.

**Keywords:** glycolysis, promoter characterization, *Saccharomyces cerevisiae*, *Saccharomyces kudriavzevii*, *Saccharomyces eubayanus*, transcription factor binding sites

## INTRODUCTION

The yeast *Saccharomyces cerevisiae* is known for its fast fermentative metabolism, which has played an important role in its domestication (Sicard and Legras, 2011). *S. cerevisiae* converts glucose to ethanol via the Embden-Meyerhof-Parnas pathway of glycolysis and the fermentative pathway, encompassing a total of 12 enzymatic steps (Barnett, 2003; Barnett and Entian, 2005). While *S. cerevisiae* can respire glucose, leading to an ATP yield of 16 moles of ATP per mole of glucose, it favors alcoholic fermentation. Indeed, even in the presence of oxygen, glucose excess triggers

ethanol formation in *S. cerevisiae* and its relatives from the *Saccharomyces* genus, a phenomenon known as the Crabtree effect (De Deken, 1966; Merico et al., 2007). To sustain the energy demand for growth and maintenance despite the low ATP yield of alcoholic fermentation (2 moles of ATP per glucose molecule), the glycolytic flux in *S. cerevisiae* can easily reach fluxes of 20–25 mmol ethanol per gram dry weight per hour (Solis-Escalante et al., 2015). This high activity of the glycolytic pathway is reflected in the remarkably high concentration of glycolytic enzymes in the cell, which can represent up to 30% of the total amount of soluble protein (Fraenkel, 2003; Carroll et al., 2011).

The genome of *S. cerevisiae* is characterized by a high genetic redundancy which can largely be attributed to a whole genome duplication event (Ohno, 1970; Wolfe and Shields, 1997). This redundancy is even more prominent among 'metabolic' genes and is remarkably elevated in the glycolytic and fermentative pathways of *S. cerevisiae* (Kellis et al., 2004; Kuepfer et al., 2005; Conant and Wolfe, 2007). These two pathways have been thoroughly investigated (and even established) in *S. cerevisiae* (Barnett, 2003; Van Heerden et al., 2015). With the exception of three steps that are catalyzed by single enzymes, i.e., phosphoglucose isomerase (Pgi1), fructose-bisphosphate aldolase (Fba1), and triosephosphate isomerase (Tpi1), the glycolytic, and fermentative steps are catalyzed by at least two and potentially up to seven isoenzymes for alcohol dehydrogenase (Adh). However, not all isoenzymes are equally important for the glycolytic and fermentative activity. With the notable exception of Pfk1 and Pfk2, two isoenzymes forming a heterooctamer that are equally important for the functionality of phosphofructokinase (Heinisch et al., 1991; Arvanitidis and Heinisch, 1994), for each step, a single isoenzyme is responsible for the bulk of the glycolytic and fermentative flux. These so-called major isoenzymes are encoded by major paralogs, which expression is strong and constitutive (i.e., *HXK2*, *TDH3*, *GPM1*, *ENO2*, *PYK1*, *PDC1*, *ADH1*) (Solis-Escalante et al., 2015). Because of these properties, glycolytic promoters are often used to drive gene expression in engineered strains (Peng et al., 2015). Conversely the expression of minor paralogs is, in most instances, far lower than the expression of the corresponding major paralogs and is condition-dependent (Ciriacy, 1979; Boer et al., 2003; Knijnenburg et al., 2009; Solis-Escalante et al., 2015). Following duplication events, redundant genes can have different fates. If their presence brings additional benefits to the cell, either in their native form or via neo-functionalization, the gene and its duplicate will be retained in the genome, otherwise the redundant copy will be lost (Kellis et al., 2004; Conant and Wolfe, 2008). The fact that the glycolytic and fermentative pathways still contain many paralogs that do not display obvious new functions suggests that they might increase fitness under specific conditions. For example, *PDC6* encoding a pyruvate decarboxylase with low sulfur amino acid content is specifically induced in sulfur limiting conditions (Fauchon et al., 2002; Boer et al., 2003). However, challenging this theory, it was recently shown that the simultaneous removal of all minor paralogs from the glycolytic and fermentative pathways had no detectable effect on *S. cerevisiae* physiology under a wide variety of conditions (Solis-Escalante et al., 2015).

The *Saccharomyces* genus consists of at least eight naturally occurring species which all evolved toward optimal performance in their niche, leading to different physiological characteristics (Replansky et al., 2008; Hittinger, 2013; Naseeb et al., 2017). For instance, *Saccharomyces kudriavzevii*, *Saccharomyces uvarum* and *Saccharomyces eubayanus* are cold-tolerant, and perform better than *S. cerevisiae* at temperatures below 20°C (Arroyo-López et al., 2010; Masneuf-Pomarède et al., 2010; Salvadó et al., 2011; Hebly et al., 2015). Strains belonging to different *Saccharomyces* species can mate and form viable hybrids, some of which play an important role in the beverage industry. For instance *Saccharomyces pastorianus*, a hybrid of *S. cerevisiae* and *S. eubayanus*, is the main lager-brewing yeast (Bond, 2009) and hybrids of *S. cerevisiae* and *S. kudriavzevii* and of *S. uvarum* and *S. eubayanus* (known as *S. bayanus*) play an important role in beer and wine fermentation (González et al., 2006; González et al., 2008; Peris et al., 2012; Nguyen and Boekhout, 2017). The cold-tolerance of *S. pastorianus* and *S. eubayanus* has undoubtedly promoted the selection of their hybrids with *S. cerevisiae* in cold environments (Belloch et al., 2008; Arroyo-López et al., 2010; Libkind et al., 2011).

In a recent study, using a unique yeast platform enabling the swapping of entire essential pathways, it was shown that *S. kudriavzevii* glycolytic and fermentative pathways could be transplanted in *S. cerevisiae* and could efficiently complement the native pathways. Expression of the full set of *S. kudriavzevii* orthologs in *S. cerevisiae*, expressed from *S. kudriavzevii* promoters, resulted in enzyme activities and physiological responses remarkably similar to the parental strain carrying a full set of native *S. cerevisiae* genes. However, the impact of *S. kudriavzevii* promoters on transcriptional activity in *S. cerevisiae* was not explored (Kuijpers et al., 2016). Despite *S. eubayanus* and *S. kudriavzevii* industrial importance and the availability of their full genome sequence, remarkably little is known about the genetic makeup and transcriptional regulation of the glycolytic and fermentative pathways.

To address this knowledge gap, the present study explores the diversity of the glycolytic and fermentative pathways within the genus *Saccharomyces*, using the industrially relevant yeasts *S. cerevisiae*, *S. eubayanus* and *S. kudriavzevii* as paradigms. More precisely, the presence and sequence similarity between paralogs in these three yeasts were explored. Cultivation in bioreactors combined with transcriptome analysis was used to evaluate the presence of dominant paralogs in *S. eubayanus* and *S. kudriavzevii* and to compare the expression levels of glycolytic and fermentative orthologs in their native context. Finally, we explored transferability of *S. kudriavzevii* and *S. eubayanus* promoters by monitoring their expression and context-dependency upon transplantation in *S. cerevisiae*.

## MATERIALS AND METHODS

### Strains and Culture Conditions

All yeast strains used in this study are derived from the CEN.PK background (Entian and Kötter, 2007) and are listed in **Table 1**. Yeast cultures for transformation and genomic DNA isolation



**TABLE 1 |** Strains table.

Strain	Genotype	Plasmid for integration	Source
<b><i>Saccharomyces cerevisiae</i></b>			
CEN.PK113-5D	<i>MATa ura3-52 HIS3 LEU2 TRP1 MAL2-8c SUC2</i>	—	Entian and Kötter, 2007
CEN.PK113-7D	<i>MATa URA3 HIS3 LEU2 TRP1 MAL2-8c SUC2</i>	—	Van Dijken et al., 2000; Entian and Kötter, 2007; Nijkamp et al., 2012; Salazar et al., 2017
CEN.PK122	<i>MATa/Mata</i>	—	Van Dijken et al., 2000; Entian and Kötter, 2007
IMX1042	<i>MATa HIS3 LEU2 TRP1 MAL2-8c SUC2 ura3:: (pHXK2sc-mRuby2-tENO2, URA3)</i>	pUDI098	This study
IMX1016	<i>MATa HIS3 LEU2 TRP1 MAL2-8c SUC2 ura3:: (pHXK2sk-mRuby2-tENO2, URA3)</i>	pUDI097	This study
IMX1102	<i>MATa HIS3 LEU2 TRP1 MAL2-8c SUC2 ura3:: (pHXK2se-mRuby2-tENO2, URA3)</i>	pUDI108	This study
IMX1068	<i>MATa HIS3 LEU2 TRP1 MAL2-8c SUC2 ura3:: (pPGI1sc-mRuby2-tENO2, URA3)</i>	pUDI101	This study
IMX1017	<i>MATa HIS3 LEU2 TRP1 MAL2-8c SUC2 ura3:: (pPGI1sk-mRuby2-tENO2, URA3)</i>	pUDI095	This study
IMX1103	<i>MATa HIS3 LEU2 TRP1 MAL2-8c SUC2 ura3:: (pPGI1se-mRuby2-tENO2, URA3)</i>	pUDI109	This study
IMX1171	<i>MATa HIS3 LEU2 TRP1 MAL2-8c SUC2 ura3:: (pPFK1sc-mRuby2-tENO2, URA3)</i>	pUDI121	This study
IMX1249	<i>MATa HIS3 LEU2 TRP1 MAL2-8c SUC2 ura3:: (pPFK1sk-mRuby2-tENO2, URA3)</i>	pUDI126	This study
IMX1174	<i>MATa HIS3 LEU2 TRP1 MAL2-8c SUC2 ura3:: (pPFK1se-mRuby2-tENO2, URA3)</i>	pUDI118	This study
IMX1175	<i>MATa HIS3 LEU2 TRP1 MAL2-8c SUC2 ura3:: (pPFK2sc-mRuby2-tENO2, URA3)</i>	pUDI131	This study
IMX1176	<i>MATa HIS3 LEU2 TRP1 MAL2-8c SUC2 ura3:: (pPFK2sk-mRuby2-tENO2, URA3)</i>	pUDI130	This study
IMX1177	<i>MATa HIS3 LEU2 TRP1 MAL2-8c SUC2 ura3:: (pPFK2se-mRuby2-tENO2, URA3)</i>	pUDI132	This study
IMX1041	<i>MATa HIS3 LEU2 TRP1 MAL2-8c SUC2 ura3:: (pFBA1sc-mRuby2-tENO2, URA3)</i>	pUDI099	This study
IMX1070	<i>MATa HIS3 LEU2 TRP1 MAL2-8c SUC2 ura3:: (pFBA1sk-mRuby2-tENO2, URA3)</i>	pUDI103	This study
IMX1097	<i>MATa HIS3 LEU2 TRP1 MAL2-8c SUC2 ura3:: (pFBA1se-mRuby2-tENO2, URA3)</i>	pUDI186	This study
IMX1132	<i>MATa HIS3 LEU2 TRP1 MAL2-8c SUC2 ura3:: (pTPI1sc-mRuby2-tENO2, URA3)</i>	pUDI114	This study
IMX1133	<i>MATa HIS3 LEU2 TRP1 MAL2-8c SUC2 ura3:: (pTPI1sk-mRuby2-tENO2, URA3)</i>	pUDI115	This study
IMX1134	<i>MATa HIS3 LEU2 TRP1 MAL2-8c SUC2 ura3:: (pTPI1se-mRuby2-tENO2, URA3)</i>	pUDI116	This study
IMX1018	<i>MATa HIS3 LEU2 TRP1 MAL2-8c SUC2 ura3:: (pTDH3sc-mRuby2-tENO2, URA3)</i>	pUDI094	This study
IMX1128	<i>MATa HIS3 LEU2 TRP1 MAL2-8c SUC2 ura3:: (pTDH3sk-mRuby2-tENO2, URA3)</i>	pUDI110	This study
IMX1130	<i>MATa HIS3 LEU2 TRP1 MAL2-8c SUC2 ura3:: (pTDH3se-mRuby2-tENO2, URA3)</i>	pUDI112	This study
IMX1043	<i>MATa HIS3 LEU2 TRP1 MAL2-8c SUC2 ura3:: (pPGK1sc-mRuby2-tENO2, URA3)</i>	pUDI100	This study
IMX1019	<i>MATa HIS3 LEU2 TRP1 MAL2-8c SUC2 ura3:: (pPGK1sk-mRuby2-tENO2, URA3)</i>	pUDI096	This study
IMX1069	<i>MATa HIS3 LEU2 TRP1 MAL2-8c SUC2 ura3:: (pPGK1se-mRuby2-tENO2, URA3)</i>	pUDI102	This study
IMX1100	<i>MATa HIS3 LEU2 TRP1 MAL2-8c SUC2 ura3:: (pGPM1sc-mRuby2-tENO2, URA3)</i>	pUDI106	This study
IMX1071	<i>MATa HIS3 LEU2 TRP1 MAL2-8c SUC2 ura3:: (pGPM1sk-mRuby2-tENO2, URA3)</i>	pUDI104	This study
IMX1101	<i>MATa HIS3 LEU2 TRP1 MAL2-8c SUC2 ura3:: (pGPM1se-mRuby2-tENO2, URA3)</i>	pUDI107	This study
IMX1178	<i>MATa HIS3 LEU2 TRP1 MAL2-8c SUC2 ura3:: (pENO2sc-mRuby2-tENO2, URA3)</i>	pUDI122	This study
IMX1299	<i>MATa HIS3 LEU2 TRP1 MAL2-8c SUC2 ura3:: (pENO2sk-mRuby2-tENO2, URA3)</i>	pUDI123	This study
IMX1180	<i>MATa HIS3 LEU2 TRP1 MAL2-8c SUC2 ura3:: (pENO2se-mRuby2-tENO2, URA3)</i>	pUDI119	This study
IMX1181	<i>MATa HIS3 LEU2 TRP1 MAL2-8c SUC2 ura3:: (pPYK1sc-mRuby2-tENO2, URA3)</i>	pUDI128	This study
IMX1182	<i>MATa HIS3 LEU2 TRP1 MAL2-8c SUC2 ura3:: (pPYK1sk-mRuby2-tENO2, URA3)</i>	pUDI127	This study
IMX1183	<i>MATa HIS3 LEU2 TRP1 MAL2-8c SUC2 ura3:: (pPYK1se-mRuby2-tENO2, URA3)</i>	pUDI129	This study
IMX1242	<i>MATa HIS3 LEU2 TRP1 MAL2-8c SUC2 ura3:: (pPDC1sc-mRuby2-tENO2, URA3)</i>	pUDI161	This study
IMX1243	<i>MATa HIS3 LEU2 TRP1 MAL2-8c SUC2 ura3:: (pPDC1sk-mRuby2-tENO2, URA3)</i>	pUDI162	This study
IMX1244	<i>MATa HIS3 LEU2 TRP1 MAL2-8c SUC2 ura3:: (pPDC1se-mRuby2-tENO2, URA3)</i>	pUDI163	This study
IMX1245	<i>MATa HIS3 LEU2 TRP1 MAL2-8c SUC2 ura3:: (pADH1sc-mRuby2-tENO2, URA3)</i>	pUDI158	This study
IMX1246	<i>MATa HIS3 LEU2 TRP1 MAL2-8c SUC2 ura3:: (pADH1sk-mRuby2-tENO2, URA3)</i>	pUDI159	This study
IMX1298	<i>MATa HIS3 LEU2 TRP1 MAL2-8c SUC2 ura3:: (pADH1se-mRuby2-tENO2, URA3)</i>	pUDI160	This study
IMX1099	<i>MATa HIS3 LEU2 TRP1 MAL2-8c SUC2 ura3:: (pACT1sc-mRuby2-tENO2, URA3)</i>	pUDI105	This study
IMX1168	<i>MATa HIS3 LEU2 TRP1 MAL2-8c SUC2 ura3:: (pTEF1sc-mRuby2-tENO2, URA3)</i>	pUDI124	This study
<b><i>Other Saccharomyces species</i></b>			
<i>S. kudriavzevii</i> CR85	<i>MATa/Mata</i>	—	Lopes et al., 2010
<i>S. eubayanus</i>	<i>MATa/Mata</i>	—	Libkind et al., 2011
CBS12357			

were grown in 500 mL shake flasks with 100 mL of complex, non-selective medium (YPD) containing 10 g L<sup>-1</sup> Bacto Yeast extract, 20 g L<sup>-1</sup> Bacto Peptone and 20 g L<sup>-1</sup> glucose. Promoter regions were obtained from the strains *S. cerevisiae* CEN.PK113-7D (Van Dijken et al., 2000; Entian and Kötter, 2007; Nijkamp et al., 2012), *S. kudriavzevii* CR85 a wild isolate from oak bark (supplied by Prof. Querol and dr. Barrio, Universitat de València, Spain) (Lopes et al., 2010) and *S. eubayanus* CBS12357 (Libkind et al., 2011). The same strains were used for transcriptome analysis, with the exception of *S. cerevisiae* for which the diploid strain CEN.PK122 was used instead of the haploid CEN.PK113-7D (Entian and Kötter, 2007). All *S. cerevisiae* strains were grown at 30°C and *S. kudriavzevii* and *S. eubayanus* at 20°C in shake flasks at 200 rpm, unless different conditions are mentioned.

All transformations were done in *S. cerevisiae* CEN.PK113-5D using the auxotrophic marker *URA3* for selection. Synthetic medium containing 3 g L<sup>-1</sup> KH<sub>2</sub>PO<sub>4</sub>, 0.5 g L<sup>-1</sup> MgSO<sub>4</sub>·7H<sub>2</sub>O, 5 g L<sup>-1</sup> (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 1 mL L<sup>-1</sup> of a trace element solution, and 1 mL L<sup>-1</sup> of a vitamin solution was used (Verduyn et al., 1992). Synthetic medium supplemented with 20 g L<sup>-1</sup> glucose (SMG) or 2% (vol/vol) ethanol (SMEtOH) was used for culture propagation where specified. For solid media 20 g L<sup>-1</sup> agar was added prior to heat sterilization. For storage and propagation of plasmids *Escherichia coli* XL1-Blue (Agilent Technologies, Santa Clara, CA, United States) was used, and grown in lysogeny broth (LB) supplemented with ampicillin (100 mg L<sup>-1</sup>) (Bertani, 1951; Bertani, 2004). For the storage of yeast and *E. coli* strains 30% or 15% (v/v) glycerol was added to exponentially growing cultures respectively, and aliquots were stored at -80°C.

## Molecular Biology Techniques

For high fidelity PCR amplification Phusion high fidelity polymerase (Thermo Scientific, Landsmeer, Netherlands) was used according to manufacturer's instructions. To improve efficiency of the PCR reactions, primer concentrations were decreased from 500 to 200 nM and the polymerase concentration was increased from 0.02 to 0.03 μL<sup>-1</sup>. PCR products were treated with 1 μL DpnI FastDigest restriction enzyme (Thermo Fisher Scientific) for 1 h at 37°C to remove residual circular templates. Afterward, the mixture was purified using GenElute™ PCR Clean-Up Kit (Sigma-Aldrich, St. Louis, MO) according to manufacturer's protocol. PCR for diagnostic purposes was done using DreamTaq PCR mastermix (Thermo Fisher Scientific) according to manufacturer's recommendations. Primers used in this study are listed in **Supplementary Tables S1, S2**. PCR products were resolved on 1% agarose gel with Tris-acetate-EDTA (TAE) buffer. Genomic DNA used as template for PCR amplification of the promoter regions was isolated using YeaStar genomic DNA kit (Zymo Research, Orange, CA) according to manufacturer's protocol. Plasmids were extracted from *E. coli* using the GenElute plasmid miniprep kit (Sigma-Aldrich) according to manufacturer's description and eluted with miliQ water. Restriction analysis of plasmids was done using FastDigest restriction enzymes with FastDigest Green Buffer (Thermo Fisher Scientific) incubating for 30 min at 37°C according to manufacturer's recommendations.

## Promoters, Plasmids and Yeast Strain Construction

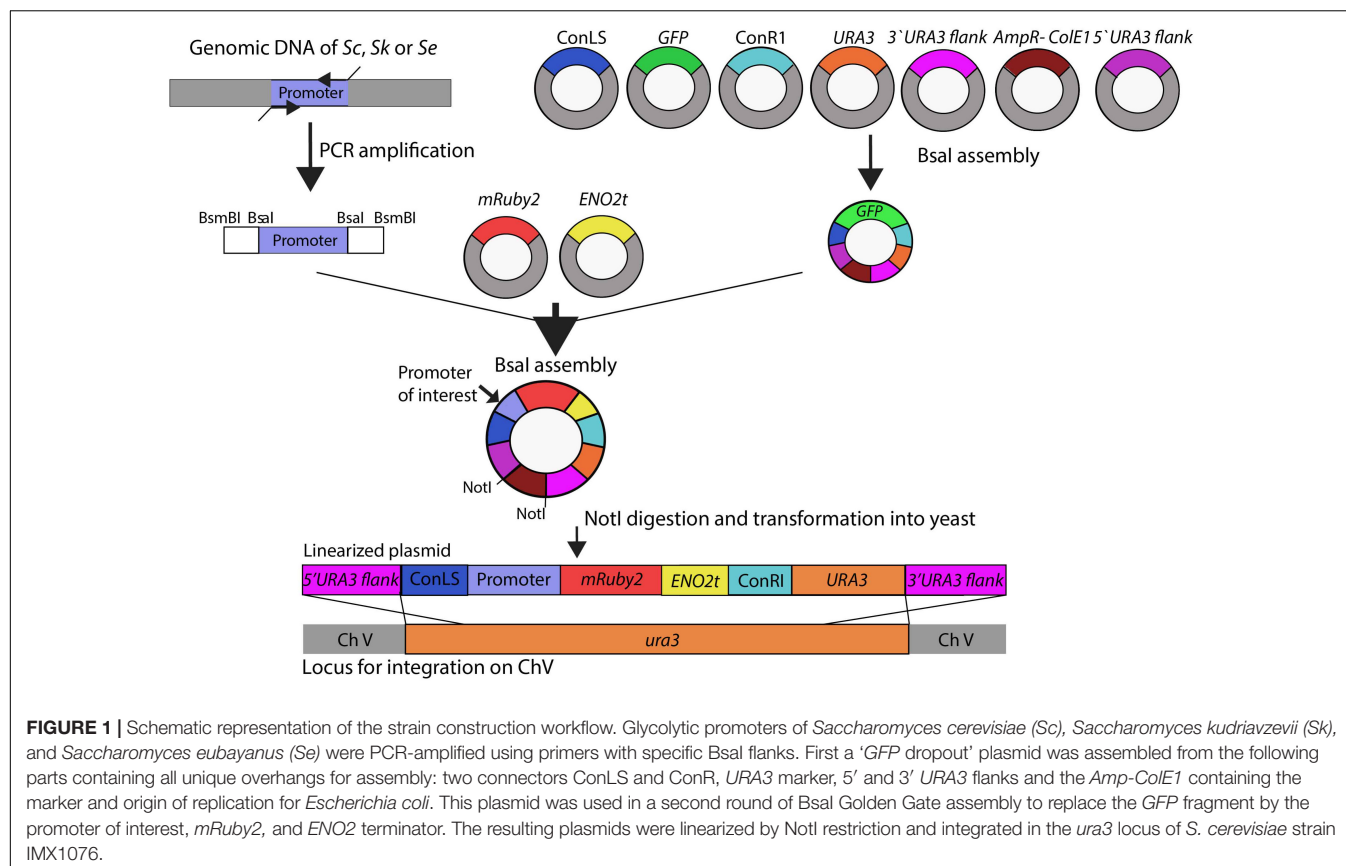
A schematic overview of the subsequent plasmid and strain construction steps is provided in **Figure 1**. Plasmids used in this study are reported in **Supplementary Table S3**. The *HXK2*, *PGI1*, *PFK1*, *PFK2*, *FBA1*, *TPI1*, *TDH3*, *PGK1*, *GPM1*, *ENO2*, *PYK1*, *PDC1* and *ADH1*, and reference *TEF1* and *ACT1* promoter regions of approximately 800 bp (see **Supplementary Table S4** for exact lengths) were PCR-amplified from *S. cerevisiae* CEN.PK113-7D, *S. kudriavzevii* CR85 and *S. eubayanus* CBS 12357 genomic DNA using primers listed in **Supplementary Table S1**. For compatibility with Golden Gate cloning, promoter sequences were flanked with BsaI and BsmBI restriction sites introduced as primer overhangs in the PCR amplification step.

The plasmid backbone was constructed by Golden Gate assembly using the collection of part plasmids provided in the Yeast Toolkit (Lee et al., 2015). To increase the efficiency of plasmid assembly, first a *GFP* dropout plasmid pUD428 was constructed containing a *URA3* marker, AmpR selection marker, bacterial origin of replication, two connector fragments and a *GFP* gene surrounded by *URA3* upstream and downstream homology flanks (**Supplementary Table S3**). The correct assembly of plasmids was checked by restriction analysis. The *GFP* dropout cassette from pUD428 was subsequently replaced by the *mRuby2* gene flanked by a promoter of interest and by the *ENO2* terminator using Golden Gate cloning with BsaI. The reaction mixture was prepared with 1 μL T4 DNA ligase buffer (Thermo Fisher Scientific), 0.5 μL T7 DNA ligase (NEB New England Biolabs, Ipswich, MA), 0.5 μL FastDigest Eco31I (BsaI) (Thermo Fisher Scientific) and 10 ng of each DNA fragment. MiliQ H<sub>2</sub>O was added to a final volume of 10 μL. The assembly was done in a thermocycler using 25 cycles of restriction and ligation: 42°C for 2 min, 16°C for 5 min, followed by a final digestion step (60°C for 10 min) and an inactivation step (80°C for 10 min). If one of the fragments contained an internal BsaI site, the final digestion and inactivation steps were omitted. 1 μL of the assembly mix was transformed to *E. coli* (XL1-Blue) according to manufacturer's description and plated on selective LB medium. Correct ligation of the promoter-*mRuby2*-terminator construct in this plasmid resulted in the loss of the *GFP* gene, which could be easily screened based on colony color. Additional plasmid confirmation was done by restriction analysis.

Prior to transformation into yeast, the constructed plasmids containing the promoter of interest, the *mRuby2* gene and the *ENO2* terminator were linearized by digestion with NotI (FastDigest, Thermo Fisher Scientific) according to manufacturer's protocol for 30 min at 37°C. 400 ng of each plasmid was digested and the mixture was directly transformed to the strain CEN.PK113-5D in which the linearized plasmid was integrated in the *ura3-52* locus. Yeast transformations were done according to Gietz and Woods (Gietz and Woods, 2002). Colonies were screened by PCR (**Supplementary Table S2**).

## Batch Cultivation in Bioreactors

Samples for transcriptome analysis of *S. cerevisiae* (CEN.PK122), *S. kudriavzevii* (CR85) and *S. eubayanus* (CBS 12357) were obtained from aerobic batch cultures in bioreactors performed in



independent duplicate. Batch cultures were performed in SMG supplemented with 0.2 g L<sup>-1</sup> antifoam Emulsion C (Sigma-Aldrich). The reactors were inoculated at a starting OD<sub>660</sub> of 0.3 with cells resuspended in demineralized water, which were obtained from exponentially growing shake flask cultures incubated at the same temperature and with the same medium as was used in the bioreactors (SMG). Cultures were performed in 2 L bioreactors (Applikon, Schiedam, The Netherlands) containing a 1.4 L working volume. The cultures were constantly stirred at 800 rpm, sparged with 700 mL min<sup>-1</sup> dried compressed air (Linde Gas Benelux, Schiedam, The Netherlands) and maintained at 30°C for *S. cerevisiae* and 25°C for *S. kudriavzevii* and *S. eubayanus*. The culture pH was kept at 5.0 during growth on glucose by automatic addition of 2M KOH.

Extracellular metabolites were determined by high-performance liquid chromatography (HPLC) analysis using a Aminex HPX-87H ion-exchange column operated at 60°C with 5 mM H<sub>2</sub>SO<sub>4</sub> as the mobile phase at a flow rate of 0.6 mL min<sup>-1</sup> (Agilent, Santa Clara). Samples were centrifuged for 3 min at 20,000 g and the supernatant was used for analysis.

Biomass dry weight was determined in analytical duplicate by filtration of 10 mL sample on filters (pore-size 0.45 μm, Whatman/GE Healthcare Life Sciences, Little Chalfont, United Kingdom) pre-dried in a microwave oven at 360 W for 20 min, as previously described (Verduyn et al., 1992). Optical density at 660 nm (OD<sub>660</sub>) was determined in a Libra S11 spectrophotometer (Biocrom, Cambridge, United Kingdom).

The CO<sub>2</sub> and O<sub>2</sub> concentration in the gas outflow was analyzed by a Rosemount NGA 2000 analyser (Baar, Switzerland), after cooling of the gas by a condenser (2°C) and drying using a PermaPure Dryer (model MD 110-8P-4; Inacom Instruments, Veenendaal, Netherlands). Sampling for transcriptome analysis was done during mid-exponential growth on glucose at a biomass concentration of approximately 1 g L<sup>-1</sup>. Sampling in liquid nitrogen and RNA extraction were performed as previously described (Piper et al., 2002).

## Promoter Activity Assay

Promoter activity measurement of the mRuby2 reporter strain library was performed in 96-well plates. Precultures were grown in 12-well plates in 1.5 mL volume in a thermoshaker (Grant-bio PHMP-4, United Kingdom) with constant shaking (800 rpm) and temperature. Precultures were grown at the temperature of the subsequent plate assay (30°C or 20°C). For the first preculture YPD medium was inoculated from glycerol stocks and grown overnight till saturation. From this culture 20 μL were transferred to new 12-well plates and the strains were grown under the conditions of interest till mid-exponential phase (corresponding to OD<sub>660</sub> of 3 to 5). Afterward the culture was centrifuged at 3000 g for 5 min, the supernatant was removed and cells were resuspended in fresh medium to an OD<sub>660</sub> of 0.3 and transferred in volumes of 100 μL to a 96-well plate (Corning™ polystyrene white/transparent bottom, Greiner Bio-One) using six replicate wells per strain. To prevent evaporation, all plates,



including preculture plates, were covered with sterile polyester acrylate sealing tape (Thermo Scientific). To supply sufficient levels of oxygen throughout the cultures, small openings were created in each well with a needle. The plate assays were performed in a plate reader (TECAN infinite M200 Pro, Tecan, Männedorf, Switzerland) with constant temperature and shaking (orbital, 1 mm). Every 20 min the optical density (OD<sub>660</sub>) and the fluorescence using excitation and emission wavelengths 559 nm/600 nm were measured. Cultures were monitored till saturation. A non-fluorescent CEN.PK113-7D strain was taken along every run to determine the background fluorescence, as well as two reference reporter strains expressing *mRuby2* from the *TEF1* and *ACT1* promoters from *S. cerevisiae*. For every well, OD<sub>660</sub> and fluorescence values from all time points during exponential growth were plotted against each other and the promoter activity was calculated as the slope of the linear regression between optical density and fluorescence.

## Flow Cytometry Analysis

*mRuby2* fluorescence intensity of individual cells from cultures grown in the TECAN plate reader was determined using flow cytometry. Mid-exponential cultures from the plate reader were diluted in Isoton II (Beckman Coulter, Brea, CA) and the fluorescence intensity was determined for 10000 cells per sample on a BD FACSARIAII (Franklin Lakes, NJ) equipped with an 561 nm excitation laser and 582/15 nm emission filter. Data were analyzed using FlowJo v10.2 (FlowJo LLC). As expected from strains in which the *mRuby2* expression system is integrated in the genome the fluorescence signal was homogeneously distributed among the yeast population (**Supplementary Figure S5**).

## Whole Genome Sequencing

To obtain genome sequences of high quality, the strain *S. kudriavzevii* CR85 was sequenced in-house both by Illumina Miseq sequencing (Illumina, San Diego, CA) and by Oxford Nanopore Technology MinION sequencing (Oxford Nanopore Technology, Oxford, United Kingdom). Genomic DNA was isolated using the Qiagen 100/G kit (Qiagen, Hilden, Germany) and the concentration was determined using Qubit® Fluorometer 2.0 (ThermoFisher Scientific). Illumina library preparation was done as described previously (Świat et al., 2017).

For Nanopore sequencing, 3 µg of genomic DNA were diluted in a total volume of 46 µL and then sheared with a g-TUBE (Covaris, Brighton, United Kingdom) to an average fragment size of 8–10 kb. The input DNA was then prepared for loading in a FLO-MIN106 flow cell with R9.4 chemistry and the 1D ligation sequencing kit (SQK-LSK108), following manufacturer's instructions with the exception of a size selection step with 0.4x (instead of 1x) AMPure beads after the End-Repair/dA tailing module and the use of 80% (instead of 70%) ethanol for washes. Raw files generated by MinKNOW were base called using Albacore (version 1.2.5; Oxford Nanopore Technology). Reads, in fastq format, with minimum length of 1000 bp were extracted, yielding 4.15 Gigabase sequence with an average read length of 4.3 kb.

*De novo* assembly was performed using Canu (v1.4, settings: genomesize = 12m) (Koren et al., 2017) producing an 11.87

Megabase genome into 20 contigs of which 13 contigs in chromosome length plus 1 mitochondrial DNA, while 3 chromosomes consisted of 2 contigs each. The contig pairs were manually joined (with 1000 N's between the contigs) into 3 chromosomes (chromosomes VII, XII, and XVI). Pilon (Walker et al., 2014) was then used to further correct assembly errors by aligning Illumina reads, using BWA (Li and Durbin, 2010) to the assembly using correction of only SNPs and short indels (–fix bases parameter). Gene annotations were performed using the MAKER2 annotation pipeline (version 2.31.9) (Holt and Yandell, 2011) using SNAP (version 2013–11–29) (Korf, 2004) and Augustus (version 3.2.3) (Stanke and Waack, 2003) as *ab initio* gene predictors. S288C EST and protein sequences were obtained from SGD (*Saccharomyces* Genome Database<sup>1</sup>) and were aligned using BLASTX (BLAST version 2.2.28+) (Camacho et al., 2009). The translated protein sequence of the final gene model was aligned using BLASTP to S288C protein Swiss-Prot database<sup>2</sup>. For CEN.PK113-7D and *S. eubayanus* CBS 12357 existing sequencing data was used (Baker et al., 2015; Salazar et al., 2017). The sequencing data are available at NCBI under bioproject accession number PRJNA480800.

## RNA Sequencing and Data Analysis

Library preparation and RNA sequencing were performed by Novogene Bioinformatics Technology Co., Ltd. (Yuen Long, Hong Kong). Sequencing was done with Illumina paired end 150 bp sequencing read system (PE150) using a 250~300 bp insert strand specific library which was prepared by Novogene. For the library preparation, as described by Novogene, mRNA enrichment was done using oligo(dT) beads. After random fragmentation of the mRNA, cDNA was synthesized from the mRNA using random hexamers primers. Afterward, second strand synthesis was done by addition of a custom second strand synthesis buffer (Illumina), dNTPs, RNase H and DNA polymerase I. Finally, after terminal repair, A ligation and adaptor ligation, the double stranded cDNA library was finalized by size selection and PCR enrichment.

The sequencing data for the three strains, *S. cerevisiae* CEN.PK122, *S. kudriavzevii* CR85 and *S. eubayanus* CBS 12357 obtained by Novogene had an average read depth of 21, 24, and 24 million reads per sample, respectively. For each sample, reads were aligned to the relevant reference genome using a two-pass STAR procedure (Dobin et al., 2013). In the first pass, we assembled a splice junction database which was used to inform the second round of alignments. As paralogs in the glycolytic pathways were highly similar, we used stricter criteria for aligning and counting reads to facilitate delineation of paralogs. Introns were allowed to be between 15 and 4000 bp, and soft clipping was disabled to prevent low quality reads from being spuriously aligned. Ambiguously mapped reads were removed. Expression was quantified per transcript using htseq-count in strict intersection mode (Anders et al., 2015). As we wished to compare gene expression across genomes, where orthologs may have different gene lengths,

<sup>1</sup><http://www.yeastgenome.org>

<sup>2</sup><http://www.ebi.ac.uk/swissprot/>



data were normalized for gene length. Therefore the average FPKM expression counts for each gene in each species were calculated (Trapnell et al., 2010). The genomes from *S. cerevisiae* CEN.PK113-7D, *S. kudriavzevii* CR85 and *S. eubayanus* CBS 12357 were used as reference NCBI BioProject accession numbers PRJNA52955, PRJNA480800, and PRJNA264003 respectively<sup>3</sup>. Data are available at Gene Expression Omnibus with accession number GSE117404. CEN.PK113-7D transcriptome data is available on Gene Expression Omnibus database under accession number GSE63884.

## Comparison of DNA Sequences

Sequences from annotated glycolytic ORF and promoters of *S. cerevisiae* CEN.PK113-7D, *S. kudriavzevii* CR85 and *S. eubayanus* CBS 12357 were used for alignments with Clone Manager 9 Professional Edition, NCBI BioProject accession numbers PRJNA52955, PRJNA480800, and PRJNA264003 respectively. For the *TP11* sequence alignment the sequences with the following accession numbers were used: CU928179 (*Z. rouxii*), HE605205 (*C. parapsilosis*), CP028453 (*Y. lipolytica*), AJ390491 (*C. albicans*), XM\_002551264 (*C. tropicalis*), AJ012317 (*K. lactis*), FR839630 (*P. pastoris*) AWR11499 (*D. bruxellensis*), XM\_018355487 (*O. parapolyomorpha*), CR380954 (*C. glabrata*), CP002711 (*A. gossypii*), AP014602 (*K. marxianus*), XM\_001642913 (*K. polysporus*), CP000501 (*S. stipitis*), XM\_002616396 (*C. lusitanae*), and CP028714 (*E. coli*).

Alignment of these sequences was performed using multiple sequence alignment in Clustal Omega (Goujon et al., 2010; Sievers et al., 2011) and the phylogenetic trees were obtained with JalView (version 2.10.4b1) using average distance and percentage identity (Waterhouse et al., 2009).

## Statistics

Statistical analysis was performed using the software IBM SPSS statistics 23 (SPSS inc. Chicago). For transcriptome data, fluorescence data and batch culture data analysis of variance (ANOVA) with Dunnett post-hoc test was performed to test if the results for *S. kudriavzevii* and *S. eubayanus* were statistically different from *S. cerevisiae*.

## RESULTS

### Genetic Makeup of the Glycolytic and Fermentative Pathways in *S. cerevisiae* and Its Close Relatives *S. kudriavzevii* and *S. eubayanus*

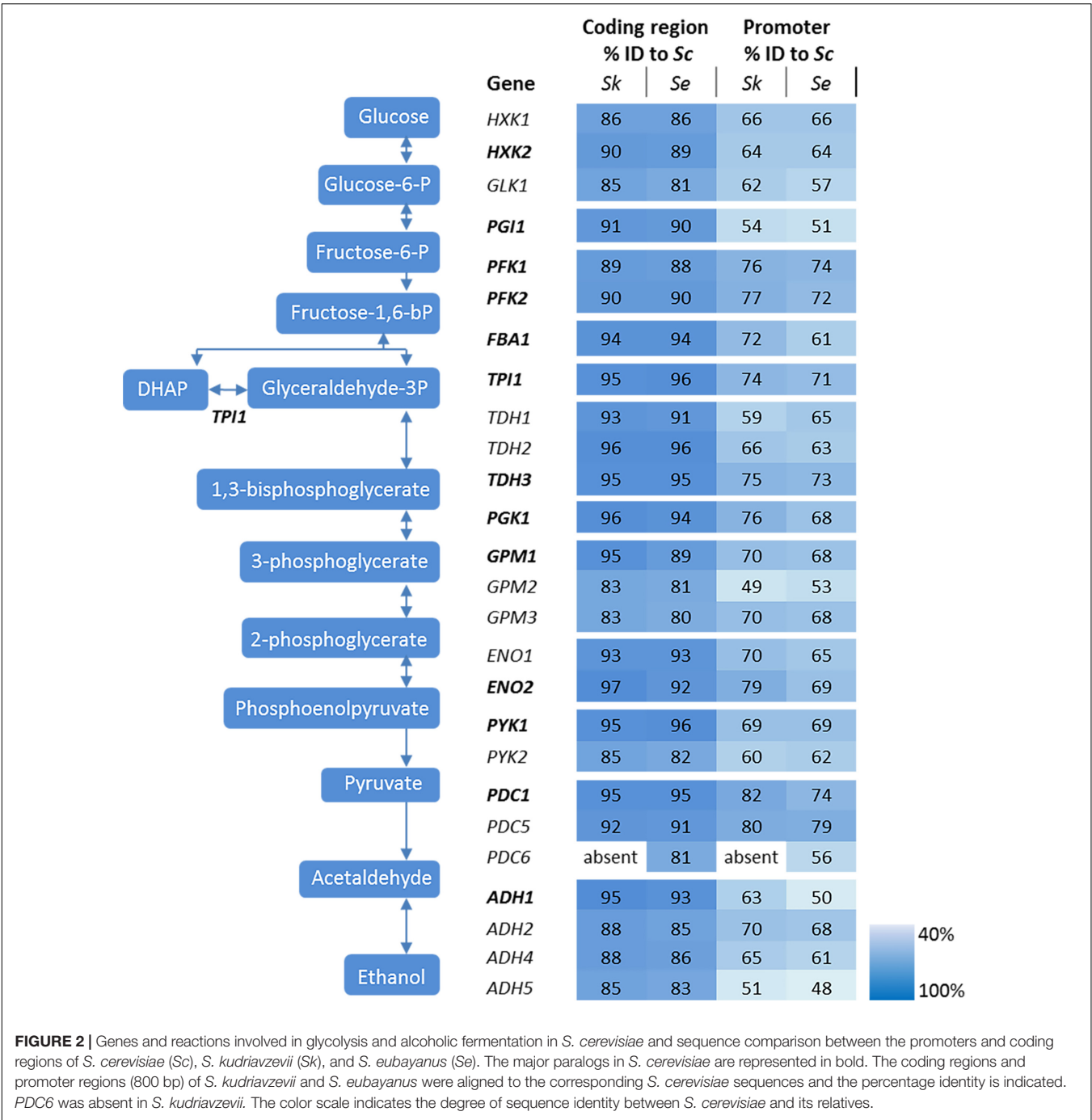
The genetic makeup of pathways involved in central carbon metabolism in *S. cerevisiae* has already been well characterized, and more particularly for glycolysis and alcoholic fermentation. The ten reactions of the glycolytic pathway and the two reactions of ethanolic fermentation in *S. cerevisiae* are catalyzed by a set of 26 enzymes encoded by 26 genes (Figure 2). High quality sequences are already available for *S. cerevisiae* and *S. eubayanus*

(Baker et al., 2015; Salazar et al., 2017). To explore these pathways in *S. kudriavzevii* the strain *S. kudriavzevii* CR85 was sequenced using both Illumina and Oxford Nanopore technologies (see Materials and Methods section and **Supplementary Table S5**). *S. cerevisiae*'s high genetic redundancy and the locations of the genes were fully mirrored in *S. kudriavzevii* and *S. eubayanus* genomes (Figure 1). The only exception was the absence of *PDC6* in *S. kudriavzevii*. While a *ScPDC6* ortholog with 81% identity was identified in *S. eubayanus*, no ortholog could be found in *S. kudriavzevii*. For all other glycolytic genes from *S. cerevisiae*, genes with 80–97% homology of the coding regions were found in *S. kudriavzevii* and *S. eubayanus* (Figure 2). Overall, genes from *S. eubayanus* were slightly more distant from their *S. cerevisiae* orthologs than genes from *S. kudriavzevii*, which is in line with earlier reports (Dujon, 2010; Shen et al., 2016; Figure 2).

In addition to the coding regions, the promoter regions were compared. Since the exact length of most promoter regions is not clearly defined, the 800 bp upstream of the coding regions were considered as promoters. Promoter sequences were substantially less conserved than the coding sequences, ranging from 43 to 78% identity when comparing *S. kudriavzevii* and *S. eubayanus* to *S. cerevisiae* promoters (Figure 2). Remarkably, some regions covering up to 45 bp were strictly conserved among the three species, whereas other parts of the promoter sequences hardly shared homology (see example of *PGK1p* on **Supplementary Figure S1**). As promoter regions are poorly defined, promoters shorter than 800 bp might be fully functional. Alignment with shorter regions might therefore increase the degree of homology between promoters. Alignments using 500 bp upstream the coding region only slightly increased the alignment percentages (up to 7%), mostly as a consequence of the enrichment for conserved transcription factor binding sites located between 100 and 500 bp upstream of the ORF (Harbison et al., 2004). Notably, orthologs with a relatively high or low degree of conservation between *S. cerevisiae* and *S. kudriavzevii* also displayed a similar pattern when comparing *S. eubayanus* to *S. cerevisiae*. For example, the *SkGPM2* and *SeGPM2* promoters both have a relatively low homology (49 and 53%) to the *ScGPM2* promoter, whereas the *SkPFK1* and *SePFK1* promoter have both a high degree (76 and 74%) of similarity to *ScPFK1*. Interestingly, the genes and promoters displaying a relatively low degree of homology between *S. cerevisiae* and its relatives, are homologs considered as minor in *S. cerevisiae* (for example *GPM2* and *PYK2*) (Figure 2). Blast searches did not identify additional glycolytic orthologs present in *S. eubayanus* or *S. kudriavzevii* but absent in *S. cerevisiae*.

The activity of a promoter strongly depends on the presence of regulatory sequences as the TATA box and other specific transcription factor binding sites. In *S. cerevisiae*, the most important glycolytic transcription factor is Gcr1, which has been experimentally shown to bind to most glycolytic promoters and to activate the expression of the corresponding genes as summarized before (Chambers et al., 1995). Gcr1 binding sites are only active when located next to DNA consensus sequences bound by Rap1 (Drazinic et al., 1996), a more pleiotropic transcription factor involved in the transcriptional regulation of a wide variety of genes including many glycolytic genes (Chambers

<sup>3</sup><https://www.ncbi.nlm.nih.gov/bioproject/>

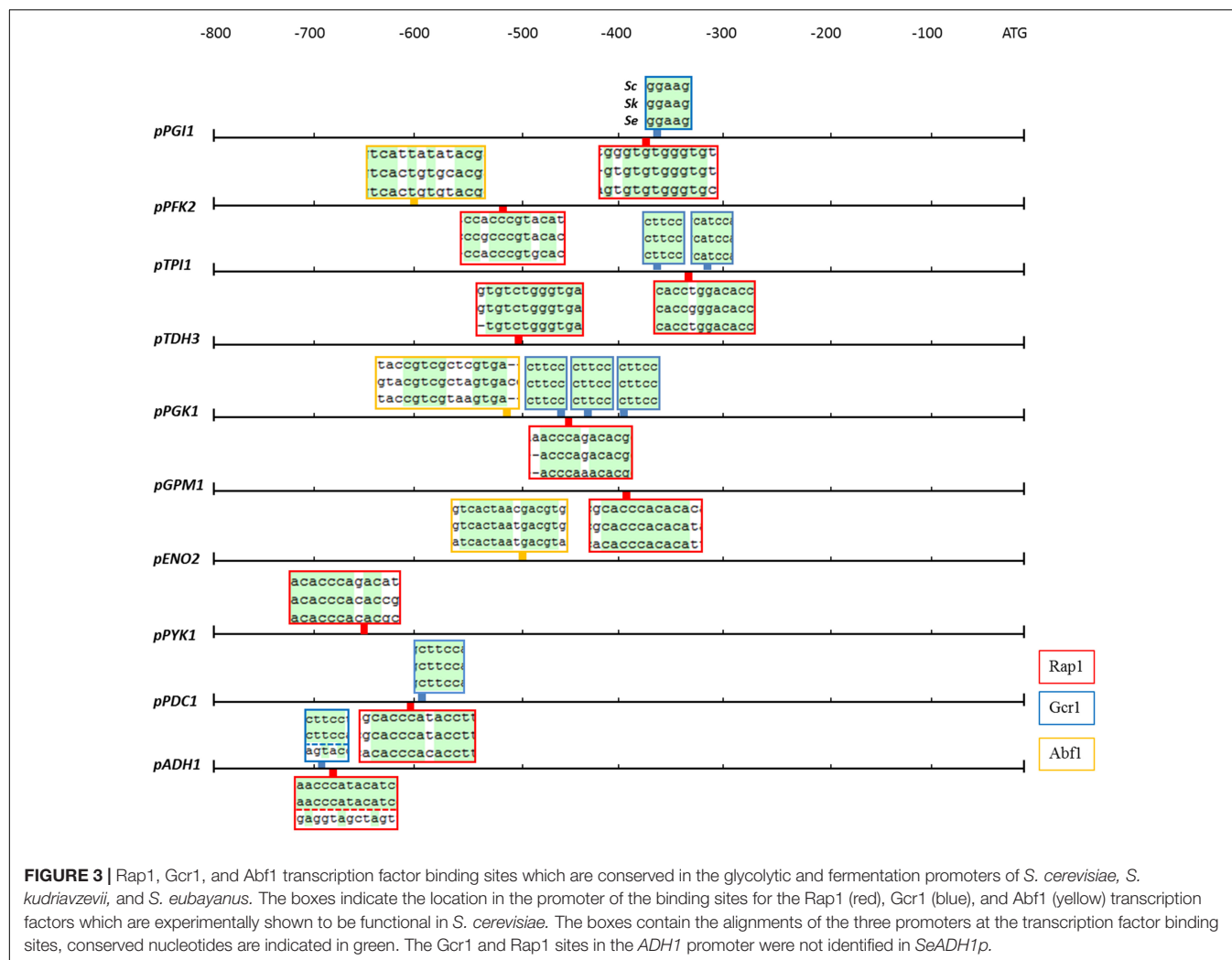


et al., 1995). Another multifunctional transcription factor is Abf1 which binds to several glycolytic promoters (Chambers et al., 1995). With a single exception, all binding sites for Rap1, Gcr1, and Abf1 which were experimentally proven to be active in *S. cerevisiae*, were conserved in *S. kudriavzevii* and *S. eubayanus* promoter regions (Figure 3). The exception was the *SeADH1* promoter in which the Rap1 and Gcr1 site which are conserved between *S. cerevisiae* and *S. kudriavzevii* could not be identified. Together with the presence and high protein similarity of the *SeRap1* (82%), *SkRap1* (86%), *SeGcr1* (85%), and *SkGcr1*

(85%), proteins with *S. cerevisiae*, these results suggested that the regulation of the glycolytic genes might be similar in the three species.

### Expression of the Glycolytic Genes During Aerobic Batch Cultivation

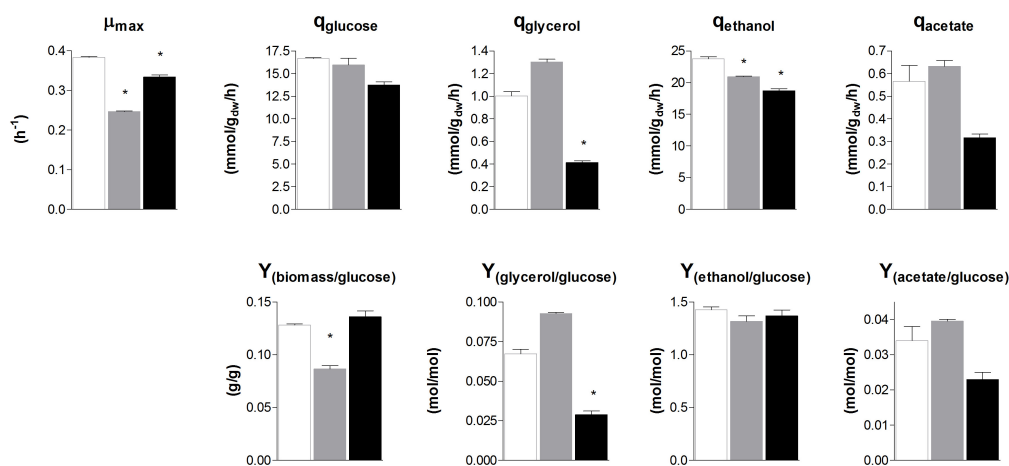
To evaluate the similarity in glycolytic and fermentative gene expression, the transcriptome of *S. cerevisiae*, *S. kudriavzevii*, and *S. eubayanus* was compared. *S. kudriavzevii* and *S. eubayanus* are



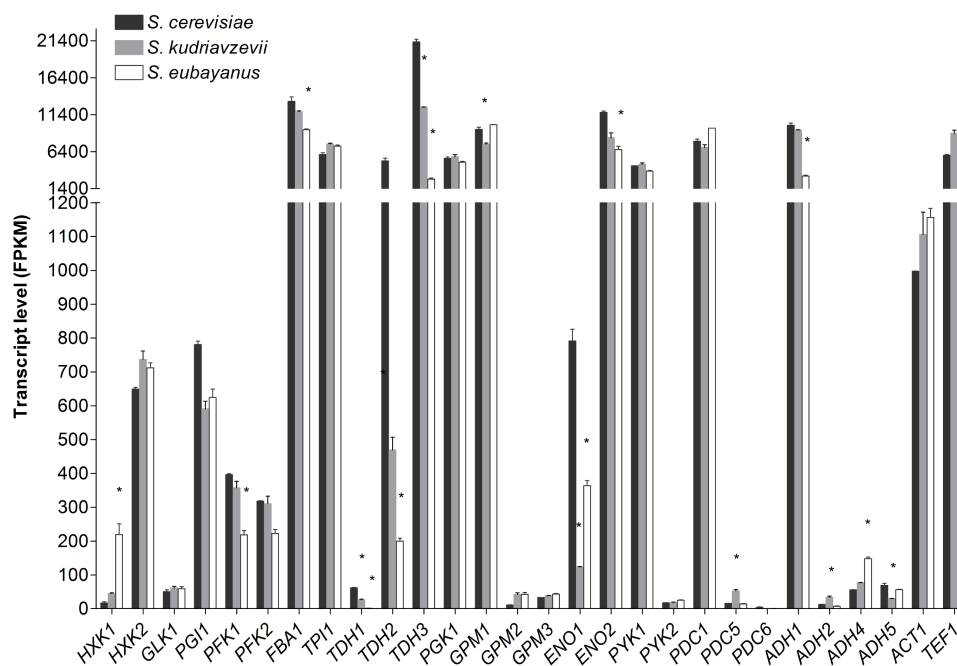
both wild isolates and both diploid (Lopes et al., 2010; Libkind et al., 2011). While many studies report the transcriptome of haploid *S. cerevisiae* strains, transcriptome data for diploid *S. cerevisiae* are scarce (Galitski et al., 1999; Li et al., 2010). To obtain comparable transcriptome datasets for the three species, the diploid CEN.PK122 strain was used. The three diploid strains were grown in aerobic batch cultures in bioreactor using minimal chemically defined medium with glucose as sole carbon source. To ensure optimal growth conditions *S. cerevisiae* was cultivated at 30°C, while its cold-tolerant relatives that have lower temperature optima were cultivated at 25°C (Arroyo-López et al., 2009; Hebly et al., 2015). Under these conditions the maximum specific growth rate of *S. cerevisiae*, *S. kudriavzevii*, and *S. eubayanus* was 0.38 h<sup>-1</sup>, 0.25 h<sup>-1</sup>, and 0.33 h<sup>-1</sup> respectively (Figure 4). Ethanol yields were similar for the three strains, but the biomass yield of *S. kudriavzevii* was significantly lower than that of its two relatives (Figure 4), which might reflect the higher relative cost of maintenance requirements at slow growth rates (Pirt, 1982). For *S. eubayanus* we observed a lower glycerol yield as compared to its relatives, which was

previously not observed under anaerobic conditions (Hebly et al., 2015).

Transcriptome analysis of *S. cerevisiae*, *S. kudriavzevii*, and *S. eubayanus* during mid-exponential growth phase revealed a remarkable similarity between the three species (Figure 5), despite differences in culture temperature. Furthermore, the major or minor classification of paralogous genes was fully conserved between the three species (Figure 5). From the genes considered as major paralogs the *SePFK1*, *SeFBA1*, *SkTDH3*, *SeTDH3*, *SkGPM1*, *SeENO2*, and *SeADH1* genes displayed significantly lower expression levels as compared to *S. cerevisiae*, although only for *SeTDH3* and *SeADH1* the difference with *S. cerevisiae* was larger than 2-fold (8 and 3-fold, respectively). For the minor paralogs slightly more variability was observed. Interestingly *SeHXX1* expression was 13-fold higher than its *S. cerevisiae* ortholog. All three *TDH* genes displayed a significantly lower expression in *S. kudriavzevii* and *S. eubayanus* as compared to *S. cerevisiae*. Likewise, for *ENO1* a lower expression was observed for *SeENO1* and even lower for *SkENO1* as compared to *ScENO1*. Finally, compared to *S. cerevisiae* a ca. 3-fold higher expression was observed for *SkPDC5* and *SeADH4*.



**FIGURE 4 |** Biomass specific rates and yields of *S. cerevisiae*, *S. kudriavzevii*, and *S. eubayanus* batch cultivations in bioreactor. The strains were grown aerobically in synthetic medium supplemented with 20 g L<sup>-1</sup> glucose. *S. cerevisiae* CEN.PK122 (white) was grown at 30°C, and *S. kudriavzevii* CR85 (gray) and *S. eubayanus* CBS 12357 (black) at 25°C. Asterisks indicate significant difference from *S. cerevisiae* (One-Way ANOVA, Dunnett *post hoc* test,  $P < 0.01$ ).



**FIGURE 5 |** Transcript levels of the diploid strains *S. cerevisiae* (black), *S. kudriavzevii* (gray), and *S. eubayanus* (white) from two biological replicates during mid-exponential growth in aerobic batch fermentations on glucose. Asterisks indicate significant difference from *S. cerevisiae* per gene (One-Way ANOVA, Dunnett *post hoc* test,  $P < 0.01$ ).

## Optimization of Microtiter Plate Assays to Monitor Promoter Strength via Fluorescent Reporters

To explore the transferability of promoters within the *Saccharomyces* genus, the promoters of the major glycolytic and fermentative genes (indicated in bold in **Figure 2**) of *S. kudriavzevii* and *S. eubayanus* were functionally characterized in *S. cerevisiae*. A library of fluorescent reporter strains in which

*mRuby2* expression was driven by heterologous promoters and, for comparison, by *S. cerevisiae* promoters, was constructed. To avoid bias due to gene copy number, the constructs were integrated in *S. cerevisiae* genome, at the *URA3* locus. The strains were cultured in 96-well plates, sealed with a transparent foil to prevent evaporation. Simultaneous monitoring of optical density and fluorescence revealed a premature saturation of the fluorescence signal as compared to biomass formation (**Supplementary Figures S2A,B**). Fluorescent proteins have a



strict requirement for molecular oxygen for the synthesis of their chromophores (Tsien, 1998). The poor oxygenation of the cultures in sealed plates combined with the competition for oxygen between cellular respiration, anabolic reactions and mRuby2 maturation could explain the early saturation of the fluorescence signal. Unfortunately, this effect is rarely reported in literature and could be easily overlooked when fluorescence is measured at only one or few time points. Plate readers are widely used as method to characterize promoters with fluorescence reporters (Davis et al., 2010; Zeevi et al., 2011; Keren et al., 2013; Lee et al., 2015) however, information provided in Materials and Methods Sections are often scarce or incomplete, which makes reproduction of data by other groups difficult. To increase oxygen transfer while preventing evaporation, a small aperture was created in each well by puncturing the seal with a needle. The presence of an aperture had a strong impact on the fluorescence intensity of the cultures, enabling to monitor the cultures for a prolonged period of time (**Supplementary Figure S2B**). Also during growth with ethanol as sole carbon source, for which oxygen requirement is substantially increased, no premature saturation of fluorescence was observed (**Supplementary Figures S2C,D**). The location of the aperture in the well did not affect the fluorescence intensity (data not shown). To further evaluate the reliability of the fluorescence signal measured by the plate reader as well as the cell-to-cell heterogeneity of the fluorescence signal, measurements were also performed by flow cytometry. Comparing these data with the plate reader data revealed a very strong correlation of the fluorescence measured with these two techniques ( $R^2 = 0.96$ , **Supplementary Figure S3**).

## Transferability and Context-Dependency of Glycolytic and Fermentative Promoters Within the *Saccharomyces* Genus

The strain library grown in SMG at 30°C not only revealed that the *S. kudriavzevii* and *S. eubayanus* promoters could drive gene expression in *S. cerevisiae*, but also that their strength was remarkably similar to the strength of their *S. cerevisiae* orthologs (**Figure 6**). Additionally, two reporter strains expressing mRuby2 from the constitutive *S. cerevisiae* *TEF1* and *ACT1* promoters were constructed and cultivated on all plates experiments. The activity of these two promoters was remarkably reproducible between independent culture replicates (**Supplementary Figure S4**).

While, due to high data reproducibility, expression driven by *S. kudriavzevii* or *S. eubayanus* promoters was in most cases considered statistically different from the expression led by their *S. cerevisiae* orthologs (student *t*-test,  $P < 0.01$ ), differences in expression larger than 1.5-fold were rarely observed. Expression of *ENO2p* and *PDC1p* of *S. kudriavzevii* and *S. eubayanus* was lower than for their *S. cerevisiae* counterparts, while *SkGPM1p*, *SeGPM1p*, and *SePYK1p* led to clearly higher expression levels than their *S. cerevisiae* homologs (**Figure 6**). These differences were not reflected in the transcript data (**Figure 5**). Conversely, the differential expression of *PFK1* and *TDH3* revealed by the

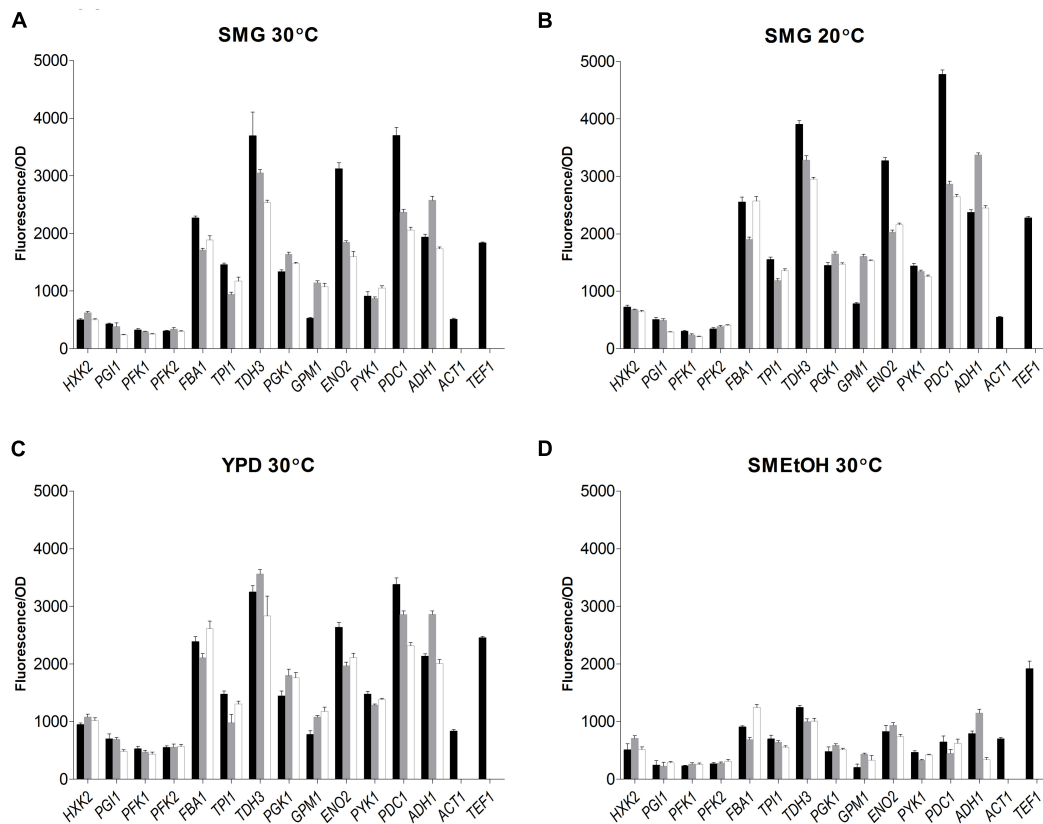
RNAseq analysis was also found in the promoter transplantation study at SMG 30°C. Overall similarities and differences between the three species in transcript levels were mirrored by promoter activity.

To test the condition dependency of promoter activity, strains were tested under several culture conditions. YPD was used as rich medium, and ethanol was used as gluconeogenic carbon source (SMEtOH). Since *S. kudriavzevii* and *S. eubayanus* have a lower optimum growth temperature and hexokinase from *S. kudriavzevii* has been proposed to have a lower temperature optimum as compared to *S. cerevisiae* (Gonçalves et al., 2011), the strains were also grown in SMG at 20°C. When grown in YPD and SMG at 20°C, all strains showed highly similar promoter activities as compared to cultures in SMG at 30°C even though the growth rates were different (SMG 30°C:  $0.34 \text{ h}^{-1}$ , SMG 20°C:  $0.15 \text{ h}^{-1}$ , YPD 30°C:  $0.36 \text{ h}^{-1}$ ). However, during growth on ethanol ( $0.13 \text{ h}^{-1}$ ) promoter activity of the three species dropped tremendously as compared to glucose-grown cultures, in stark contrast with the fluorescence of the reference strains (*TEF1p* and *ACT1p*) that remained remarkably constant for all cultivation conditions. Nevertheless, also on SMEtOH *S. kudriavzevii* and *S. eubayanus* promoters showed expression levels very similar to their *S. cerevisiae* orthologs.

## DISCUSSION

In this study we showed that the genetic makeup of the glycolytic and fermentative pathways is highly conserved among *S. cerevisiae*, *S. kudriavzevii*, and *S. eubayanus*. For 11 out of 12 reactions, the exact same number of paralogs was found in the three species, reflecting that species divergence took place after whole genome and post-whole genome duplications. The only exception was the absence of the minor paralog *PDC6* in the *S. kudriavzevii* CR85 genome. In agreement with this observation, the presence of a pseudogene in *S. kudriavzevii* strains IFO1802 and ZP591 consisting of only about 15% of the full *PDC6* gene length has been reported (Scannell et al., 2011). At the transcript level a strong conservation was also observed, suggesting that the classification between major and minor paralogs, confirmed in *S. cerevisiae* by mutant studies, could be extended to *S. kudriavzevii* and *S. eubayanus*. The slightly lower degree of conservation of minor paralogs (e.g., *GPM2*, *PYK2*, *ENO1*, *TDH1*, *TDH2*, *PDC6*, *ADH2*, *ADH4*, *ADH5*) is in line with the previously reported accelerated evolution of the *PYK2* and *ADH5* as compared to their *PYK1* and *ADH1* paralogs (Kellis et al., 2004).

The glycolytic pathway is known to be highly conserved compared to most other pathways (Fothergill-Gilmore and Michels, 1993; Webster, 2003). Recently it was shown that glycolytic coding regions from *E. coli* could replace the corresponding yeast genes (Kachroo et al., 2017). For promoter regions the conservation is in general lower as compared to coding regions, but a stronger conservation was found for the glycolytic promoters in the *Saccharomyces* genus than for other promoter regions (Kuang et al., 2017). Combined with the remarkable conservation of binding sites



**FIGURE 6 |** Promoter activity of the major glycolytic promoters from *S. cerevisiae* (black), *S. kudriavzevii* (gray), and *S. eubayanus* (white) expressing *mRuby2* in *S. cerevisiae*. During exponential growth in SMG (A), SMG 20°C (B), YPD (C), and SMETOH (D) fluorescence and optical density were measured every 20 min and promoter activity was calculated as the slope of the linear regression between optical density and fluorescence. Two reference strains expressing *mRuby2* from the *ScTEF1* and *ScACT1* promoters were taken along in every plate. Error bars represent the standard deviation of the mean of six biological replicates.

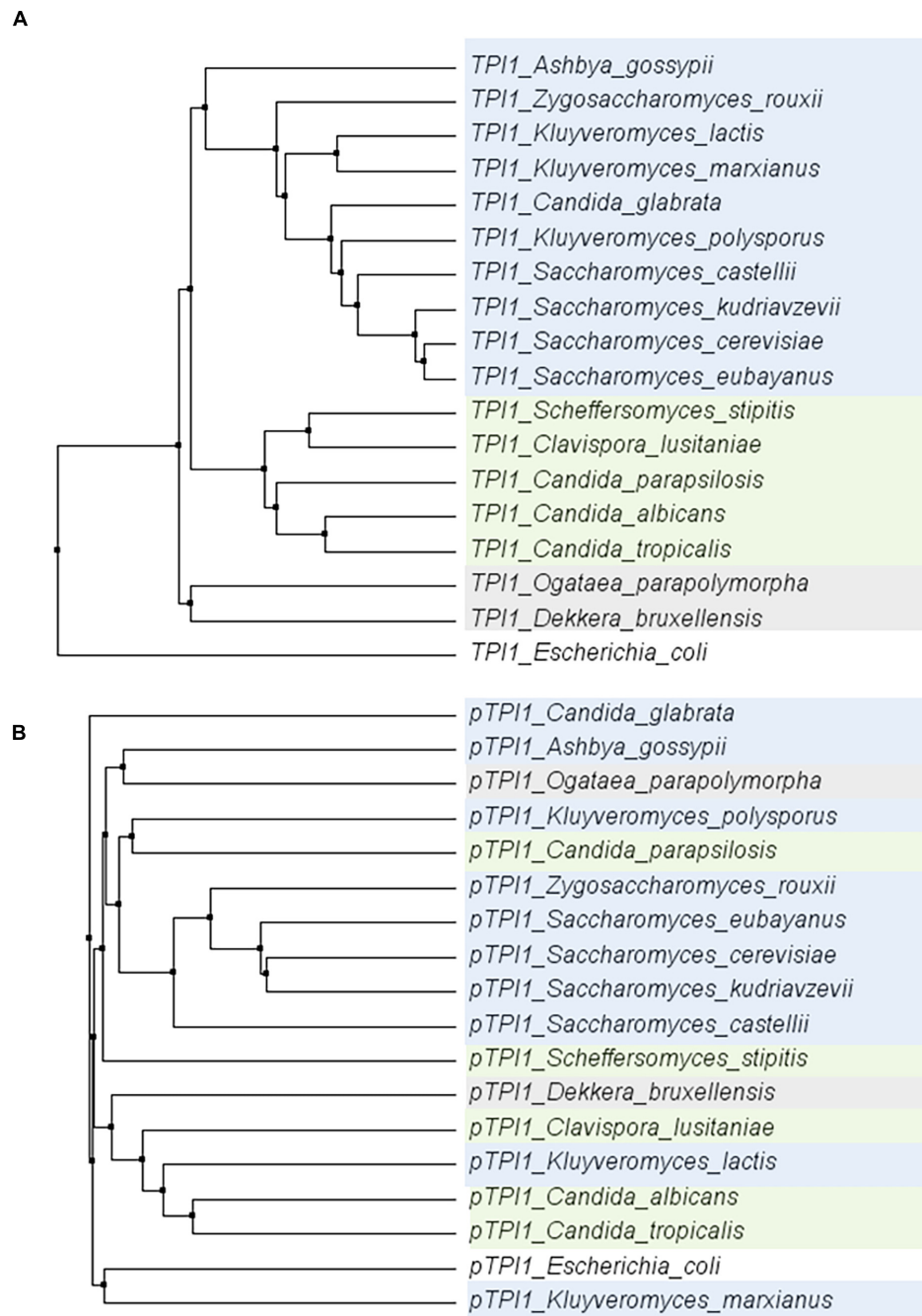
for major transcriptional regulators (i.e., Rap1, Gcr1, and Abf1), these observations suggested a very similar transcriptional regulation of glycolytic and fermentative genes across the three species. Accordingly, transcriptome data showed a remarkable conservation in expression for the majority of glycolytic and fermentative genes in their native context. It is noteworthy that transcript levels of glycolytic and fermentative genes of these three diploid species were highly similar to the transcript levels of the haploid *S. cerevisiae* CEN.PK113-7D cultivated in the same condition as the *S. cerevisiae* diploid (Solis-Escalante et al., 2015). The similarity in gene expression and the conservation of the main transcription factor binding sites in the three species suggested the possibility to introduce the promoters in *S. cerevisiae*, expecting similar regulation.

Until now a limited number of examples of heterologous glycolytic promoters driving gene expression in *S. cerevisiae* is available. Recently, it was shown that *S. kudriavzevii* glycolytic and fermentative promoters could drive gene expression in *S. cerevisiae* (Kuijpers et al., 2016). More recently, it was shown that the *ADH2* promoter of several *Saccharomyces* species could drive gene expression in *S. cerevisiae* (Harvey et al., 2018). Further, the glycolytic genes *PFK1*, *PFK2* and *PYK1* of the more distantly related yeast *Hanseniaspora*

*uvarum*, expressed from their native promoters were shown to complement their *S. cerevisiae* orthologs (Langenberg et al., 2017).

To explore the conservation of glycolytic genes in a broader context, the sequence of the *TPI1* gene was compared across a set of 18 species within the *Saccharomycotina* subphylum (Dujon, 2010). Within this subphylum, the coding region of *TPI1* was highly conserved (ranging from 64,7 to 96,3% identity to *S. cerevisiae*), while the promoters generally displayed a much weaker similarity (ranging from 28,5%–71,1% identity to *S. cerevisiae*) (Figure 7). These observations are in line with studies reporting the loss of the gene encoding the Gcr1 transcription factor and the gain of new function by Rap1 in the CTG clade yeast *Candida albicans* (Askew et al., 2009; Lavoie et al., 2010; Weirauch and Hughes, 2010). Indeed using the MEME suite motif discovery tool (Bailey et al., 2009) gave only hits for Rap1 and Gcr1 motifs in the *Saccharomyces* genus.

The present study shows the ability of all the major glycolytic promoters of *S. kudriavzevii* and *S. eubayanus* to drive gene expression in *S. cerevisiae* with similar strength and condition-dependency. Since many hybrids occur between *S. cerevisiae*  $\times$  *S. eubayanus* and *S. cerevisiae*  $\times$  *S. kudriavzevii*, it is



**FIGURE 7 |** Phylogenetic trees of the alignments of the *TPI1* genes (ORF) **(A)** and promoters (800 bp) **(B)** for a set of yeast species from the *Saccharomycotina* phylum. The coding regions are strongly conserved, whereas there is hardly any conservation among promoter regions. *E. coli* was used as an outgroup. Color indicates groups as defined in Dujon (2010).

not surprising that promoters are functional in *S. cerevisiae*. However, the similarity we found in promoter activities for most promoters transplanted to *S. cerevisiae* under different conditions is remarkable and indicates a strong conservation of the glycolytic regulatory mechanisms for *S. cerevisiae*, *S. kudriavzevii*, and *S. eubayanus* (Supplementary Figure S6). In general, these data do not correlate very well with the

transcript data (Supplementary Figure S7). This can most likely be explained by the relatively low dynamic measurement range of the plate reader compared to RNAseq, differences in cultivation conditions, length of promoters, choice of site for genomic integration (Bai Flagfeldt et al., 2009) or differences in regulatory sequences in the promoters. During growth on ethanol a strong decrease in promoter activity was observed.

This is in agreement with the previously reported drop in enzymatic activity of glycolysis during growth on ethanol (Peter Smits et al., 2000; Van Hoek et al., 2000).

*S. cerevisiae*'s proficiency in assembling and functionally expressing large (heterologous) pathways has propelled this yeast as preferred host for the production of complex molecules such as isoprenoids or opioids (Paddon et al., 2013; Galanie et al., 2015). However, the successful expression of these pathways depends on the availability of suitable promoters. While *S. cerevisiae* has one of the most furnished molecular toolbox, the number of constitutive and well characterized promoters remains limited. Since *S. cerevisiae*'s extremely efficient homologous recombination renders strains with repeated usage of promoter sequences genetically unstable (Manivasakam et al., 1995), this shortage of promoters presents a hurdle for extensive strain construction programs. While a lot of effort is invested in the design of synthetic promoters and transcription amplifiers (Redden and Alper, 2015; Rantasalo et al., 2016; Machens et al., 2017; Naseri et al., 2017), using slightly distant but functional orthologous promoters presents an attractive alternative (Naesby et al., 2009; Harvey et al., 2018). Usage of especially the *S. eubayanus* promoters, which are slightly more distant from *S. cerevisiae* than the *S. kudriavzevii* promoters, would reduce the length of the sequences being 100% identical to the native *S. cerevisiae* promoters. The minimum length which was found to be needed for efficient homologous recombination in *S. cerevisiae* was 30 bp with an optimal efficiency at a length of 60 bp or more (Manivasakam et al., 1995; Hua et al., 1997). In the *S. eubayanus* promoters, with one exception for the *PFK2* promoter, the longest sequence being identical to *S. cerevisiae* was found to be 34 bp. Usage of the *S. eubayanus* promoters would therefore substantially decrease the risk of instability and undesired recombination events during strain construction programs.

## CONCLUSION

This study brings new insight in the genetic makeup and expression of glycolytic and fermentative genes in *S. eubayanus* and *S. kudriavzevii*. It also expands the molecular toolbox for *S. cerevisiae*, but also for its two relatives, with a set of strong,

constitutive promoters. Furthermore, combining Illumina and Oxford Nanopore technologies, the present study offers a high quality sequence for *S. kudriavzevii* CR85, available from NCBI (PRJNA480800). Finally, the full set of transcript levels for the three diploid strains grown in tightly controlled conditions is available via GEO (See Materials and Methods section) and can be mined to compare species-specific regulation of gene expression beyond the glycolytic and fermentative pathways.

## AUTHOR CONTRIBUTIONS

FB, J-MD, and PD-L designed the research. FB and SD performed the experiments. MvdB and TG performed the transcriptome analysis. MvdB performed the sequence analysis. FB, SD, J-MD, and PD-L prepared the manuscript. All authors read and approved the final manuscript.

## FUNDING

This project was funded by the AdLibYeast European Research Council (ERC) consolidator 648141 grant awarded to PD-L.

## ACKNOWLEDGMENTS

We thank Rik Brouwer for his contribution to strain construction, Mark Bisschops for advice and help with fermentations and data analysis, Marijke Luttkik for technical support, and advice for the TECAN plate reader and flow cytometry analysis and Pilar de la Torre for performing the whole genome sequencing. Furthermore, we thank Jack Pronk for his advice and Raúl Ortiz Merino for constructive comments on the manuscript. We thank Eladio Barrio Esparducer for kindly providing *S. kudriavzevii* CR85.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2018.00504/full#supplementary-material>

## REFERENCES

- Anders, S., Pyl, P. T., and Huber, W. (2015). HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* 31, 166–169. doi: 10.1093/bioinformatics/btu638
- Arroyo-López, F. N., Orlic, S., Querol, A., and Barrio, E. (2009). Effects of temperature, pH and sugar concentration on the growth parameters of *Saccharomyces cerevisiae*, *S. kudriavzevii* and their interspecific hybrid. *Int. J. Food Microbiol.* 131, 120–127. doi: 10.1016/j.jfoodmicro.2009.01.035
- Arroyo-López, F. N., Pérez-Torrado, R., Querol, A., and Barrio, E. (2010). Modulation of the glycerol and ethanol syntheses in the yeast *Saccharomyces kudriavzevii* differs from that exhibited by *Saccharomyces cerevisiae* and their hybrid. *Food Microbiol.* 27, 628–637. doi: 10.1016/j.fm.2010.02.001
- Arvanitidis, A., and Heinisch, J. J. (1994). Studies on the function of yeast phosphofructokinase subunits by *in vitro* mutagenesis. *J. Biol. Chem.* 269, 8911–8918.
- Askew, C., Sellam, A., Epp, E., Hogue, H., Mullick, A., Nantel, A., et al. (2009). Transcriptional regulation of carbohydrate metabolism in the human pathogen *Candida albicans*. *PLoS Path.* 5:e1000612. doi: 10.1371/journal.ppat.1000612
- Bai Flagfeldt, D., Siewers, V., Huang, L., and Nielsen, J. (2009). Characterization of chromosomal integration sites for heterologous gene expression in *Saccharomyces cerevisiae*. *Yeast* 26, 545–551. doi: 10.1002/yea.1705
- Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., et al. (2009). MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* 37, W202–W208. doi: 10.1093/nar/gkp335
- Baker, E., Wang, B., Bellora, N., Peris, D., Hulfachor, A. B., Koshalek, J. A., et al. (2015). The genome sequence of *Saccharomyces eubayanus* and



- the domestication of lager-brewing yeasts. *Mol. Biol. Evol.* 32, 2818–2831. doi: 10.1093/molbev/msv168
- Barnett, J. A. (2003). A history of research on yeasts 5: the fermentation pathway. *Yeast* 20, 509–543. doi: 10.1002/yea.986
- Barnett, J. A., and Entian, K. D. (2005). A history of research on yeasts 9: regulation of sugar metabolism. *Yeast* 22, 835–894. doi: 10.1002/yea.1249
- Belloch, C., Orlic, S., Barrio, E., and Querol, A. (2008). Fermentative stress adaptation of hybrids within the *Saccharomyces sensu stricto* complex. *Int. J. Food Microbiol.* 122, 188–195. doi: 10.1016/j.ijfoodmicro.2007.11.083
- Bertani, G. (1951). Studies on lysogeny I: the mode of phage liberation by lysogenic *Escherichia coli*. *J. Bacteriol.* 62, 293–300.
- Bertani, G. (2004). Lysogeny at mid-twentieth century: P1, P2, and other experimental systems. *J. Bacteriol.* 186, 595–600. doi: 10.1128/JB.186.3.595-600.2004
- Boer, V. M., De Winde, J. H., Pronk, J. T., and Piper, M. D. (2003). The genome-wide transcriptional responses of *Saccharomyces cerevisiae* grown on glucose in aerobic chemostat cultures limited for carbon, nitrogen, phosphorus, or sulfur. *J. Biol. Chem.* 278, 3265–3274. doi: 10.1074/jbc.M209759200
- Bond, U. (2009). The genomes of lager yeasts. *Adv. Appl. Microbiol.* 69, 159–182. doi: 10.1016/S0065-2164(09)69006-7
- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., et al. (2009). BLAST+: architecture and applications. *BMC Bioinformatics* 10:421. doi: 10.1186/1471-2105-10-421
- Carroll, K. M., Simpson, D. M., Eysers, C. E., Knight, C. G., Brownridge, P., Dunn, W. B., et al. (2011). Absolute quantification of the glycolytic pathway in yeast: deployment of a complete QconCAT approach. *Mol. Cell. Proteomics* 10:M111.007633. doi: 10.1074/mcp.M111.007633
- Chambers, A., Packham, E. A., and Graham, I. R. (1995). Control of glycolytic gene expression in the budding yeast (*Saccharomyces cerevisiae*). *Curr. Genet.* 29, 1–9. doi: 10.1007/BF00313187
- Ciriacy, M. (1979). Isolation and characterization of further cis- and trans-acting regulatory elements involved in the synthesis of glucose-repressible alcohol dehydrogenase (ADHI) in *Saccharomyces cerevisiae*. *Mol. Gen. Genet.* 176, 427–431. doi: 10.1007/BF00333107
- Conant, G. C., and Wolfe, K. H. (2007). Increased glycolytic flux as an outcome of whole – genome duplication in yeast. *Mol. Syst. Biol.* 3:129. doi: 10.1038/msb4100170
- Conant, G. C., and Wolfe, K. H. (2008). Turning a hobby into a job: how duplicated genes find new functions. *Nat. Rev. Genet.* 9, 938–950. doi: 10.1038/nrg2482
- Davis, J. H., Rubin, A. J., and Sauer, R. T. (2010). Design, construction and characterization of a set of insulated bacterial promoters. *Nucleic Acids Res.* 39, 1131–1141. doi: 10.1093/nar/gkq810
- De Deken, R. (1966). The Crabtree effect: a regulatory system in yeast. *Microbiology* 44, 149–156. doi: 10.1099/00221287-44-2-149
- Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., et al. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21. doi: 10.1093/bioinformatics/bts635
- Drazinic, C. M., Smerage, J. B., López, M. C., and Baker, H. V. (1996). Activation mechanism of the multifunctional transcription factor repressor-activator protein 1 (Rap1p). *Mol. Cell. Biol.* 16, 3187–3196. doi: 10.1128/MCB.16.6.3187
- Dujon, B. (2010). Yeast evolutionary genomics. *Nat. Rev. Genet.* 11, 512–524. doi: 10.1038/nrg2811
- Entian, K.-D., and Kötter, P. (2007). 25 Yeast genetic strain and plasmid collections. *Methods Microbiol.* 36, 629–666. doi: 10.1016/S0580-9517(06)36025-4
- Fauchon, M., Lagniel, G., Aude, J.-C., Lombardia, L., Soularue, P., Petat, C., et al. (2002). Sulfur sparing in the yeast proteome in response to sulfur demand. *Mol. Cell* 9, 713–723. doi: 10.1016/S1097-2765(02)00500-2
- Fothergill-Gilmore, L. A., and Michels, P. A. (1993). Evolution of glycolysis. *Prog. Biophys. Mol. Biol.* 59, 105–235. doi: 10.1016/0079-6107(93)90001-Z
- Fraenkel, D. G. (2003). The top genes: on the distance from transcript to function in yeast glycolysis. *Curr. Opin. Microbiol.* 6, 198–201. doi: 10.1016/S1369-5274(03)00023-7
- Galanie, S., Thodey, K., Trenchard, I. J., Interrante, M. F., and Smolke, C. D. (2015). Complete biosynthesis of opioids in yeast. *Science* 349, 1095–1100. doi: 10.1126/science.aac9373
- Galitski, T., Saldanha, A. J., Styles, C. A., Lander, E. S., and Fink, G. R. (1999). Ploidy regulation of gene expression. *Science* 285, 251–254. doi: 10.1126/science.285.5425.251
- Gietz, R. D., and Woods, R. A. (2002). Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method. *Methods Enzymol.* 350, 87–96. doi: 10.1016/S0076-6879(02)50957-5
- Gonçalves, P., Valério, E., Correia, C., de Almeida, J. M., and Sampaio, J. P. (2011). Evidence for divergent evolution of growth temperature preference in sympatric *Saccharomyces* species. *PLoS One* 6:e20739. doi: 10.1371/journal.pone.0020739
- González, S. S., Barrio, E., Gafner, J., and Querol, A. (2006). Natural hybrids from *Saccharomyces cerevisiae*, *Saccharomyces bayanus* and *Saccharomyces kudriavzevii* in wine fermentations. *FEMS Yeast Res.* 6, 1221–1234. doi: 10.1111/j.1567-1364.2006.00126.x
- González, S. S., Barrio, E., and Querol, A. (2008). Molecular characterization of new natural hybrids of *Saccharomyces cerevisiae* and *S. kudriavzevii* in brewing. *Appl. Environ. Microbiol.* 74, 2314–2320. doi: 10.1128/AEM.01867-07
- Goujon, M., McWilliam, H., Li, W., Valentin, F., Squizzato, S., Paern, J., et al. (2010). A new bioinformatics analysis tools framework at EMBL-EBI. *Nucleic Acids Res.* 38, W695–W699. doi: 10.1093/nar/gkq313
- Harbison, C. T., Gordon, D. B., Lee, T. I., Rinaldi, N. J., Macisaac, K. D., Danford, T. W., et al. (2004). Transcriptional regulatory code of a eukaryotic genome. *Nature* 431, 99–104. doi: 10.1038/nature02800
- Harvey, C. J., Tang, M., Schlecht, U., Horecka, J., Fischer, C. R., Lin, H.-C., et al. (2018). HEx: a heterologous expression platform for the discovery of fungal natural products. *Sci. Adv.* 4:eaar5459. doi: 10.1126/sciadv.aar5459
- Hebly, M., Brickwedde, A., Bolat, I., Driessen, M. R., de Hulster, E. A., van den Broek, M., et al. (2015). *S. cerevisiae* × *S. eubayanus* interspecific hybrid, the best of both worlds and beyond. *FEMS Yeast Res.* 15:fov005. doi: 10.1093/femsyr/fov005
- Heinisch, J., Vogelsang, K., and Hollenberg, C. P. (1991). Transcriptional control of yeast phosphofructokinase gene expression. *FEBS Lett.* 289, 77–82. doi: 10.1016/0014-5793(91)80912-M
- Hittinger, C. T. (2013). *Saccharomyces* diversity and evolution: a budding model genus. *Trends Genet.* 29, 309–317. doi: 10.1016/j.tig.2013.01.002
- Holt, C., and Yandell, M. (2011). MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics* 12:491. doi: 10.1186/1471-2105-12-491
- Hua, S.-B., Qiu, M., Chan, E., Zhu, L., and Luo, Y. (1997). Minimum length of sequence homology required for *in vivo* cloning by homologous recombination in yeast. *Plasmid* 38, 91–96. doi: 10.1006/plas.1997.1305
- Kachroo, A. H., Laurent, J. M., Akhmetov, A., Szilagi-Jones, M., McWhite, C. D., Zhao, A., et al. (2017). Systematic bacterialization of yeast genes identifies a near-universally swappable pathway. *Elife*. 6:e25093. doi: 10.7554/eLife.25093
- Kellis, M., Birren, B. W., and Lander, E. S. (2004). Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*. *Nature* 428, 617–624. doi: 10.1038/nature02424
- Keren, L., Zackay, O., Lotan-Pompan, M., Barenholz, U., Dekel, E., Sasson, V., et al. (2013). Promoters maintain their relative activity levels under different growth conditions. *Mol. Syst. Biol.* 9:701. doi: 10.1038/msb.2013.59
- Knijnenburg, T. A., Daran, J.-M. G., van den Broek, M. A., Daran-Lapujade, P. A., de Winde, J. H., Pronk, J. T., et al. (2009). Combinatorial effects of environmental parameters on transcriptional regulation in *Saccharomyces cerevisiae*: a quantitative analysis of a compendium of chemostat-based transcriptome data. *BMC Genomics* 10:53. doi: 10.1186/1471-2164-10-53
- Koren, S., Walenz, B. P., Berlin, K., Miller, J. R., Bergman, N. H., and Phillippy, A. M. (2017). Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.* 27, 722–736. doi: 10.1101/gr.215087.116
- Korf, I. (2004). Gene finding in novel genomes. *BMC Bioinformatics* 5:59. doi: 10.1186/1471-2105-5-59
- Kuang, Z., Pinglay, S., Ji, H., and Boeke, J. D. (2017). Msn2/4 regulate expression of glycolytic enzymes and control transition from quiescence to growth. *Elife* 6:e29938. doi: 10.7554/eLife.29938
- Kuefer, L., Sauer, U., and Blank, L. M. (2005). Metabolic functions of duplicate genes in *Saccharomyces cerevisiae*. *Genome Res.* 15, 1421–1430. doi: 10.1101/gr.3992505
- Kuijpers, N. G., Solis-Escalante, D., Luttk, M. A., Bisschops, M. M., Boonekamp, F. J., van den Broek, M., et al. (2016). Pathway swapping: toward modular engineering of essential cellular processes. *Proc. Natl. Acad. Sci. U.S.A.* 113, 15060–15065. doi: 10.1073/pnas.1606701113

- Langenberg, A.-K., Bink, F. J., Wolff, L., Walter, S., von Wallbrunn, C., Grossmann, M., et al. (2017). Glycolytic functions are conserved in the genome of the wine yeast *Hanseniaspora uvarum* and pyruvate kinase limits its capacity for alcoholic fermentation. *Appl. Environ. Microbiol.* 83:e01580-17. doi: 10.1128/AEM.01580-17
- Lavoie, H., Hogues, H., Mallick, J., Sellam, A., Nantel, A., and Whiteway, M. (2010). Evolutionary tinkering with conserved components of a transcriptional regulatory network. *PLoS Biol.* 8:e1000329. doi: 10.1371/journal.pbio.1000329
- Lee, M. E., DeLoache, W. C., Cervantes, B., and Dueber, J. E. (2015). A highly characterized yeast toolkit for modular, multipart assembly. *ACS Synth. Biol.* 4, 975–986. doi: 10.1021/sb500366v
- Li, B.-Z., Cheng, J.-S., Ding, M.-Z., and Yuan, Y.-J. (2010). Transcriptome analysis of differential responses of diploid and haploid yeast to ethanol stress. *J. Biotechnol.* 148, 194–203. doi: 10.1016/j.jbiotec.2010.06.013
- Li, H., and Durbin, R. (2010). Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics* 26, 589–595. doi: 10.1093/bioinformatics/btp698
- Libkind, D., Hittinger, C. T., Valério, E., Gonçalves, C., Dover, J., Johnston, M., et al. (2011). Microbe domestication and the identification of the wild genetic stock of lager-brewing yeast. *Proc. Natl. Acad. Sci. U.S.A.* 108, 14539–14544. doi: 10.1073/pnas.1105430108
- Lopes, C. A., Barrio, E., and Querol, A. (2010). Natural hybrids of *S. cerevisiae* x *S. kudriavzevii* share alleles with European wild populations of *Saccharomyces kudriavzevii*. *FEMS Yeast Res.* 10, 412–421. doi: 10.1111/j.1567-1364.2010.00614.x
- Machens, F., Balazadeh, S., Mueller-Roeber, B., and Messerschmidt, K. (2017). Synthetic promoters and transcription factors for heterologous protein expression in *Saccharomyces cerevisiae*. *Front. Bioeng. Biotechnol.* 5:63. doi: 10.3389/fbioe.2017.00063
- Manivasakam, P., Weber, S. C., McElver, J., and Schiestl, R. H. (1995). Microhomology mediated PCR targeting in *Saccharomyces cerevisiae*. *Nucleic Acids Res.* 23, 2799–2800. doi: 10.1093/nar/23.14.2799
- Masneuf-Pomarède, I., Bely, M., Marullo, P., Lonvaud-Funel, A., and Dubourdieu, D. (2010). Reassessment of phenotypic traits for *Saccharomyces bayanus* var. *uvarum* wine yeast strains. *Int. J. Food Microbiol.* 139, 79–86. doi: 10.1016/j.ijfoodmicro.2010.01.038
- Merico, A., Sulo, P., Piškur, J., and Compagno, C. (2007). Fermentative lifestyle in yeasts belonging to the *Saccharomyces* complex. *FEBS J.* 274, 976–989. doi: 10.1111/j.1742-4658.2007.05645.x
- Naesby, M., Nielsen, S. V., Nielsen, C. A., Green, T., Tange, T. Ø, Simón, E., et al. (2009). Yeast artificial chromosomes employed for random assembly of biosynthetic pathways and production of diverse compounds in *Saccharomyces cerevisiae*. *Microb. Cell Fact.* 8:45. doi: 10.1186/1475-2859-8-45
- Naseeb, S., James, S. A., Alsammar, H., Michaels, C. J., Gini, B., Nueno-Palop, C., et al. (2017). *Saccharomyces jurei* sp. nov., Isolation and genetic identification of a novel yeast species from *Quercus robur*. *Int. J. Syst. Evol. Microbiol.* 67, 2046–2052. doi: 10.1099/ijsem.0.002013
- Naseri, G., Balazadeh, S., Machens, F., Kamranfar, I., Messerschmidt, K., and Mueller-Roeber, B. (2017). Plant-derived transcription factors for orthologous regulation of gene expression in the yeast *Saccharomyces cerevisiae*. *ACS Synth. Biol.* 6, 1742–1756. doi: 10.1021/acssynbio.7b00094
- Nguyen, H. V., and Boekhout, T. (2017). Characterization of *Saccharomyces uvarum* (Beijerinck, 1898) and related hybrids: assessment of molecular markers that predict the parent and hybrid genomes and a proposal to name yeast hybrids. *FEMS Yeast Res.* 17:2. doi: 10.1093/femsyr/fox014
- Nijkamp, J. F., van den Broek, M., Datema, E., de Kok, S., Bosman, L., Luttik, M. A., et al. (2012). De novo sequencing, assembly and analysis of the genome of the laboratory strain *Saccharomyces cerevisiae* CEN. PK113-7D, a model for modern industrial biotechnology. *Microb. Cell Fact.* 11:36. doi: 10.1186/1475-2859-11-36
- Ohno, S. (1970). *Evolution by Gene Duplication*. New York, NY: Springer. doi: 10.1007/978-3-642-86659-3
- Paddon, C. J., Westfall, P. J., Pitera, D. J., Benjamin, K., Fisher, K., McPhee, D., et al. (2013). High-level semi-synthetic production of the potent antimalarial artemisinin. *Nature* 496, 528–532. doi: 10.1038/nature12051
- Peng, B., Williams, T. C., Henry, M., Nielsen, L. K., and Vickers, C. E. (2015). Controlling heterologous gene expression in yeast cell factories on different carbon substrates and across the diauxic shift: a comparison of yeast promoter activities. *Microb. Cell Fact.* 14:91. doi: 10.1186/s12934-015-0278-5
- Peris, D., Lopes, C. A., Belloch, C., Querol, A., and Barrio, E. (2012). Comparative genomics among *Saccharomyces cerevisiae* × *Saccharomyces kudriavzevii* natural hybrid strains isolated from wine and beer reveals different origins. *BMC Genomics* 13:407. doi: 10.1186/1471-2164-13-407
- Peter Smits, H., Hauf, J., Müller, S., Hobley, T. J., Zimmermann, F. K., Hahn-Hägerdal, B., et al. (2000). Simultaneous overexpression of enzymes of the lower part of glycolysis can enhance the fermentative capacity of *Saccharomyces cerevisiae*. *Yeast* 16, 1325–1334. doi: 10.1002/1097-0061(200010)16:14<1325::AID-YEA627>3.0.CO;2-E
- Piper, M. D., Daran-Lapujade, P., Bro, C., Regenber, B., Knudsen, S., Nielsen, J., et al. (2002). Reproducibility of oligonucleotide microarray transcriptome analyses an interlaboratory comparison using chemostat cultures of *Saccharomyces cerevisiae*. *J. Biol. Chem.* 277, 37001–37008. doi: 10.1074/jbc.M204490200
- Pirt, S. (1982). Maintenance energy: a general model for energy-limited and energy-sufficient growth. *Arch. Microbiol.* 133, 300–302. doi: 10.1007/BF00521294
- Rantasalo, A., Czeizler, E., Virtanen, R., Rousu, J., Lähdesmäki, H., Penttilä, M., et al. (2016). Synthetic transcription amplifier system for orthogonal control of gene expression in *Saccharomyces cerevisiae*. *PLoS One* 11:e0148320. doi: 10.1371/journal.pone.0148320
- Redden, H., and Alper, H. S. (2015). The development and characterization of synthetic minimal yeast promoters. *Nat. Commun.* 6:7810. doi: 10.1038/ncomms8810
- Replansky, T., Koufopanou, V., Greig, D., and Bell, G. (2008). *Saccharomyces sensu stricto* as a model system for evolution and ecology. *Trends Ecol. Evol.* 23, 494–501. doi: 10.1016/j.tree.2008.05.005
- Salazar, A. N., Gorter de Vries, A. R., van den Broek, M., Wijsman, M., de la Torre Cortés, P., Brickwedde, A., et al. (2017). Nanopore sequencing enables near-complete de novo assembly of *Saccharomyces cerevisiae* reference strain CEN. PK113-7D. *FEMS Yeast Res.* 17:fox074. doi: 10.1093/femsyr/fox074
- Salvador, Z., Arroyo-López, F., Guillamón, J., Salazar, G., Querol, A., and Barrio, E. (2011). Temperature adaptation markedly determines evolution within the genus *Saccharomyces*. *Appl. Environ. Microbiol.* 77, 2292–2302. doi: 10.1128/AEM.01861-10
- Scannell, D. R., Zill, O. A., Rokas, A., Payen, C., Dunham, M. J., Eisen, M. B., et al. (2011). The awesome power of yeast evolutionary genetics: new genome sequences and strain resources for the *Saccharomyces sensu stricto* genus. *G3* 1, 11–25. doi: 10.1534/g3.111.000273
- Shen, X.-X., Zhou, X., Kominek, J., Kurtzman, C. P., Hittinger, C. T., and Rokas, A. (2016). Reconstructing the backbone of the *Saccharomycotina* yeast phylogeny using genome-scale data. *G3* 6, 3927–3939. doi: 10.1534/g3.116.034744
- Sicard, D., and Legras, J.-L. (2011). Bread, beer and wine: yeast domestication in the *Saccharomyces sensu stricto* complex. *C. R. Biol.* 334, 229–236. doi: 10.1016/j.crvi.2010.12.016
- Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., et al. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* 7:539. doi: 10.1038/msb.2011.75
- Solis-Escalante, D., Kuijpers, N. G., Barrajon-Simancas, N., van den Broek, M., Pronk, J. T., Daran, J. M., et al. (2015). A minimal set of glycolytic genes reveals strong redundancies in *Saccharomyces cerevisiae* central metabolism. *Eukaryot. Cell* 14, 804–816. doi: 10.1128/EC.00064-15
- Stanke, M., and Waack, S. (2003). Gene prediction with a hidden Markov model and a new intron submodel. *Bioinformatics* 19, ii215–ii225. doi: 10.1093/bioinformatics/btg1080
- Świat, M. A., Dashko, S., den Ridder, M., Wijsman, M., van der Oost, J., Daran, J.-M., et al. (2017). FnCpf1: a novel and efficient genome editing tool for *Saccharomyces cerevisiae*. *Nucleic Acids Res.* 45, 12585–12598. doi: 10.1093/nar/gkx1007
- Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., Van Baren, M. J., et al. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* 28, 511–515. doi: 10.1038/nbt.1621
- Tsien, R. Y. (1998). The green fluorescent protein. *Annu. Rev. Biochem.* 67, 509–544. doi: 10.1146/annurev.biochem.67.1.509
- Van Dijken, J., Bauer, J., Brambilla, L., Duboc, P., Francois, J., Gancedo, C., et al. (2000). An interlaboratory comparison of physiological and genetic properties

- of four *Saccharomyces cerevisiae* strains. *Enzyme Microb. Technol.* 26, 706–714. doi: 10.1016/S0141-0229(00)00162-9
- Van Heerden, J. H., Bruggeman, F. J., and Teusink, B. (2015). Multi-tasking of biosynthetic and energetic functions of glycolysis explained by supply and demand logic. *Bioessays* 37, 34–45. doi: 10.1002/bies.201400108
- Van Hoek, P., Van Dijken, J. P., and Pronk, J. T. (2000). Regulation of fermentative capacity and levels of glycolytic enzymes in chemostat cultures of *Saccharomyces cerevisiae*. *Enzyme Microb. Technol.* 26, 724–736. doi: 10.1016/S0141-0229(00)00164-2
- Verduyn, C., Postma, E., Scheffers, W. A., and Van Dijken, J. P. (1992). Effect of benzoic acid on metabolic fluxes in yeasts: a continuous-culture study on the regulation of respiration and alcoholic fermentation. *Yeast* 8, 501–517. doi: 10.1002/yea.320080703
- Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., et al. (2014). Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 9:e112963. doi: 10.1371/journal.pone.0112963
- Waterhouse, A. M., Procter, J. B., Martin, D. M., Clamp, M., and Barton, G. J. (2009). Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25, 1189–1191. doi: 10.1093/bioinformatics/btp033
- Webster, K. A. (2003). Evolution of the coordinate regulation of glycolytic enzyme genes by hypoxia. *J. Exp. Biol.* 206, 2911–2922. doi: 10.1242/jeb.00516
- Weirauch, M. T., and Hughes, T. R. (2010). Conserved expression without conserved regulatory sequence: the more things change, the more they stay the same. *Trends Genet.* 26, 66–74. doi: 10.1016/j.tig.2009.12.002
- Wolfe, K. H., and Shields, D. C. (1997). Molecular evidence for an ancient duplication of the entire yeast genome. *Nature* 387, 708–713. doi: 10.1038/42711
- Zeevi, D., Sharon, E., Lotan-Pompan, M., Lubling, Y., Shipony, Z., Raveh-Sadka, T., et al. (2011). Compensation for differences in gene copy number among yeast ribosomal proteins is encoded within their promoters. *Genome Res.* 21, 2114–2128. doi: 10.1101/gr.119669.110
- Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Boonekamp, Dashko, van den Broek, Gehrmann, Daran and Daran-Lapujade. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Comparative Genomics Between *Saccharomyces kudriavzevii* and *S. cerevisiae* Applied to Identify Mechanisms Involved in Adaptation

Laura G. Macías<sup>1,2</sup>, Miguel Morard<sup>1,2</sup>, Christina Toft<sup>1,2†</sup> and Eladio Barrio<sup>1,2\*</sup>

<sup>1</sup> Departament de Genètica, Universitat de València, Valencia, Spain, <sup>2</sup> Departamento de Biotecnología, Instituto de Agroquímica y Tecnología de Alimentos IATA, CSIC, Valencia, Spain

## OPEN ACCESS

### Edited by:

Ed Louis,  
University of Leicester,  
United Kingdom

### Reviewed by:

Samina Naseeb,  
The University of Manchester,  
United Kingdom  
Gilles Fischer,  
Sorbonne Université (CNRS), France

### \*Correspondence:

Eladio Barrio  
Eladio.Barrío@uv.es

### † Present address:

Christina Toft,  
Institute for Integrative and Systems  
Biology, Universitat de València  
and CSIC, Valencia, Spain

### Specialty section:

This article was submitted to  
Evolutionary and Genomic  
Microbiology,  
a section of the journal  
Frontiers in Genetics

Received: 21 September 2018

Accepted: 21 February 2019

Published: 13 March 2019

### Citation:

Macías LG, Morard M, Toft C and  
Barrio E (2019) Comparative  
Genomics Between *Saccharomyces*  
*kudriavzevii* and *S. cerevisiae* Applied  
to Identify Mechanisms Involved  
in Adaptation. *Front. Genet.* 10:187.  
doi: 10.3389/fgene.2019.00187

Yeasts belonging to the *Saccharomyces* genus play an important role in human-driven fermentations. The species *S. cerevisiae* has been widely studied because it is the dominant yeast in most fermentations and it has been widely used as a model eukaryotic organism. Recently, other species of the *Saccharomyces* genus are gaining interest to solve the new challenges that the fermentation industry are facing. One of these species is *S. kudriavzevii*, which exhibits interesting physiological properties compared to *S. cerevisiae*, such as a better adaptation to grow at low temperatures, a higher glycerol synthesis and lower ethanol production. The aim of this study is to understand the molecular basis behind these phenotypic differences of biotechnological interest by using a species-based comparative genomics approach. In this work, we sequenced, assembled and annotated two new genomes of *S. kudriavzevii*. We used a combination of different statistical methods to identify functional divergence, signatures of positive selection and acceleration of substitution rates at specific amino acid sites of proteins in *S. kudriavzevii* when compared to *S. cerevisiae*, and vice versa. We provide a list of candidate genes in which positive selection could be acting during the evolution of both *S. cerevisiae* and *S. kudriavzevii* clades. Some of them could be related to certain important differences in metabolism previously reported by other authors such as *DAL3* and *ARO4*, involved in nitrogen assimilation and amino acid biosynthesis. In addition, three of those genes (*FBA1*, *ZIP1*, and *RQC2*) showed accelerated evolutionary rates in *Sk* branch. Finally, genes of the riboflavin biosynthesis were also among those genes with a significant higher rate of nucleotide substitution and those proteins have amino acid positions contributing to functional divergence.

**Keywords:** *Saccharomyces cerevisiae*, *S. kudriavzevii*, comparative genomics, positive selection, functional divergence, evolutionary rate

## INTRODUCTION

How species have adapted to new environments by the action of natural selection shaping their genomes is a key question in modern biology since Charles Darwin proposed the theory of natural selection to explain the origin of adaptations. The Modern Synthesis (Neo-Darwinism), reconciling Darwin's theory of evolution and Mendelian genetics, was based on the idea that

**Abbreviations:** GO, gene ontology; *Sc*, *Saccharomyces cerevisiae*; *Sk*, *Saccharomyces kudriavzevii*; SSD, small-scale duplication; WGD, whole genome duplication.



most natural populations contain enough genetic variation, generated by mutation, to respond to any sort of selection, and explained adaptation as the gradual evolution resulting from changes in the frequencies of the genetic variants acted upon by natural selection. However, with the proposal of the neutral theory of molecular evolution (Kimura, 1983), it has widely been assumed that most mutations are neutral or deleterious, depending on their functional constraints. In contrast, advantageous mutations constitute a very small fraction of the total but are responsible for adaptation. As deleterious mutations are removed by purifying selection, most genetic variation, both within-species polymorphisms and between-species divergence, is the result of the action of genetic drift, with a negligible contribution of the rare beneficial mutations fixed by positive selection (Kimura, 1983). In recent years, several authors propose conciliation between neutralism and selectionism by considering that fixed neutral mutations can become advantageous by shifts in the selective pressures, and hence, promote later evolutionary adaptation (Wagner, 2008). According to Michael Lynch (2007), p. 375: “the non-adaptive force of random genetic drift set the stage for future paths of adaptive evolution in novel ways that would not otherwise be possible.”

In the genomic era, an important challenge is to determine whether patterns of genome variation can be explained by random genetic drift or selection. However, the rapid acquisition of more and more genome sequences, together with the development and improvement of statistical methods for comparative genomics, allow us to unveil the evolutionary forces responsible for adaptation at the molecular level.

The genus *Saccharomyces* is composed of eight species (Boynton and Greig, 2014): *S. arboricola*, *S. cerevisiae*, *S. eubayanus*, *S. kudriavzevii*, *S. mikatae*, *S. paradoxus*, *S. uvarum*, and the recently described *S. jurei* (Naseeb et al., 2017). Yeasts belonging to this genus have mostly been isolated in wild environments. The exception is *S. cerevisiae* (Sc), one of the most well-studied microorganisms, which has also been found in a wide range of human-manipulated fermentative environments such as wine, cider, sake, beer, bread, etc., as well as in traditional fermentations (Liti et al., 2009; Gallone et al., 2016; Peter et al., 2018). In a lesser extent, *S. uvarum* (Su) is also present in wine and cider fermentations from regions of cold climate, where coexists or even replaces *S. cerevisiae* (Almeida et al., 2014; Rodríguez et al., 2017). In addition, different types of interspecific *Saccharomyces* hybrids have also been isolated in fermentations from cold regions (González et al., 2006; Morales and Dujon, 2012; Pérez-Través et al., 2014). Another interesting species from this genus is *S. kudriavzevii* (Sk). This species is isolated only from wild environments, such as oak barks and decayed leaves in Asia (Naumov et al., 2000, 2013) and Europe (Sampaio and Gonçalves, 2008; Lopes et al., 2010; Erny et al., 2012). Although Sk has never been found in fermentations, its double hybrids with Sc and triple with Sc and Su appear, and even dominate, in wine, beer and cider fermentations in regions of cold climates (Peris et al., 2018).

To understand the contribution of the Sk parent to its hybrids, several comparative physiological studies between Sc and Sk have been performed (González et al., 2007; Belloch et al., 2008;

Arroyo-López et al., 2009; Gangl et al., 2009). This way, these results indicate that hybrids acquired the high alcohol tolerance trait of Sc (Arroyo-López et al., 2010b), and the better adaptation to grow at low temperatures of Sk (Salvadó et al., 2011). These physiological differences have been related to modifications in the components of lipid membrane of both species (Tronchoni et al., 2012), and in the production of glycerol (Arroyo-López et al., 2010a). The lower ethanol yield and the higher glycerol synthesis, together with differences in the aroma production (Stribny et al., 2015) and an optimal growth under low pH (Arroyo-López et al., 2009) indicate that Sk and its hybrids are good potential candidates for future applications in the wine industry (Alonso-Del-Real et al., 2017; Pérez-Torrado et al., 2018).

At the same time, different studies have been performed to unravel the genetic basis responsible for the phenotypic differences observed between Sc and Sk, especially for those that study the low temperature adaptation. The analysis of the glycerol synthesis pathway showed that the higher glycerol production in *S. kudriavzevii* is due to an enhanced enzymatic activity of its glycerol-3-phosphate dehydrogenase Gpd1p (Oliveira et al., 2014). A transcriptomic study revealed that Sk exhibits a higher ability to initiate the translation of crucial genes in cold adaptation (Tronchoni et al., 2014). A systems biology study applied to both Sc and Sk revealed that pathways such as lipid, oxidoreductase and vitamin metabolism were directly involved with the fitness of these species at low temperatures (Paget et al., 2014). Main phenotypic differences between Sk and Sc are described but most of genes responsible for these phenotypes remain unknown.

In the present study, we applied for the first time diverse comparative approaches to study adaptive differences and functional divergence between both *Saccharomyces* species at genome-wide level. We sequenced, *de novo* assembled and annotated two new genomes of Sk strains isolated from Spanish tree barks. We used complete genome sequences of these strains as well as those from two Sk previously sequenced (Scannell et al., 2011), and four genome sequences from representative Sc strains, to identify selective shifts in a set of orthologous genes in both Sc and Sk leading branches. Functional divergence among orthologous proteins was also quantified leading to the identification of the most functional divergent pathways between Sk and Sc.

## MATERIALS AND METHODS

### Assembly and Annotation

*Saccharomyces kudriavzevii* strains CR85 and CA111 were isolated in a previous work (Lopes et al., 2010) and their genomes were sequenced in this study. These strains were sequenced by Illumina MiSeq with paired-end 300 bp reads. In addition, Sk CR85 was also sequenced using Roche 454 shotgun sequencing and paired-end reads of 8 kb.

*De novo* assembly of the Sk CR85 genome was carried out using MIRA v3.4.1.1<sup>1</sup> and GS *de novo* Assembler (Roche/454 Life

<sup>1</sup><https://sourceforge.net/p/mira-assembler/wiki/Home/>

Sciences, Branford, CT, United States). Manual checking and corrections of the assembly were done using Consed (Gordon and Green, 2013).

Assembly of CA111 strain was performed using Velvet v1.1 (Zerbino and Birney, 2008) to determine the best k-mer value and then Sopra v1.4. (Dayarian et al., 2010) was used for *de novo* assembly. To get the scaffolds into chromosome structure, ultra-scaffolds were generated with an in-house script, which orders the contigs according to their homology to a reference genome, in our case *S. cerevisiae* S288c. A whole genome aligner MUMmer (Kurtz et al., 2004) was used to generate this information. Final assembly sizes of 11.75 and 11.89 Mb were obtained for *Sk* CR85 and *Sk* CA111, respectively, (Supplementary Table S1).

Reannotations of *Sk* NBRC 1802 and ZP 591 genomes (Scannell et al., 2011) were also performed due to problems with the original annotations. Two approaches were used for the annotation of the four *Sk* genomes: first, a transfer of annotations from *Sc* S288c (Goffeau et al., 1996) by using RATT tool (Otto et al., 2011), and second, a novel gene prediction with Augustus (Stanke and Morgenstern, 2005). Finally, the annotations were manually verified by using Artemis (Rutherford et al., 2000). With this pipeline, 5664 genes in NBRC 1802 strain were annotated, 5575 in ZP591 strain, 5623 in CR85, and 5492 in CA111.

## Orthology and Alignment

We also used four well-annotated genome sequences from different populations of *S. cerevisiae* (Liti et al., 2009), as representative strains of this species (Table 1). The genome sequence of *Torulaspora delbrueckii* (Gordon et al., 2011) was used as outgroup. This species was selected because it diverged from the *Saccharomyces* genus before the Whole Genome Duplication (WGD) event (Wolfe and Shields, 1997). This was done to ensure the use of orthologous reference sequences in the analyses, which is not necessarily true if a post-WGD

species is selected as outgroup due to differential loss of paralogous (ohnologous) genes (Scannell et al., 2007). Orthology among the three species was defined according to synteny information available in the Yeast Genome Order Browser (YGOB) (Byrne and Wolfe, 2005).

Alignments for all orthologous sequences were obtained using Mafft v7.221 (Kato and Standley, 2013). A total number of 4164 orthologous genes were found in common among the three species. In some cases, as *T. delbrueckii* was a pre-WGD species, the same gene sequence was aligned against two different gene sequences of *Saccharomyces* genomes, those duplicated genes generated by the WGD event, according to the YGOB.

## Signatures of Positive Selection

To identify genes being potentially under positive selection in both *Sk* and *Sc* branches, we performed a comparison of the likelihood scores of selection models implemented in the branch-site CodeML software of the PAML package, version 4.5 (Yang, 2007). The branch-site test was used to detect positive selection acting at specific codons in a defined branch of a phylogenetic tree. This branch is known as the foreground and the rest as background branches. The branch-site test compares a model considering three fractions of codon sites with a null model with only two fractions of codons. In the three-fraction model, the first fraction ( $p_0$ ) evolved in both foreground and background branches with a non-synonymous/synonymous substitution ratio of  $\omega_0 < 1$  (purifying selection), the second ( $p_1$ ) with  $\omega_1 = 1$  (neutral) in both sets of branches, and the third ( $p_2$ ) evolved with  $\omega_0 > 1$  (positive selection) in the foreground branch but with  $\omega_0 < 1$  or  $\omega_1 = 1$  in the background branches. The null model considers only two fractions, one evolved with  $\omega_0 < 1$  and the other with  $\omega_1 = 1$  in both sets of branches. Since this is a species-based method, we first set as the foreground branch the one leading to the *Sk* clade. Then, we repeated the analysis by setting as foreground the *Sc* clade branch. Both analyses were performed using *T. delbrueckii* as an outgroup species.

All genes whose Likelihood Ratio Test (LRT)  $\chi^2$  analysis, with one degree of freedom (the difference of free parameters between models), reached *p*-values lower than 0.05 were considered significant, and those genes containing a fraction of codons with  $\omega > 1$  were selected as gene candidates to be under positive selection. In these genes, Bayesian posterior probabilities for site classes were estimated, with the Bayes Empirical Bayes (BEB) method (Yang et al., 2005), to identify amino acid sites under positive selection.

## Testing Constant Rate of Evolution

The molecular constant rate of evolution was tested for all orthologous genes analyzed in both *Sc* and *Sk* species, by using *T. delbrueckii* as outgroup. Tajima relative rate test (Tajima, 1993) was implemented by using an in-house built Python script. A singleton was defined as a change in the nucleotide sequence specific for every one of the three species included in the alignment. Number of observed singletons in each gene alignment was calculated according to the formulas:

$$m_1 = \sum_i \sum_{j \neq i} n_{ij} \quad m_2 = \sum_i \sum_{j \neq i} n_{ji} \quad m_3 = \sum_i \sum_{j \neq i} n_{jji} \quad (1)$$

**TABLE 1** | List of strains and sources of the genomic sequences used in this study.

Strain	Origin	Source	Reference
<i>S. kudriavzevii</i> CR85	Ciudad Real, Spain	<i>Quercus ilex</i> bark	This study
<i>S. kudriavzevii</i> CA111	Castellón, Spain	<i>Quercus ilex</i> bark	This study
<i>S. kudriavzevii</i> ZP591	Cast. Vide, Portugal	<i>Quercus pyrenaica</i>	Scannell et al., 2011
<i>S. kudriavzevii</i> NBRC1802	Japan	Decayed leaf	Scannell et al., 2011
<i>S. cerevisiae</i> T73	Alicante, Spain	Wine	Morard et al., 2019
<i>S. cerevisiae</i> S288c	–	Laboratory	Goffeau et al., 1996
<i>S. cerevisiae</i> YPS128	Pennsylvania, United States	<i>Quercus</i> forest soil	Liti et al., 2009
<i>S. cerevisiae</i> Y9	Indonesia	Ragi (similar to sake)	Liti et al., 2009
<i>T. delbrueckii</i> CBS1146	Unknown	Unknown	Gordon et al., 2011

where  $i$  is the variable position in the alignment and  $j$  is the nucleotide that is conserved in two out of three sequences of the alignment.  $m_1$ ,  $m_2$ , and  $m_3$  are the total number of singletons in one alignment for *Sc*, *Sk* and the outgroup species, respectively.

Under the molecular clock hypothesis, the number of singletons in *Sc* and *Sk* species are expected to be the same, therefore, the expected singletons according to this hypothesis was calculated as:

$$E(m_1) = E(m_2) = (m_1 + m_2) / 2 \quad (2)$$

For every nucleotide alignment, number of *Sc* and *Sk* singletons was calculated and it was compared with the number of expected changes under the molecular clock hypothesis. A  $\chi^2$  test with one degree of freedom was applied to assess whether the difference between the observed and the expected singletons was significant and, if so, the molecular clock hypothesis was rejected.

## Functional Divergence

In this work, functional divergence type I was identified. This type of functional divergence involves the change in selection constraints acting at specific amino acid sites of a protein in a specific phylogenetic clade (which will be defined as clade-of-interest) when compared to another clade. A method previously described by Toft et al. (2009), was used to identify amino acid sites which have diverged significantly from the output sequence in a clade of interest with respect to the homologous sites in a second clade. This test was performed twice by defining as the clade-of-interest *Sk* or *Sc*. Once all divergent amino acid sites were obtained, results were filtered by Grantham's scores (Grantham, 1974), to quantify the biochemical divergence between *Sc* and *Sk* amino acids. Scores of 120 and higher were considered for further analyses as sites that have radically changed in *Sk* when compared to *Sc* and which might have functional importance for the protein, these results were normalized by the protein length.

We also tested whether there was any genome region enriched in proteins showing evidence of functional divergence. This task was assessed by checking if the mean of normalized functional divergence values from non-overlapping windows of ten genes fall within the 95% confidence interval resulting from generating a random distribution after sampling  $10^6$  times ten genes from the whole set of genes analyzed.

Finally, functional divergence was also determined in terms of domain architecture. SUPERFAMILY hidden Markov models available in the SUPERFAMILY database (ver. 1.75., last accessed February 20, 2018) (Gough et al., 2001) were used for domain assignment according to the Structural Classification of Proteins (SCOP) database to get domain annotations for every *Sk* and *Sc* orthologous pair of proteins using the same criteria as described in Grassi et al. (2010). Orthologous pairs with identical domain architecture, which exhibit no domain architecture functional divergence, were annotated as class A. Orthologous carrying similar domain architectures but differed in domain copy number were annotated as class B. Finally, class C contained *Sc-Sk* orthologous whose domain architectures differed in the presence or absence of one or more domains.

## Duplicated Genes

A careful examination of duplicated genes was done after performing all analyses previously mentioned. We defined duplicated gene pairs as the resulting best reciprocal hits from all-against-all BLAST (Altschul et al., 1990) searches using BLASTP with an  $E$ -value cut-off of  $1 \cdot 10^{-5}$  and a bit score cut-off of 50. Duplicated pairs were then classified as ohnologous gene pairs, generated by the whole genome duplication event (WGDs) according to the Yeast Gene Order Browser (YGOB) list (Byrne and Wolfe, 2005). All other paralogous gene duplicates were considered as derived from small-scale duplications (SSDs).

## Gene Ontology and Pathway Enrichment Analyses

For every analysis previously mentioned, a list of candidate genes was obtained. Gene ontology (GO) term and pathway enrichments were performed using the Gene List tool available in the *Saccharomyces* Genome Database<sup>2</sup>, by considering the list of all 4164 aligned genes used in this study as background population. Results were filtered by a  $p$ -value lower than 0.05 after a Holm-Bonferroni test correction (Aickin and Gensler, 1996).

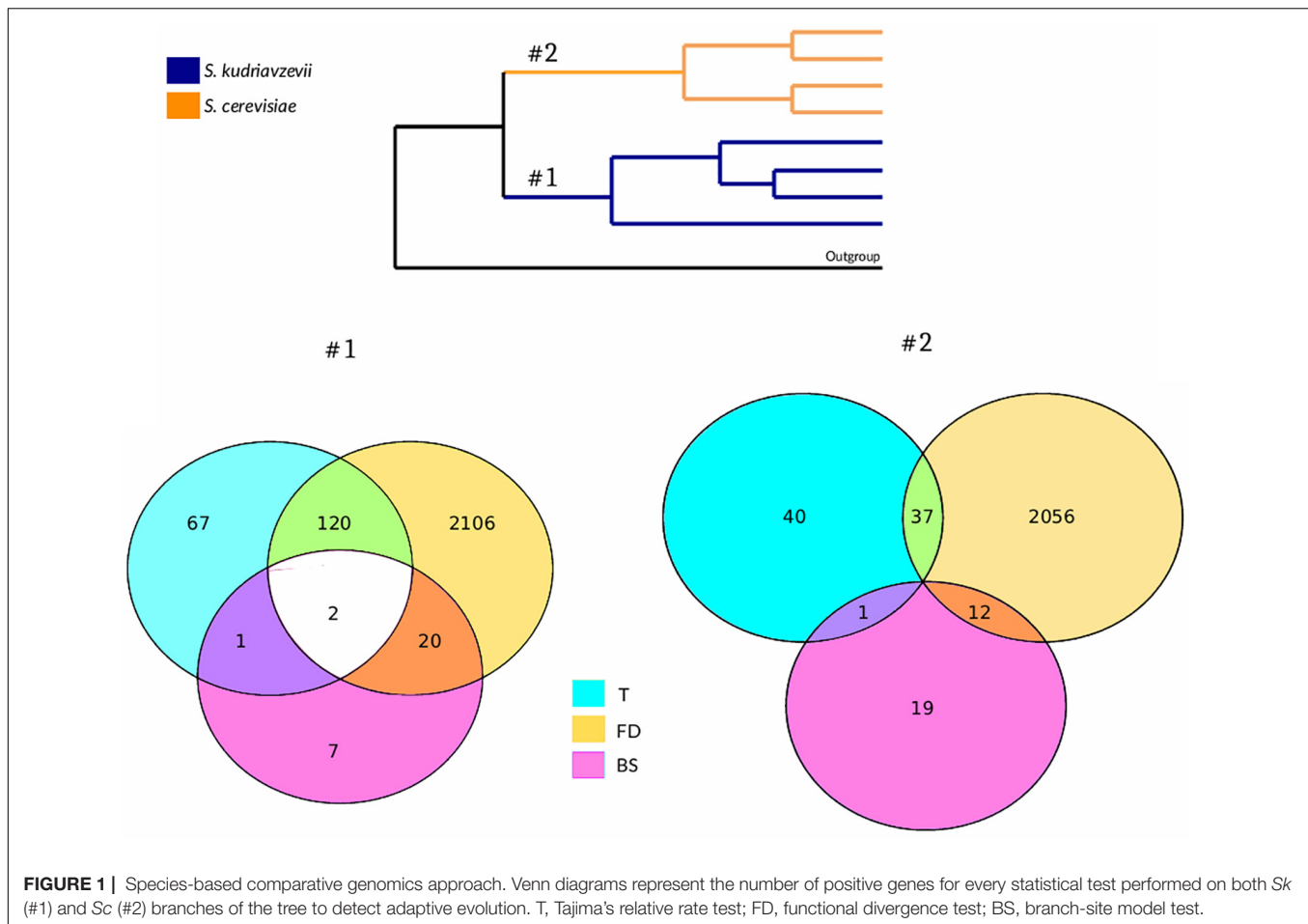
## RESULTS

A species-based comparative genomics approach has been applied to investigate the genetic basis behind the main phenotypic differences already reported between *Sk* and *Sc*. As representatives of *Sk*, we included four complete genome assemblies from strains of different origins. Those from strains NBRC 1802 and ZP 591 were publicly available from a previous study (Scannell et al., 2011) and the other two, corresponding to strains isolated from oak bark samples taken in different locations of Spain, were sequenced, *de novo* assembled and annotated for the present study. Our assembly and annotation pipeline, that combines transfer of annotation and *de novo* gene prediction with a final accurate manual correction, allowed us to provide *Sk* high-quality annotation avoiding common errors such as paralogs mislabelling, coming from the sole use of automatic annotation pipelines. For this reason, NBRC 1802 and ZP 591 genome assemblies were also re-annotated using the same pipeline. In the case of *Sc*, we included in the analyses well-annotated genomes of four strains as representatives of the main lineages (Liti et al., 2009; Peter et al., 2018).

## Differential Adaptive Evolution Between *S. cerevisiae* and *S. kudriavzevii*

The presence of signatures of adaptive evolution in coding genes was tested in both species by using three different approaches: branch-site test of selection, Tajima's rate of evolution test and functional divergence test, which results are summarized in **Figure 1**. GO and pathway enrichment analyses performed for genes showing a positive result simultaneously for more than one of the tests mentioned revealed no significant results.

<sup>2</sup><https://yeastmine.yeastgenome.org/yeastmine/bag.do>



Using the branch-site model, we obtained 30 genes under positive selection when the branch leading to *Sk* was considered as foreground branch. Additionally, 32 genes were found under positive selection when *Sc* was set as the foreground branch (**Supplementary Tables S2, S4**). Neither GO nor pathway enrichment were obtained for these lists. Only two genes, *FRT2* and *RQC2*, showed evidence of positive selection on both branches.

Tajima's relative rate test was applied to detect higher rate of nucleotide substitution at specific coding sequences. Using this test, 190 genes in *Sk* branch and 78 genes in *Sc* branch were obtained (**Supplementary Table S2**). The difference between the numbers of genes detected on both branches was significant (Fisher's exact test:  $F = 2.5$ ,  $p\text{-value} = 2.71e^{-12}$ ). No GO term enrichment was found for none of the two lists. No pathway enrichment was found for *Sc* branch results whereas an enrichment of genes belonging to riboflavin pathway (*RIB2*, *RIB3*, *RIB5*, and *FMN1*) was found in *Sk* branch results. In the *Sk* branch, three genes showed an acceleration in evolutionary rates and were found to be under positive selection: *FBA1*, *ZIP1*, and *RQC2*, while in the *Sc* branch, only one gene (*STE24*) was detected in both statistical tests.

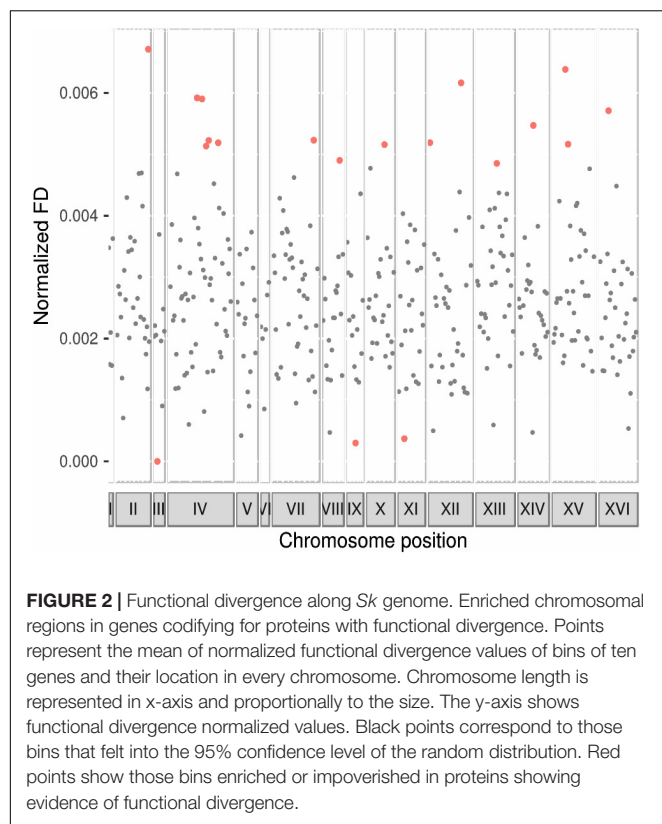
A set of proteins showing evidence of functional divergence and the specific amino acid sites that are contributing to

this phenomenon was obtained for both *Saccharomyces* species (**Supplementary Table S2**). Using this approach, 2248 proteins out of 4164 analyzed (54%) showed evidence of functional divergence when *Sk* was compared to their *Sc* orthologous proteins. On the other hand, 2105 proteins (~50%) were found to be under functional divergence when *Sc* was set as the clade-of-interest.

To assess whether there was any region in the genome of *Sk* showing an enrichment in genes codifying for functionally divergent proteins, we evaluated chromosomal regions in non-overlapping windows of ten genes (**Figure 2** and **Supplementary Table S5**). One region containing ten genes in chromosomes III and IX were impoverished in proteins showing functional divergence in the *Sk* clade. One region in each of the chromosomes II, VII, VIII, X, XI, XIV, and XVI, two regions in chromosomes XII and XV, and five regions in chromosome IV were enriched in functionally divergent proteins.

In addition, functional divergence was evaluated in terms of protein domain architecture. Domain-based functional analysis leads us to get an additional perspective on the possible biological differences between *Sk* and *Sc* orthologous proteins. SCOP domains were assigned for every *Sk*-*Sc* orthologous pair of proteins (**Supplementary Table S2**). A total number of 2544 proteins were annotated with SCOP domains for *Sc* and 2550





for *Sk*. Of them, 2402 proteins were classified as class A as they showed no evidence of functional divergence in protein domain architecture. Other 54 proteins were classified as class B because they differed in copy number domain. Finally, 96 proteins were classified in class C as they carried domain architectures which differed in the presence or absence of any of the domains of their orthologous gene. No GO or pathway enrichment were found for groups B or C. Two genes belonging to category B, *SEC7* and *SMC1*, and five genes from category C, *YNL144C*, *FAR1*, *PRP19*, *SMC4* and *ZIP1*, showed accelerated evolutionary rates as well.

Contribution of genes under functional divergence to every metabolic pathway was analyzed to identify those pathways more functionally divergent in *Sk* and *Sc* (Figure 3 and Supplementary Figure S1). Amino acid biosynthesis, glycerophospholipid metabolism, GPI-anchor biosynthesis, N-glycan biosynthesis and purine and pyrimidine metabolisms were found between the pathways containing a higher number of functionally divergent proteins in both *Sk* and *Sc*.

We also assessed the significance of the differences between *Sc* and *Sk* in normalized functional divergence values belonging to the different pathways (Supplementary Figure S2). In the *Sk* branch, proteins related to metabolism of riboflavin and biosynthesis of various types of N-glycans showed highly normalized functional divergence values and their difference with the same values calculated for the *Sc* branch was found to be significant (Figure 3, bottom panel).

Finally, functional divergence was evaluated in pairs of duplicated sequences (WGDs and SSDs) coming from gene

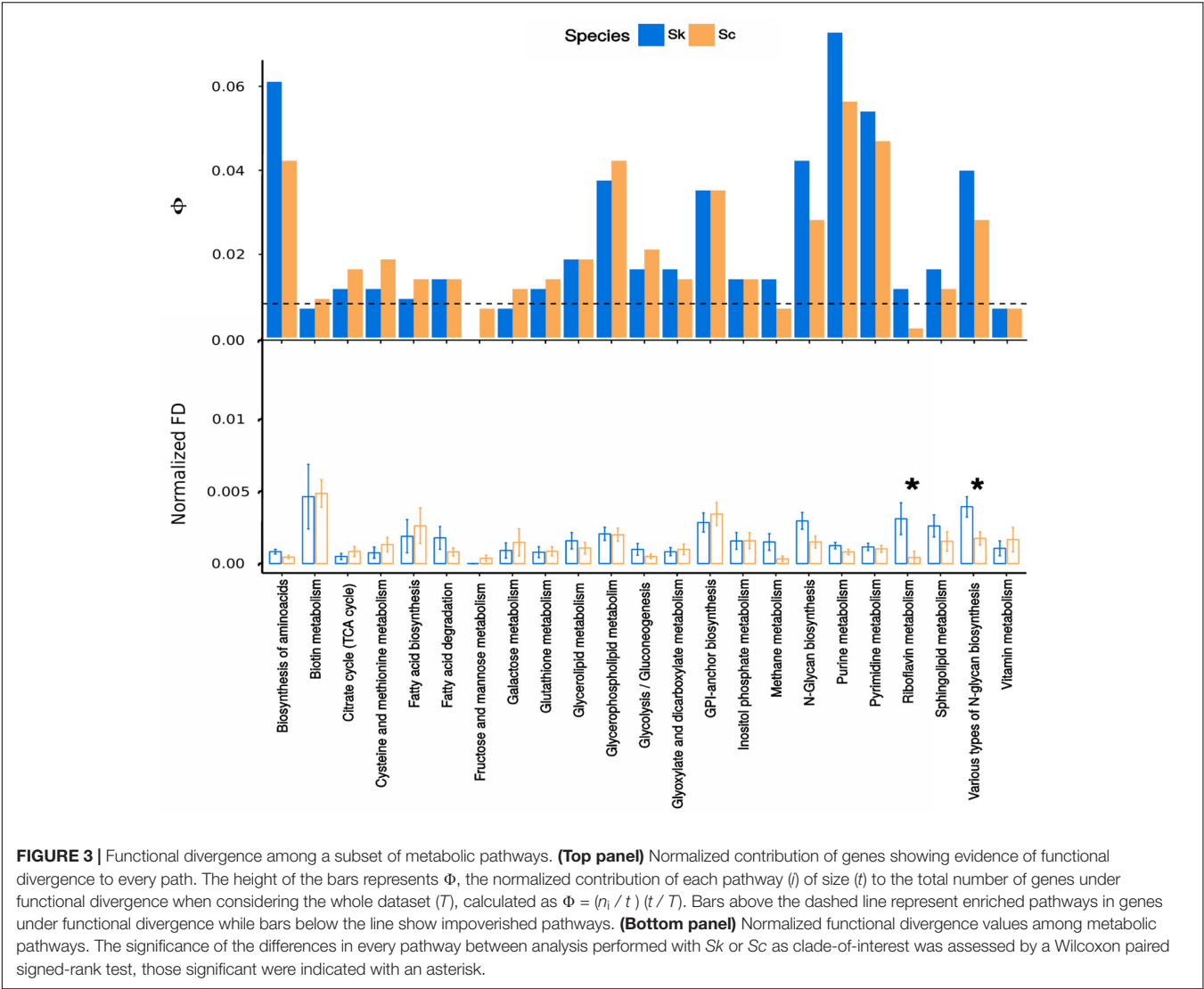
duplication events (Table 2). There were not significant differences in the ratio between singletons in any of the duplicated sequences. We observed more WGDs than SSDs cases of functional divergence ( $F = 1.08$ ), although this difference was not significant ( $p$ -value = 0.66).

### Evidence of Adaptive Evolution in Genes Related With Known Physiological Differences Between the Two *Saccharomyces* Species

Detecting traces of positive selection could be challenging. As mentioned above, very few genes were obtained with the statistical methods applied for detecting adaptive evolution such as the branch site test and the Tajima's relative rate analysis. Contrastingly, at least half of the proteins analyzed for both species showed amino acid positions that could be contributing to functional divergence. In this section, we want to highlight those genes detected in our analysis that could have a role in the phenotypic differences between *S. cerevisiae* and *S. kudriavzevii* according to physiological characterizations performed in our group. Thus, these two *Saccharomyces* species show differences in their carbon metabolism (Arroyo-López et al., 2010a; López-Malo et al., 2013; Oliveira et al., 2014). In our branch-test analysis, we detected adaptive evolution in gene *FBA1*, an essential gene encoding a fructose 1,6-bisphosphate aldolase (Schwelberger et al., 1989). This enzyme has a crucial role in the glycolysis pathway, it catalyzes the conversion of a high-energy hexose, fructose 1,6-bisphosphate, into two interconvertible phosphorylated trioses, glyceraldehyde-3-phosphate and dihydroxyacetone-phosphate, just at the branching point where these trioses can be directed to the end of glycolysis and ethanol fermentation or to the synthesis of glycerol. This gene also showed acceleration of evolutionary rate in *Sk* branch according to the Tajima's relative rate test results.

Differences in nitrogen metabolism and aroma synthesis have also been reported (Gamero et al., 2015; Stribny et al., 2015). One of the genes under positive selection in the *Sk* branch is *ARO4*, which encodes a 3-deoxy-D-arabino-heptulosonate-7-phosphate synthase that catalyzes the first step in aromatic amino acid biosynthesis (Künzler et al., 1992). Another gene is *DAL3*, which codifies for a ureidoglycolate lyase with a role in the third step of allantoin degradation (Yoo et al., 1985). *DAL3* belongs to the allantoin cluster (Wong and Wolfe, 2005) together with the genes *DAL1*, *DCG1*, *DAL2*, *DAL5*, *DAL7*, and *DUR1,2*. Although *DAL5* and *DAL7* were not included in the set of 4164 genes analyzed because they were missing in some genomes, the rest of genes of the allantoin cluster included in the analyses encode proteins that showed signals of functional divergence (*Dur1,2p*, *Dal1p*, *Dal2p*, *Dal3p*, and *Dcg1p*).

The riboflavin pathway was found to be enriched for genes showing accelerated evolutionary rates in *Sk* branch. Riboflavin is required for the synthesis of the cofactors flavin mononucleotide (FMN) and flavin adenine dinucleotide (FAD) (Ghisla and Massey, 1989). Additionally, four genes involved in this pathway, *RIB2*, *RIB3*, *RIB5*, and *RIB7*, encoded proteins that showed functional divergence for *Sk*.



Adaptive Evolution in Genes for Which No Previous Physiological Data Is Available

Our approach also detected several genes for which no experimental data about physiological differences is available, and therefore, can be the subject of future studies. For instance,

**TABLE 2 |** Number of genes with a positive result in positive selection, functional divergence and Tajima's relative rate test analyses.

Type of gene	Genes under positive selection		Genes rejecting molecular clock hypothesis		Genes showing functional divergence		Total genes analyzed
	Sc	Sk	Sc	Sk	Sc	Sk	
Singletons	25	26	65	157	1734	1849	3439
SSD	2	1	6	19	151	169	330
WGD	5	3	7	15	221	230	395

Genes are classified in singletons and duplicates by WGD or SSD.

*FRT2* and *RQC2* were found under positive selection in both *Sk* and *Sc* branches. *FRT2*, also known as *HPH2*, encodes for a membrane protein of the endoplasmic reticulum (Burri and Lithgow, 2004), which interacts with the protein encoded by its paralog *FRT1* (*HPH1*), duplicated after the WGD. Although their functions are not well known, both paralogs have been associated to physiological stress response as they can promote growth when there is a high concentration of  $\text{Na}^+$  in the environment (Heath et al., 2004). *RQC2* encodes for a component of the ribosome quality control (RQC) complex which takes part in the degradation of aberrant nascent proteins (Brandman et al., 2012), and also has a role in the recruitment of alanine-to-threonine- charged tRNAs (Shen et al., 2015). It has also been reported that *RQC2* is responsible for communicating translation stress signal to the heat shock transcription factor *HSF1* (Brandman et al., 2012).

Finally, *ZIP1* not only has been found under positive selection in *Sk* branch, it also showed accelerated evolutionary rates. In addition, amino acid positions contributing to functional

divergence and different SCOP domains have been observed. *Sk ZIP1* encodes a Zip1p protein carrying a tropomyosin domain, which it is not present in *Sc ZIP1*. This gene encodes a transverse filament protein that conforms the synaptonemal complex and it is required for meiotic chromosome synapsis, acting as a molecular zipper to facilitate the interaction between homologous chromosomes (Sym et al., 1993). This is correlated with the GO enrichment results of functionally divergent proteins on both *Sc* and *Sk* clades, which revealed an enrichment in biological processes such as cell cycle and cellular component like cellular bud neck (**Supplementary Table S3**).

## DISCUSSION

*Saccharomyces kudriavzevii* is a species from the *Saccharomyces* genus isolated from natural environments such as tree barks and decayed leaves (Boynton and Greig, 2014). On the contrary, *Sc* is a species very well known, isolated from a wide range of environments and frequently related to human-driven industrial processes (Goddard and Greig, 2015). Although *Sk* ecological niche is still not well understood, phenotypic differences existing between *Sk* and *Sc* have been addressed in previous studies (reviewed in Pérez-Torrado et al., 2018).

In an attempt to understand the genetic basis behind the main phenotypic differences between *Sk* and *Sc* we have proposed to trace the genomic changes occurred as a consequence of the adaptation of these species to the different environments.

Despite this study relied on a small set of *Sk* genomes, we have assessed some general insights into the genetic differences between the well-studied *Sc* and *Sk*. Understanding the evolutionary process of the adaptation of *Sk* and *Sc* requires a pluralistic approach. This way we have applied methods to detect signatures of strong selection in coding sequences combined with differences observed at protein level such as functional divergence and accelerated rates of substitution.

The positive selection analyses revealed three genes related to metabolism that might be good candidates to explain differences between both species: *FBA1*, *ARO4* and *DAL3*. As mentioned, *FBA1* is involved in the synthesis of dihydroxyacetone phosphate, the precursor of the glycerol synthesis. Previous studies have shown how *Sk* is able to produce higher amounts of glycerol when compared to *Sc* (Arroyo-López et al., 2010a). Here we proposed that the positive selection observed in this gene together with the acceleration in the evolutionary rates, as shown after the performance of the Tajima's relative rate test, are signatures of adaptation and the *Sk* version of the *FBA1* may have an importance in the synthesis of glycerol in the cell (Oliveira et al., 2014). In addition, the reaction catalyzed by this enzyme has been proposed as cold-favoring (Paget et al., 2014), so the thermal stability of this enzyme could be also an important factor to take into account to explain the patterns of adaptation observed in this gene (Gonçalves et al., 2011).

*ARO4* is involved in aromatic amino acid biosynthesis. Previous works have demonstrated the differences in amino acid metabolism among closely related *Saccharomyces* species and the ability of *Sk* to produce different amounts of aroma compounds

such as higher alcohols and acetate esters from amino acid precursors (Stribny et al., 2015). Therefore, we proposed that the evidence of positive selection observed in this gene could be related to the phenotype already mentioned.

*DAL3* is part of the allantoin gene cluster (Wong and Wolfe, 2005), and it is involved in the allantoin degradation pathway. Allantoin has been found in similar environments as those in which *Sk* has been isolated, like tree bark exudates, and it has been demonstrated to have an important effect on the fitness of yeast living in natural environments (Filteau et al., 2016). This nitrogen source, especially when it is limited, has been shown to cause a rapid effect in the yeast genomes due to environmental adaptation (Gresham et al., 2011). The evidence of selection acting on this gene, together with the fact that the whole cluster showed functional divergence, could explain that *Sk* is better adapted to natural environments in which allantoin is more frequently found rather than in human-related environments.

Differences in functional divergence values revealed that proteins belonging to metabolism of riboflavin pathway were significantly different in *Sk* than in *Sc*. *RIB2*, *RIB3*, *RIB5* and *FMN1* were also found to have their evolutionary rates accelerated when compared to *Sc*. Positions contributing to protein functional divergence were also found to be related to protein structure stability. A previous systems biology study which used these two species of yeasts because of their differences in temperature growth revealed that genes related to riboflavin were potentially affected by cold temperature because vitamins might have an important role at low temperatures (Paget et al., 2014).

The analysis of functional divergence in *Sk* also revealed a high number of genes involved in cellular response to osmotic and oxidative stress and sphingolipid metabolic pathway. Sphingolipids play very important roles in yeasts, being involved in signal transmission, cell recognition, regulation of endocytosis, ubiquitin-dependent proteolysis, cytoskeletal dynamics, cell cycle, translation, post-translational protein modification, and heat stress response (Coward and Obeid, 2007).

In this work, we have increased the number of *Sk* genomes, which allowed us to conduct comparative analyses to unveil some of the mechanisms involved in the differential adaptation of *Sc* and *Sk*. We used methods making different assumptions just to validate the reliability of our results and their interpretation. The inferred cases of positive selection deserve further research, especially with the experimental testing of functional divergence.

## DATA AVAILABILITY

The whole genome sequence datasets generated from this study were deposited in the European Nucleotide Archive (ENA) under accession number PRJEB31099.

## AUTHOR CONTRIBUTIONS

EB and CT conceived and designed the study. LM and MM performed all the analyses under CT and EB supervision.

LM and CT wrote the first versions of the article. EB wrote the final version.

## FUNDING

This work was funded by grant AGL2015-67504-C3-3-R from the Spanish Government and European Union ERDF-FEDER to EB and by ERA CoBioTech MeMBrane Project PCI2018-093190. LM was supported by the aforementioned grant associated to EB. MM was supported by a Ph.D. student contract ACIF/2015/194 from the Regional Government of Valencia. CT acknowledged a “Juan de la Cierva” postdoctoral contract JCI-2012-14056 from the Spanish Government.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2019.00187/full#supplementary-material>

**TABLE S1** | Assembly statistics for *Sk* CA111 and CR85 genome assemblies.

**TABLE S2** | Positive selection, evolutionary rate test and functional divergence results for all genes analyzed in both *Sk* and *Sc* species. PS, positive selection;

FD, number of positions contributing to functional divergence in every protein; SCOP, annotations with SCOP domains as explained in Methods; NA (not available), protein domain annotation was not possible; Tajima's test, *Sk*: acceleration of evolutionary rates leading to *Sk* branch; *Sc*, acceleration of evolutionary rates leading to *Sc* branch.

**TABLE S3** | GO term enrichment for genes showing evidence of functional divergence in *Sk* and *Sc* branch.

**TABLE S4** | Amino acids sites under positive selection according to branch-site model and BEB method. Specific amino acids sites and the probability of being under positive selection were retrieved for those candidates obtained with the branch-site model.

**TABLE S5** | Enriched/Impoverished chromosome regions in proteins with functional divergence evidence.

**FIGURE S1** | Functional divergence among metabolic pathways. Normalized contribution of genes showing evidence of functional divergence to every path. The height of the bars represents  $\Phi$ , the normalized contribution of each pathway ( $i$ ) of size ( $t$ ) to the total number of genes under functional divergence when considering the whole dataset ( $T$ ), calculated as  $\Phi = (n_i / t) (t / T)$ . Bars above the dashed line represent enriched pathways in genes under functional divergence while bars below the line show impoverished pathways. B, biosynthesis; M, metabolism; D, degradation; aa, amino acid.

**FIGURE S2** | *Sk* vs. *Sc* differences in functional divergence among metabolic pathways. Normalized functional divergence values among metabolic pathways. The significance of the differences in every pathway between analysis performed with *Sk* or *Sc* as clade-of-interest was assessed by a Wilcoxon paired signed-rank test, those significant were indicated with an “\*.” B, biosynthesis; M, metabolism; D, degradation; aa, amino acid.

## REFERENCES

- Aickin, M., and Gensler, H. (1996). Adjusting for multiple testing when reporting research results: the bonferroni vs holm methods. *Am. J. Public Health* 86, 726–728. doi: 10.2105/AJPH.86.5.726
- Almeida, P., Gonçalves, C., Teixeira, S., Libkind, D., Bontrager, M., Masneuf-Pomarède, I., et al. (2014). A gondwanan imprint on global diversity and domestication of wine and cider yeast *Saccharomyces uvarum*. *Nat. Commun.* 5:4044. doi: 10.1038/ncomms5044
- Alonso-Del-Real, J., Contreras-Ruiz, A., Castiglioni, G. L., Barrio, E., and Querol, A. (2017). The use of mixed populations of *Saccharomyces cerevisiae* and *S. kudriavzevii* to reduce ethanol content in wine: limited aeration, inoculum proportions, and sequential inoculation. *Front. Microbiol.* 8:2087. doi: 10.3389/fmicb.2017.02087
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410. doi: 10.1016/S0022-2836(05)80360-2
- Arroyo-López, F. N., Orlic, S., Querol, A., and Barrio, E. (2009). Effects of temperature, pH and sugar concentration on the growth parameters of *Saccharomyces cerevisiae*, *S. kudriavzevii* and their interspecific hybrids. *Int. J. Food Microbiol.* 131, 120–127. doi: 10.1016/j.ijfoodmicro.2009.01.035
- Arroyo-López, F. N., Pérez-Torrado, R., Querol, A., and Barrio, E. (2010a). Modulation of the glycerol and ethanol syntheses in the yeast *Saccharomyces kudriavzevii* differs from that exhibited by *S. cerevisiae* and their hybrid. *Food Microbiol.* 27, 628–637. doi: 10.1016/j.fm.2010.02.001
- Arroyo-López, F. N., Salvadó, Z., Tronchoni, J., Guillamón, J. M., Barrio, E., and Querol, A. (2010b). Susceptibility and resistance to ethanol in *Saccharomyces* strains isolated from wild and fermentative environments. *Yeast* 27:1015. doi: 10.1002/yea.1809
- Belloch, C., Orlic, S., Barrio, E., and Querol, A. (2008). Fermentative stress adaptation of hybrids within the *Saccharomyces sensu stricto* complex. *Int. J. Food Microbiol.* 122, 188–195. doi: 10.1016/j.ijfoodmicro.2007.11.083
- Boynton, P. J., and Greig, D. (2014). The ecology and evolution of non-domesticated *Saccharomyces* species. *Yeast* 31, 449–462. doi: 10.1002/yea.3040
- Brandman, O., Stewart-Ornstein, J., Wong, D., Larson, A., Williams, C. C., Li, G.-W., et al. (2012). A ribosome-bound quality control complex triggers degradation of nascent peptides and signals translation stress. *Cell* 151, 1042–1054. doi: 10.1016/j.cell.2012.10.044
- Burri, L., and Lithgow, T. (2004). A complete set of SNAREs in yeast. *Traffic* 5, 45–52. doi: 10.1046/j.1600-0854.2003.00151.x
- Byrne, K. P., and Wolfe, K. H. (2005). The Yeast gene order browser: combining curated homology and syntenic context reveals gene fate in polyploid species. *Genome Res.* 15, 1456–1461. doi: 10.1101/gr.3672305
- Cowart, L. A., and Obeid, L. M. (2007). Yeast sphingolipids: recent developments in understanding biosynthesis, regulation, and function. *Biochim. Biophys. Acta* 1771, 421–431. doi: 10.1016/j.bbalip.2006.08.005
- Dayarian, A., Michael, T. P., and Sengupta, A. M. (2010). SOPRA: Scaffolding algorithm for paired reads via statistical optimization. *BMC Bioinformatics* 11:345. doi: 10.1186/1471-2105-11-345
- Erny, C., Raoult, P., Alais, A., Butterlin, G., Delobel, P., Matei-Radoi, F., et al. (2012). Ecological success of a group of *Saccharomyces cerevisiae*/*Saccharomyces kudriavzevii* hybrids in the northern european wine making environment. *Appl. Environ. Microbiol.* 78, 3256–3265. doi: 10.1128/AEM.06752-11
- Filteau, M., Charron, G., and Landry, C. R. (2016). Identification of the fitness determinants of budding yeast on a natural substrate. *Isme J.* 11:959. doi: 10.1038/ismej.2016.170
- Gallone, B., Steensels, J., Prahl, T., Soriaga, L., Saels, V., Herrera-Malaver, B., et al. (2016). Domestication and divergence of *Saccharomyces cerevisiae* beer yeasts. *Cell* 166, 1397.e16–1410.e16. doi: 10.1016/j.cell.2016.08.020
- Gamero, A., Belloch, C., and Querol, A. (2015). Genomic and transcriptomic analysis of aroma synthesis in two hybrids between *Saccharomyces cerevisiae* and *S. kudriavzevii* in winemaking conditions. *Microb. Cell Fact.* 14:128. doi: 10.1186/s12934-015-0314-5
- Gangl, H., Batusic, M., Tschek, G., Tiefenbrunner, W., Hack, C., and Lopandic, K. (2009). Exceptional fermentation characteristics of natural hybrids from *Saccharomyces cerevisiae*, *S. kudriavzevii*. *New Biotechnol.* 25, 244–251. doi: 10.1016/j.nbt.2008.10.001
- Ghisla, S., and Massey, V. (1989). Mechanisms of flavoprotein-catalyzed reactions. *Eur. J. Biochem.* 181, 1–17. doi: 10.1111/j.1432-1033.1989.tb14688.x



- Goddard, M. R., and Greig, D. (2015). *Saccharomyces cerevisiae*: a nomadic yeast with no niche? *FEMS Yeast Res.* 15:fov009. doi: 10.1093/femsyr/fov009
- Goffeau, A., Barrell, B. G., Bussey, H., Davis, R. W., Dujon, B., Feldmann, H., et al. (1996). Life with 6000 genes. *Science* 274, 546–567. doi: 10.1126/science.274.5287.546
- Gonçalves, P., Valério, E., Correia, C., De Almeida, J. M. G. C., and Sampaio, J. P. (2011). Evidence for divergent evolution of growth temperature preference in sympatric *Saccharomyces* species. *PLoS One* 6:e20739. doi: 10.1371/journal.pone.0020739
- González, S. S., Barrio, E., Gafner, J., and Querol, A. (2006). Natural hybrids from *Saccharomyces cerevisiae*, *Saccharomyces bayanus* and *Saccharomyces kudriavzevii* in wine fermentations. *FEMS Yeast Res.* 6, 1221–1234. doi: 10.1111/j.1567-1364.2006.00126.x
- González, S. S., Gallo, L., Climent, M. D., Barrio, E., and Querol, A. (2007). Enological characterization of natural hybrids from *Saccharomyces cerevisiae* and *S. kudriavzevii*. *Int. J. Food Microbiol.* 116, 11–18. doi: 10.1016/j.ijfoodmicro.2006.10.047
- Gordon, D., and Green, P. (2013). Consed: a graphical editor for next-generation sequencing. *Bioinformatics* 29, 2936–2937. doi: 10.1093/bioinformatics/btt515
- Gordon, J. L., Armisen, D., Proux-Wéra, E., Óhéigeartaigh, S. S., Byrne, K. P., and Wolfe, K. H. (2011). Evolutionary erosion of yeast sex chromosomes by mating-type switching accidents. *Proc. Natl. Acad. Sci. U. S. A.* 108, 20024–20029. doi: 10.1073/pnas.1112808108
- Gough, J., Karplus, K., Hughey, R., and Chothia, C. (2001). Assignment of homology to genome sequences using a library of hidden markov models that represent all proteins of known structure. Edited by G. Von Heijne. *J. Mol. Biol.* 313, 903–919. doi: 10.1006/jmbi.2001.5080
- Grantham, R. (1974). Amino acid difference formula to help explain protein evolution. *Science* 185, 862–864. doi: 10.1126/science.185.4154.862
- Grassi, L., Fusco, D., Sellerio, A., Corà, D., Bassetti, B., Caselle, M., et al. (2010). Identity and divergence of protein domain architectures after the yeast whole-genome duplication event. *Mol. Biosyst.* 6, 2305–2315. doi: 10.1039/C003507F
- Gresham, D., Boer, V. M., Caudy, A., Ziv, N., Brandt, N. J., Storey, J. D., et al. (2011). System-level analysis of genes and functions affecting survival during nutrient starvation in *Saccharomyces cerevisiae*. *Genetics* 187, 299–317. doi: 10.1534/genetics.110.120766
- Heath, V. L., Shaw, S. L., Roy, S., and Cyert, M. S. (2004). Hph1p and Hph2p, novel components of calcineurin-mediated stress responses in *Saccharomyces cerevisiae*. *Eukaryot. Cell* 3, 695–704. doi: 10.1128/ec.3.3.695-704.2004
- Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. doi: 10.1093/molbev/mst010
- Kimura, M. (1983). *The Neutral Theory of Molecular Evolution*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511623486
- Künzler, M., Paravicini, G., Egli, C. M., Irniger, S., and Braus, G. H. (1992). Cloning, primary structure and regulation of the ARO4 gene, encoding the tyrosine-inhibited 3-deoxy-D-arabino-heptulosonate-7-phosphate synthase from *Saccharomyces cerevisiae*. *Gene* 113, 67–74. doi: 10.1016/0378-1119(92)90670-K
- Kurtz, S., Phillippy, A., Delcher, A. L., Smoot, M., Shumway, M., Antonescu, C., et al. (2004). Versatile and open software for comparing large genomes. *Genome Biol.* 5:R12. doi: 10.1186/gb-2004-5-2-r12
- Liti, G., Carter, D. M., Moses, A. M., Warringer, J., Parts, L., James, S. A., et al. (2009). Population genomics of domestic and wild yeasts. *Nature* 458, 337–341. doi: 10.1038/nature07743
- Lopes, C. A., Barrio, E., and Querol, A. (2010). Natural hybrids of *Saccharomyces cerevisiae* x *Saccharomyces kudriavzevii* share alleles with European wild populations of *S. kudriavzevii*. *FEMS Yeast Res.* 10, 412–421. doi: 10.1002/yea.1493
- López-Malo, M., Querol, A., and Guillamón, J. M. (2013). Metabolomic comparison of *Saccharomyces cerevisiae* and the cryotolerant species *S. bayanus* var. *uvarum* and *S. kudriavzevii* during wine fermentation at low temperature. *PLoS One* 8:e60135. doi: 10.1371/journal.pone.0060135
- Lynch, M. (2007). *The Origins of Genome Architecture*. Sunderland, MA: Sinauer Associates, Inc.
- Morales, L., and Dujon, B. (2012). Evolutionary role of interspecies hybridization and genetic exchanges in yeasts. *Microbiol. Mol. Biol. Rev.* 76, 721–739. doi: 10.1128/MMBR.00022-12
- Morard, M., Macías, L. G., Adam, A. C., Lairón-Peris, M., Pérez-Torrado, R., Toft, C., et al. (2019). Aneuploidy and ethanol tolerance in *Saccharomyces cerevisiae*. *Front. Genet.* 10:82. doi: 10.3389/FGENE.2019.00082
- Naseeb, S., James, S. A., Alsammar, H., Michaels, C. J., Gini, B., Nueno-Palop, C., et al. (2017). *Saccharomyces jurei* sp. nov., isolation and genetic identification of a novel yeast species from *Quercus robur*. *Int. J. Syst. Evol. Microbiol.* 67, 2046–2052. doi: 10.1099/ijsem.0.002013
- Naumov, G. I., James, S. A., Naumova, E. S., Louis, E. J., and Roberts, I. N. (2000). Three new species in the *Saccharomyces sensu stricto* complex: *Saccharomyces cariocanus*, *Saccharomyces kudriavzevii* and *Saccharomyces mikatae*. *Int. J. Syst. Evol. Microbiol.* 50, 1931–1942. doi: 10.1099/00207713-50-5-1931
- Naumov, G. I., Lee, C. F., and Naumova, E. S. (2013). Molecular genetic diversity of the *Saccharomyces* yeasts in taiwan: *Saccharomyces arboricola*, *Saccharomyces cerevisiae* and *Saccharomyces kudriavzevii*. *Antonie van Leeuwenhoek* 103, 217–228. doi: 10.1007/s10482-012-9803-2
- Oliveira, B. M., Barrio, E., Querol, A., and Pérez-Torrado, R. (2014). Enhanced enzymatic activity of glycerol-3-phosphate dehydrogenase from the cryophilic *Saccharomyces kudriavzevii*. *PLoS One* 9:e87290. doi: 10.1371/journal.pone.0087290
- Otto, T. D., Dillon, G. P., Degraeve, W. S., and Berriman, M. (2011). RATT: rapid annotation transfer tool. *Nucleic Acids Res.* 39:e57. doi: 10.1093/nar/gkq1268
- Paget, C. M., Schwartz, J. M., and Delneri, D. (2014). Environmental systems biology of cold-tolerant phenotype in *Saccharomyces* species adapted to grow at different temperatures. *Mol. Ecol.* 23, 5241–5257. doi: 10.1111/mec.12930
- Pérez-Torrado, R., Barrio, E., and Querol, A. (2018). Alternative yeasts for winemaking: *Saccharomyces non-cerevisiae* and its hybrids. *Crit. Rev. Food Sci. Nutr.* 58, 1780–1790. doi: 10.1080/10408398.2017.1285751
- Pérez-Través, L., Lopes, C. A., Querol, A., and Barrio, E. (2014). On the complexity of the *Saccharomyces bayanus* taxon: hybridization and potential hybrid speciation. *PLoS One* 9:e93729. doi: 10.1371/journal.pone.0093729
- Peris, D., Pérez-Torrado, R., Hittinger, C. T., Barrio, E., and Querol, A. (2018). On the origins and industrial applications of *Saccharomyces cerevisiae* x *Saccharomyces kudriavzevii* hybrids. *Yeast* 35, 51–69. doi: 10.1002/yea.3283
- Peter, J., De Chiara, M., Friedrich, A., Yue, J.-X., Pflieger, D., Bergström, A., et al. (2018). Genome evolution across 1,011 *Saccharomyces cerevisiae* isolates. *Nature* 556, 339–344. doi: 10.1038/s41586-018-0030-5
- Rodríguez, M. E., Pérez-Través, L., Sangorrín, M. P., Barrio, E., Querol, A., and Lopes, C. A. (2017). *Saccharomyces uvarum* is responsible for the traditional fermentation of apple chicha in patagonia. *FEMS Yeast Res.* 17:fow109. doi: 10.1093/femsyr/fow109
- Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M.-A., et al. (2000). Artemis: sequence visualization and annotation. *Bioinformatics* 16, 944–945. doi: 10.1093/bioinformatics/16.10.944
- Salvadó, Z., Arroyo-López, F. N., Guillamón, J. M., Salazar, G., Querol, A., and Barrio, E. (2011). Temperature adaptation markedly determines evolution within the genus *Saccharomyces*. *Appl. Environ. Microbiol.* 77, 2292–2302. doi: 10.1128/aem.01861-10
- Sampaio, J. P., and Gonçalves, P. (2008). Natural populations of *Saccharomyces kudriavzevii* in Portugal are associated with oak bark and sympatric with *S. cerevisiae* and *S. paradoxus*. *Appl. Environ. Microbiol.* 74, 2144–2152. doi: 10.1128/AEM.02396-07
- Scannell, D. R., Frank, A. C., Conant, G. C., Byrne, K. P., Woolfit, M., and Wolfe, K. H. (2007). Independent sorting-out of thousands of duplicated gene pairs in two yeast species descended from a whole-genome duplication. *Proc. Natl. Acad. Sci. U S A.* 104, 8397–8402. doi: 10.1073/pnas.0608218104
- Scannell, D. R., Zill, O. A., Rokas, A., Payen, C., Dunham, M. J., Eisen, M. B., et al. (2011). The awesome power of yeast evolutionary genetics: new genome sequences and strain resources for the *Saccharomyces sensu stricto* genus. *G3 Genes Genom. Genet.* 1, 11–25. doi: 10.1534/g3.111.000273
- Schwelberger, H. G., Kohlwein, S. D., and Paltauf, F. (1989). Molecular cloning, primary structure and disruption of the structural gene of aldolase from *Saccharomyces cerevisiae*. *Eur. J. Biochem.* 180, 301–308. doi: 10.1111/j.1432-1033.1989.tb14648.x

- Shen, P. S., Park, J., Qin, Y., Li, X., Parsawar, K., Larson, M. H., et al. (2015). Rqc2p and 60S ribosomal subunits mediate mRNA-independent elongation of nascent chains. *Science* 347, 75–78. doi: 10.1126/science.1259724
- Stanke, M., and Morgenstern, B. (2005). AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Res.* 33, W465–W467. doi: 10.1093/nar/gki458
- Stribny, J., Gamero, A., Perez-Torrado, R., and Querol, A. (2015). *Saccharomyces kudriavzevii* and *Saccharomyces uvarum* differ from *Saccharomyces cerevisiae* during the production of aroma-active higher alcohols and acetate esters using their amino acidic precursors. *Int. J. Food Microbiol.* 205, 41–46. doi: 10.1016/j.ijfoodmicro.2015.04.003
- Sym, M., Engebrecht, J., and Roeder, G. S. (1993). ZIP1 is a synaptonemal complex protein required for meiotic chromosome synapsis. *Cell* 72, 365–378. doi: 10.1016/0092-8674(93)90114-6
- Tajima, F. (1993). Simple methods for testing the molecular evolutionary clock hypothesis. *Genetics* 135, 599–607.
- Toft, C., Williams, T. A., and Fares, M. A. (2009). Genome-wide functional divergence after the symbiosis of *Proteobacteria* with insects unraveled through a novel computational approach. *PLoS Comp. Biol.* 5:e1000344. doi: 10.1371/journal.pcbi.1000344
- Tronchoni, J., Medina, V., Guillamón, J. M., Querol, A., and Pérez-Torrado, R. (2014). Transcriptomics of cryophilic *Saccharomyces kudriavzevii* reveals the key role of gene translation efficiency in cold stress adaptations. *BMC Genomics* 15:432. doi: 10.1186/1471-2164-15-432
- Tronchoni, J., Rozès, N., Querol, A., and Guillamón, J. M. (2012). Lipid composition of wine strains of *Saccharomyces kudriavzevii* and *Saccharomyces cerevisiae* grown at low temperature. *Int. J. Food Microbiol.* 155, 191–198. doi: 10.1016/j.ijfoodmicro.2012.02.004
- Wagner, A. (2008). Neutralism and selectionism: a network-based reconciliation. *Nat. Rev. Genet.* 9:965. doi: 10.1038/nrg2473
- Wolfe, K. H., and Shields, D. C. (1997). Molecular evidence for an ancient duplication of the entire yeast genome. *Nature* 387, 708–713. doi: 10.1038/42711
- Wong, S., and Wolfe, K. H. (2005). Birth of a metabolic gene cluster in yeast by adaptive gene relocation. *Nat. Genet.* 37, 777–782. doi: 10.1038/ng1584
- Yang, Z. (2007). PAML 4: a program package for phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* 24, 1586–1591. doi: 10.1093/molbev/msm088
- Yang, Z., Wong, W. S. W., and Nielsen, R. (2005). Bayes empirical bayes inference of amino acid sites under positive selection. *Mol. Biol. Evol.* 22, 1107–1118. doi: 10.1093/molbev/msi097
- Yoo, H. S., Genbauffe, F. S., and Cooper, T. G. (1985). Identification of the ureidoglycolate hydrolase gene in the *DAL* gene cluster of *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* 5, 2279–2288. doi: 10.1128/mcb.5.9.2279
- Zerbino, D. R., and Birney, E. (2008). Velvet: Algorithms for de novo short read assembly using de bruijn graphs. *Genome Res.* 18, 821–829. doi: 10.1101/gr.074492.107

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Macías, Morard, Toft and Barrio. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# A Quasi-Domesticated Relic Hybrid Population of *Saccharomyces cerevisiae* × *S. paradoxus* Adapted to Olive Brine

Ana Pontes<sup>1</sup>, Neža Čadež<sup>2</sup>, Paula Gonçalves<sup>1</sup> and José Paulo Sampaio<sup>1\*</sup>

<sup>1</sup> UCIBIO-REQUIMTE, Departamento de Ciências da Vida, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Caparica, Portugal, <sup>2</sup> Biotechnical Faculty, University of Ljubljana, Ljubljana, Slovenia

## OPEN ACCESS

### Edited by:

Jean Marie François,  
UMR5504 Laboratoire d'Ingénierie  
des Systèmes Biologiques et des  
Procédés (LISBP), France

### Reviewed by:

Jordi Tronchoni,  
Instituto de Ciencias de la Vid y del  
Vino (ICVV), Spain  
Alexander DeLuna,  
Centro de Investigación y de Estudios  
Avanzados (CINVESTAV), Mexico

### \*Correspondence:

José Paulo Sampaio  
jss@fct.unl.pt

### Specialty section:

This article was submitted to  
Evolutionary and Genomic  
Microbiology,  
a section of the journal  
Frontiers in Genetics

**Received:** 12 January 2019

**Accepted:** 30 April 2019

**Published:** 29 May 2019

### Citation:

Pontes A, Čadež N, Gonçalves P  
and Sampaio JP (2019) A  
Quasi-Domesticated Relic Hybrid  
Population of *Saccharomyces*  
*cerevisiae* × *S. paradoxus* Adapted  
to Olive Brine. *Front. Genet.* 10:449.  
doi: 10.3389/fgene.2019.00449

The adaptation of the yeast *Saccharomyces cerevisiae* to man-made environments for the fermentation of foodstuffs and beverages illustrates the scientific, social, and economic relevance of microbe domestication. Here we address a yet unexplored aspect of *S. cerevisiae* domestication, that of the emergence of lineages harboring some domestication signatures but that do not fit completely in the archetype of a domesticated yeast, by studying *S. cerevisiae* strains associated with processed olives, namely table olives, olive brine, olive oil, and alpechin. We confirmed earlier observations that reported that the Olives population results from a hybridization between *S. cerevisiae* and *S. paradoxus*. We concluded that the olive hybrids form a monophyletic lineage and that the *S. cerevisiae* progenitor belonged to the wine population of this species. We propose that homoploid hybridization gave rise to a diploid hybrid genome, which subsequently underwent the loss of most of the *S. paradoxus* sub-genome. Such a massive loss of heterozygosity was probably driven by adaptation to the new niche. We observed that olive strains are more fit to grow and survive in olive brine than control *S. cerevisiae* wine strains and that they appear to be adapted to cope with the presence of NaCl in olive brine through expansion of copy number of *ENA* genes. We also investigated whether the *S. paradoxus* HXT alleles retained by the Olives population were likely to contribute to the observed superior ability of these strains to consume sugars in brine. Our experiments indicate that sugar consumption profiles in the presence of NaCl are different between members of the Olives and Wine populations and only when cells are cultivated in nutritional conditions that support adaptation of their proteome to the high salt environment, which suggests that the observed differences are due to a better overall fitness of olives strains in the presence of high NaCl concentrations. Although relic olive hybrids exhibit several characteristics of a domesticated lineage, tangible benefits to humans cannot be associated with their phenotypes. These strains can be seen as a case of adaptation without positive or negative consequences to humans, that we define as a quasi-domestication.

**Keywords:** yeast, *Saccharomyces cerevisiae*, hybridization, microbe population genomics, microbiology of olive brine

## INTRODUCTION

The domestication of plants and animals was a major revolution in human history because it drove the emergence civilizations with the associated demographic and technological consequences that last until today (Diamond, 1997). In many instances domestication represents a dramatic case of adaptive divergence in response to human selection (Doebley et al., 2006; Ross-Ibarra et al., 2007). Domestication consists in the selective and controlled propagation an organism that genetically acquires modifications that not only distinguish it from its wild ancestors, but also make it more useful to humans (Diamond, 2002). Since archeological and biomolecular evidence indicates that fermented beverages reminiscent of rice wine were produced as far back as 9,000 years ago in China (McGovern, 2009) and the forebear of modern beer was consumed 8,000 years ago in Sumeria (Hornsey, 2003), the case of microbe domestication has a context and a time scale comparable to the much better understood aspects of plant and animal domestication. In fact, the mechanisms and consequences of artificial selection of microbes such as yeasts, carried out in most cases unconsciously by innumerable generations of brewers, winemakers and other artisans, are starting to be understood (Almeida et al., 2015, 2017; Gallone et al., 2016; Gonçalves et al., 2016; Barbosa et al., 2018; Duan et al., 2018; Legras et al., 2018; Peter et al., 2018). Although the mechanisms that gave rise to the phenotypes of domesticated strains are currently the focus of intense scientific inquiry, the detailed comprehension of the multiple transformations that gave rise to domesticated yeast lineages is far from been achieved.

Besides the obvious cases of the emergence of wine, beer or sake variants, several other presumably domesticated lineages of *Saccharomyces cerevisiae* have also been recently revealed (Duan et al., 2018; Preiss et al., 2018). Moreover, the proximity to humans appears to have elicited the emergence of a new niche to which *S. cerevisiae* is adapting to, as the isolation of commensal (Angebaunt et al., 2013) and opportunistic (Enache-Angoulvant and Hennequin, 2005; Munoz et al., 2005) strains suggests. Recently we and others have explored the domestication space of *S. cerevisiae* (Almeida et al., 2015; Gallone et al., 2016; Gonçalves et al., 2016; Barbosa et al., 2018; Legras et al., 2018) and one of the main findings is that the domestication routes of this yeast are multiple and independent, and most remain poorly known. The picture that is gradually emerging depicts a complex population structure rich of different wild populations, most of them showing geographical partitioning, and numerous domesticated populations, each associated with a given fermented product and showing specific adaptations related to the type of fermentation in which they participate (Duan et al., 2018; Legras et al., 2018; Peter et al., 2018). This complex scenario is further complicated by the occurrence of admixture between certain populations that gives rise to mosaic genotypes (Tilakaratna and Bensasson, 2017) and to transitions from primary to secondary domestications (Barbosa et al., 2018). Here we address a yet unexplored aspect of *S. cerevisiae* domestication – that of the emergence of lineages harboring some domestication signatures but that do not fit completely in the definition of a domesticated yeast because artificial

selection, even if unintentional, is not easy to accommodate with the emergence of a phenotype that provides identifiable benefits to humans. Specifically we survey a set of *S. cerevisiae* strains associated with the maturation of table olives, where they occur spontaneously without propagation from one batch to the other. Olive brine strains were first studied by Cromie et al. (2013) using restriction site-associated sequencing (RAD-seq), a reduced genome sequencing strategy. These authors detected a small group of *S. cerevisiae* strains isolated from European olives that clustered next to wine strains and were defined by a unique set of sequence variants not present in other populations. Subsequently, Strobe et al. (2015), in a genomics survey of 100 *S. cerevisiae* strains, found that a restricted group of three strains that included YJM 1252 (=PYCC 6732 = CBS 3081), isolated from alpechin (olive mill wastes) in Spain had a considerable number of ORFs (>200) from *S. paradoxus*. The other two strains were YJM 1078 (NRRL YB-4348 = PYCC 8028) and YJM 248 (NRRL Y-12659 = CBS 2910 = PYCC 8034) isolated in Portugal in the 1950's from human feces. In a more recent study involving the genome analyses of more than 1000 *S. cerevisiae* strains, Peter et al. (2018) identified 17 strains in a so-called alpechin clade sharing the already reported *S. paradoxus* contribution. Because the strains originating from the olive niche have never been studied separately but rather on comprehensive surveys that included hundreds or thousands of strains from other provenances, thus precluding their detailed analysis, we carried out an investigation on the ecological, physiological and genomic particularities of olive strains aiming at understanding their origins and specific adaptations within the global framework of yeast domestication. We found that olive strains are more fit to grow and survive in olive brine than control *S. cerevisiae* wine strains and that they appear to be adapted to cope with the presence of NaCl in olive brine. Moreover, the ecological range of these strains includes the processed olives niche but not olive trees or olives in natural conditions. We postulate that an ancient hybridization between a *S. cerevisiae* wine strain and *S. paradoxus*, provided the genetic diversity that allowed the adaptation to the new niche and that this process was accompanied by the adaptive loss of most the non-*cerevisiae* sub-genome.

## MATERIALS AND METHODS

### Yeast Isolation, Identification, and Crosses

The isolation of *Saccharomyces* strains was conducted using a selective enrichment protocol previously described (Sampaio and Gonçalves, 2008). For the strains isolated in Slovenia, samples were directly used for yeast isolation without enrichment. Preliminary species-level identifications were performed by sequencing the D1/D2 region of the 26S rDNA. Crosses involved ascospore micro-manipulation. Positive crosses between two parental strains were confirmed by sequencing of *GAL1* and confirmation of the expected heterozygous sites. For each cross, interspecific spore viability was determined by examining 200 ascospores.



## Olive Brine Medium

Olives of the Oblica variety (approx. 200 g) collected in Évora, Portugal were used to prepare olive brine (200ml H<sub>2</sub>O, 8% NaCl w/v) during 3 months at 17°C. After this period, the brine was sterilized by filtration and kept at 4°C. This brine was analyzed by HPLC to identify and quantify the sugars and sugar-like compounds present. Sugar concentrations were determined using a carbohydrate analysis column (250 mm × 4 mm + Aminotrap, Dionex Carbopac PA10; DIONEX ICS3000). The column was kept at 25°C and 0.1M NaOH was used as the mobile phase at 1 ml min<sup>-1</sup>. For the phenolic compounds the concentrations (%) were determined using a Waters Novapack C18 15 mm column (DIONEX ICS3000). The column was kept at 30°C and 2% methanol was used as the mobile phase at 0.5 ml min<sup>-1</sup>.

## Absolute Fitness in Olive Brine

Two experiments were done independently for the group of strains selected. Pre-cultures (20 ml in 50 ml flasks) grown for 24 h in YNB (Yeast Nitrogen Base, Difco) supplemented with 1% (w/v) glucose (incubation at 25°C) were used to inoculate approximately  $1 \times 10^5$  cells/ml in a volume of 2 ml of olive brine in 2 ml micro-centrifuge tubes. Cells were grown in batch cultures for 70 days at 17°C without shaking, and cell viability was estimated by performing regular plate counts after preliminary counts in a hemocytometer. Statistical significance was tested using an unpaired *t*-test with Welch's correction, implemented in GraphPad Prism v5 (*p*-value cut-off <0.01). At the end of the experiment, the supernatants of two cultures from the Olives population and two cultures from the Wine population were randomly chosen for HPLC analysis to determine the residual concentrations of sugars.

## Growth Rates in NaCl

Strains were pre-grown overnight in liquid YPD medium [yeast extract 1% (w/v), peptone 2% (w/v), D-glucose 2% (w/v), at 25°C and were subsequently transferred to fresh medium (20 ml YPD or YPD supplemented with NaCl 6% or 8% w/v) and incubated with orbital shaking (180 r.p.m.) at 30°C in 50 ml flasks]. The initial OD<sub>640nm</sub> was 0.1 – 0.2. Growth rates were calculated in the exponential phase using OD<sub>640nm</sub> measurements.

## Sugar Consumption in Phosphate Buffer

Strains were pre-grown overnight at 25°C in YNB medium supplemented with 1% (w/v) glucose. Cells were then transferred to phosphate buffer at pH 5.8 (30 ml in 50 ml flasks) supplemented with 0.6% (w/v) glucose 0.1% (w/v), fructose, and 8% (w/v) NaCl, to mimic the conditions in olive brine and incubated at 20°C. A similar experiment was also conducted in the absence of NaCl. Sugar consumption was monitored for 10 days by HPLC. Extracellular concentrations of fructose and glucose were determined using a carbohydrate analysis column (300 mm × 7 mm, Thermo HyperREZ XP Carbohydrate Ca++; KNAUER Smartline) and a differential refractometer. The column

was kept at 85°C and H<sub>2</sub>O was used as a mobile phase at 0.6 ml min<sup>-1</sup>.

## Genome Sequencing, Read Alignment, and Genotype Calling

DNA was extracted from overnight grown cultures of monosporic or single-cell derivatives and paired-end Illumina MiSeq 250 bp genomic reads were obtained after sequencing for 500 cycles. Genomic data for other strains was obtained from the NCBI-SRA archive and from the *Saccharomyces* Genome Resequencing Project v2 (SGRP2) (Bergström et al., 2014). When only finished genome sequences were available in public databases (NCBI), the corresponding error-free Illumina reads were simulated using dwgsm<sup>1</sup>. Reads for each isolate were mapped to the *S. cerevisiae* reference genome (UCSC version sacCer3) using SMALT v0.7.5 aligner<sup>2</sup>. The reference Index was built with a word length of 13 and a sampling step size of 2 (−k 13 −s 2). An intensive search for alignments (−x) was performed during the mapping step with the random assignment of ambiguous alignments switched off (−r −1) and the base quality threshold for the look-up of the hash index set to 10 (−q 10). With these settings, SMALT v0.7.5 only reports the best unique gapped alignment for each read. For the paired-end data the insert size distribution was inferred with the “sample” command of SMALT prior to mapping. Conversion of SAM format to BAM, sorting, indexing, several mapping statistics, and consensus genotype calling were performed using the tools available in the SAMtools package v1.18 (Li et al., 2009) and as described previously (Almeida et al., 2014). Multiple sequence alignments for each reference chromosome were generated from the resulting fasta files. For downstream analysis, all bases with Phred quality score below Q40 (equivalent to a 99.99% base call accuracy) or ambiguous base calls were converted to “N.” For obtaining the *S. cerevisiae* and *S. paradoxus* sub-genomes of the hybrid strains, reads for each strain were mapped to an extended *Saccharomyces* spp. reference with assembled sequences from the genomes for *S. cerevisiae* (UCSC version sacCer3), *S. paradoxus*, *S. mikatae*, *S. kudriavzevii*, *S. uvarum* (Scannell et al., 2011), and *S. arboricolus* (Liti et al., 2013).

## Phylogenetic Inference and Divergence Across the Genome

Chromosomal single nucleotide polymorphisms (SNPs) were extracted from multiple sequence alignments only if the evaluated site was represented by unambiguous high confidence alleles in at least 85% of the strains. SNPs were then concatenated to generate a whole-genome SNP alignment. The phylogeny was inferred using maximum likelihood as implemented in IQ-TREE v 1.6.7 (Nguyen et al., 2015) using an empirically determined substitution model, SH-like approximate likelihood ratio test (1000 replications) (Guindon et al., 2010), and rooted with *S. paradoxus*. The phylogeny

<sup>1</sup><https://sourceforge.net/projects/dnaa/files/dwgsm/>

<sup>2</sup><http://www.sanger.ac.uk/resources/software/smalt/>

was visualized using ITOL, version 3.0 (Letunic and Bork, 2016). Whole-genome levels of divergence were estimated using Variscan v2.0 (Hutter et al., 2006). Divergence was calculated for each mapped strain in comparison with the reference genome of *S. cerevisiae* using RunMode 21. The results were processed using a 10 kb sliding window with 10 kb step increments.

## Screening for Non-*S. cerevisiae* Genes

Evidence of the presence of genes from other *Saccharomyces* species was investigated by mapping the reads to a combined reference that includes the annotated coding sequences of *S. arboricola*, *S. cerevisiae*, *S. kudriavzevii*, *S. mikatae*, *S. paradoxus*, and *S. uvarum* (Scannell et al., 2011; Liti et al., 2013). Reads were mapped to this combined reference using BWA V0.6.2 (Li and Durbin, 2009) with default parameters, but setting the quality threshold to 10 ( $-q\ 10$ ). SAMtools V1.1852 (Li et al., 2009) was then used for manipulation of the resulting BAM files. Only ORFs with orthologs unambiguously annotated in all the species were analyzed. An ORF was considered to have a foreign origin to *S. cerevisiae* if its coverage was higher at least one-fourth of the median whole genome coverage for a given strain. The ORF coverage was defined as the product of the total number of mapped reads to a given ORF by the read length, dividing by the sum of all the ORFs length (considering only ORFs that have at least 25% of reads mapped to, when comparing to the orthologous ORF with the highest number of reads). This measure was taken to control spurious alignment counts. The coverage threshold allowed for some heterogeneity in the read counts and for the eventual presence of a foreign ORF together with the native *S. cerevisiae* ORF. For some of the *S. paradoxus* genes detected in the hybrid genomes, their assignment to this species was confirmed with phylogenetic analyses involving homologous sequences from other *Saccharomyces* species.

## Gene Ontology Analyses and Survey of Specific Genes and of Gene Copy Number Variation

Standard gene ontology (GO) term find was performed with the GO TERM FINDER tool v0.83, available at SGD, using a  $p$ -value cut-off of  $<0.01$ . We performed *de novo* genome assemblies using SPAdes v3.11.1. Prior to assembly, reads were processed with Trimmomatic v0.36 based on quality score threshold of 20 for windowed trimming, discarding reads less than 100 bp in length or harboring ambiguities. To retrieve genes of interest, a local BLAST database was set up for each genome. Copy number variation of the two *CUP1* genes (*CUP1-1* and *CUP1-2*) and the three *ENA* genes (*ENA1*, *ENA2*, and *ENA5*) was investigated using CNVnator (Abyzov et al., 2011) on mapped genomes and using *ACT1* as control. The query sequences were defined by the coordinates in the reference sequence of *S. cerevisiae* for the coding regions of the genes of interest. The results obtained were manually validated by checking the chromosomal context of the hits using UGENE (Okonechnikov et al., 2012)

and by analyzing the copy number of the genes flanking the genes of interest.

## Data Availability

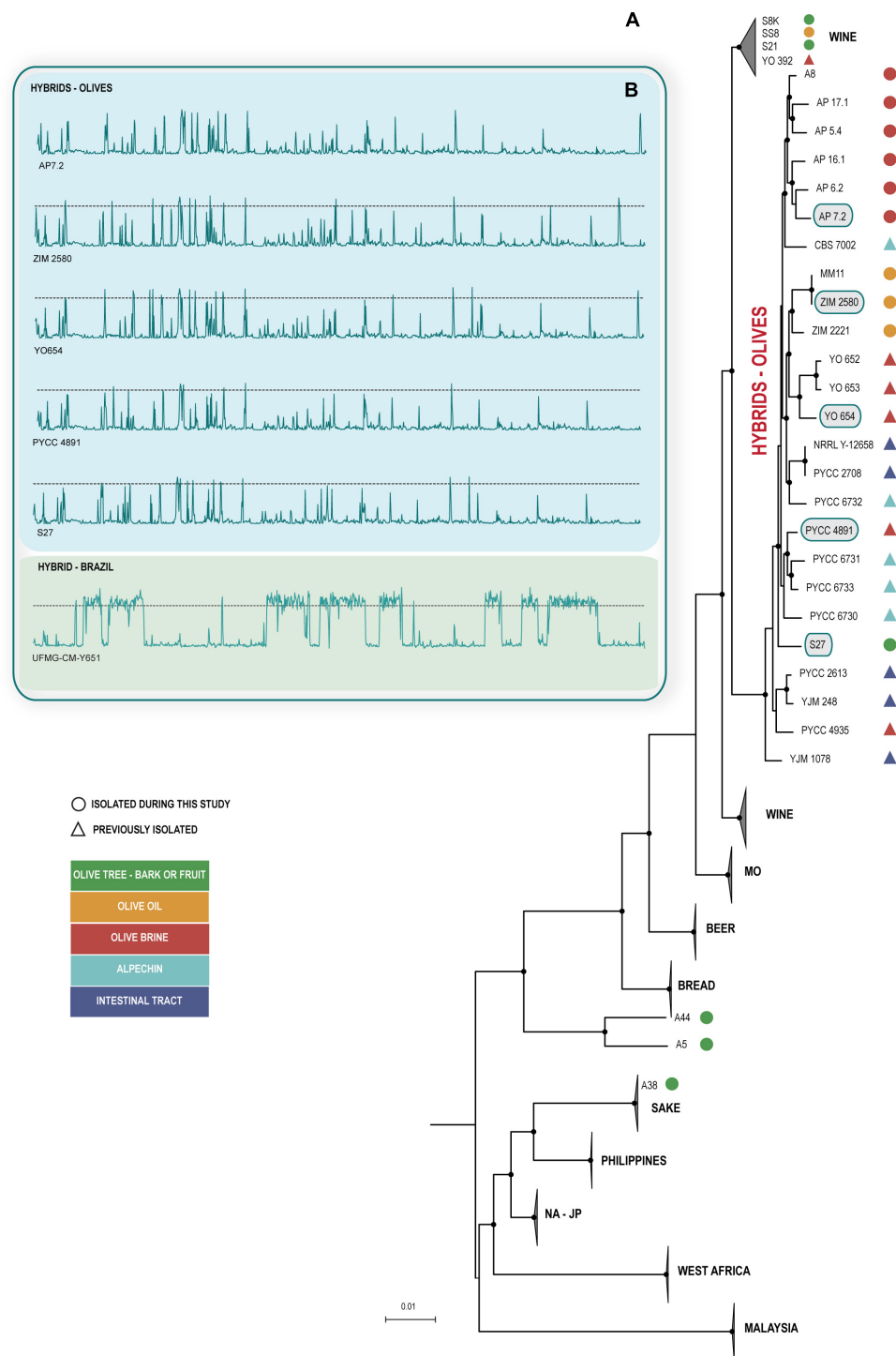
Genome sequence data have been deposited in the European Nucleotide Archive (ENA) database under the accession code PRJEB30431.

## RESULTS

### Ecology and Phylogeny

Given earlier reports on the occurrence of *S. cerevisiae* in association with table olives (Arroyo-López et al., 2008; Cromie et al., 2013; Bonatsou et al., 2018) and with alpechin (Santa María, 1958), we asked if the original source of these yeasts was the olive tree itself. For this reason we conducted an isolation program employing samples of olive tree bark, leaves, fruits and soil underneath the trees, and a selective enrichment protocol for yeasts of the genus *Saccharomyces*. In parallel, processed products such as olive oil and olive brine from table olives were also investigated. In total 163 samples from olive trees were investigated, together with 53 samples from olive oil and 7 samples from olive brine. Although the number of samples collected from olive trees was much higher, the frequency of isolation of *Saccharomyces* spp. was very low (3.7%), and only six strains were collected. The frequency of isolation in olive oil was higher (7.6%), and yielded four strains but was still markedly lower than that of olive brine (85.7%, six strains). Therefore, in total 16 new strains of *Saccharomyces* spp. were isolated from the olive niche.

Next, we obtained draft genome sequences of the new isolates in order to ascertain if they belonged to *S. cerevisiae* and, if so, to determine to which population they belonged to. As shown in **Figure 1A**, the new isolates were all identified as members of *S. cerevisiae*, thus showing that *S. paradoxus* was not isolated during our survey. Interestingly, the *S. cerevisiae* strains were found to belong to different populations. The six strains isolated directly from olive tree bark or ripe olives did not belong to the same population (**Figure 1A** and **Supplementary Table S1**). A single strain had substantial genomic contributions from *S. paradoxus* and, accordingly, was assigned to the Olives clade. Two strains belonged to the Wine clade, one strain to the Sake clade and two additional strains occupied an isolated position in the phylogeny and subsequent analyses showed that they had “mosaic” genomes with major contributions from the Wine and North American – Japan clades. With respect to the four strains isolated from olive oil, three of them had *S. paradoxus* contributions and belonged to the Olives clade. The remaining strain belonged to the Wine clade. For the strains isolated from olive brine, more homogeneous results were obtained and all of them were found to belong to the Olives clade (**Figure 1A**). Our phylogenetic analysis included also other strains that belong to the Olives clade and that had been isolated as far back as 1957 from olives, olive brine, alpechin, and from the gut or feces of humans and pigs (van Uden and Assis-Lopes, 1957; Santa María, 1958, 1962). It



**FIGURE 1 |** Hybrid olive strains form a monophyletic group and have a similar genomic organization. **(A)** Phylogenetic placement of hybrid olive strains among the known lineages of *Saccharomyces cerevisiae* (MO, Mediterranean oaks; NA-JP, North America – Japan). Whole-genome phylogenetic tree constructed after discarding *S. paradoxus* regions in all genomes. The phylogeny was inferred from 93 sequences and 913590 SNPs using the Maximum Likelihood method as implemented in IQ-TREE with the TVM+F+G4 model of sequence evolution and was rooted with *S. paradoxus*. Branch lengths correspond to the expected number of substitutions per site and black dots in tree nodes depict bootstrap support values above 85% (1000 replicates). Strains isolated from the olive niche are distinguished based on the specific isolation source (see color codes). **(B)** Similar divergence plots of the genomes of selected hybrid strains (highlighted in the phylogeny) to the reference genome of *S. paradoxus* CBS 432. The dotted lines depict the 10% divergence threshold that represents the average divergence between *S. cerevisiae* and *S. paradoxus*. The substantially distinct divergence plot of a Brazilian *S. cerevisiae* × *S. paradoxus* hybrid strain (UFMG-CM-Y651) previously reported by us (Barbosa et al., 2016) is included for comparison.

is noteworthy that the 25 *S. cerevisiae* × *S. paradoxus* hybrid strains isolated during this study or in previous studies formed a monophyllum even if the phylogeny of **Figure 1A** was prepared only with *S. cerevisiae* ORFs, thus avoiding the strong bias that would be introduced if *S. paradoxus* ORFs were considered. This suggests that all hybrid strains share the same *S. cerevisiae* ancestor irrespective of the geographical origin and particular substrate from which they were collected, a possibility also supported by the divergence plots depicting the *S. cerevisiae* and *S. paradoxus* blocks along the genome, that were similar for all strains of the Olives clade (**Figure 1B**). Moreover it appears the *S. cerevisiae* ancestor of the hybrids was a member of the wine population.

In conclusion, the results from our ecological survey do not support the hypothesis that the members of the Olives clade reside in the olive tree environment. Although the possibility that such strains are associated with olive trees cannot be entirely ruled out, it appears more likely that the ecological niche of this clade are processed olives and their products like olive oil, alpechin, which is the corresponding waste product, and table olives /olive brine. Also, the occurrence of hybrid strains in the intestinal tract is of notice. Besides the two strains (YJM 248 and YJM 1078) already reported in Strope et al. (2015) we found three additional strains from this source (PYCC 2613, PYCC 2708, and PYCC 8033).

## Fertility

Most strains (70%) of the Olives clade were sexually competent (**Supplementary Table S1**). Spore viability for two strains in this clade (PYCC 4935 and YO652) ranged 95.5 – 96% and a cross between them was also fertile (90% spore viability), thus suggesting that sexual recombination within the clade can occur. Also, a cross between YO 652 and the wine strain EXF 6719 (97% spore viability) had an ascospore fertility of 87%, thus indicating that sexual contact between the Olives and Wine populations appears not to be significantly hampered.

## Genomic Analysis

In order to characterize the genomic nature of the hybrids, we analyzed in detail 23 strains (**Supplementary Dataset S1**). We detected a total of 540 *S. paradoxus* ORFs and between 193 and 314 *S. paradoxus* ORFs per strain, with 103 ORFs being shared among all the strains. The *S. paradoxus* ORFs originated in the European population of this species (**Supplementary Figure S2**), thus suggesting that the hybridization event occurred in this continent. The co-existence of *S. cerevisiae* and *S. paradoxus* alleles for a given ORF was not frequent. In total 148 ORFs (27.4% of the total number of *S. paradoxus* ORFs) were found to occur in that configuration in at least one strain. One strain was devoid of ORFs represented in the genome by alleles belonging to the two species and 16 strains had only two to four ORFs with *S. cerevisiae* and *S. paradoxus* alleles. Together, these strains represent 74% of the total number of strains analyzed. Three strains had between 12 and 25 ORFs with *S. cerevisiae* and *S. paradoxus* alleles and another three strains had between 34 and 54 ORFs with *S. cerevisiae* and *S. paradoxus* alleles. The distribution of strains

having more heterozygous ORFs did not show any association with the isolation substrate or with the phylogeny.

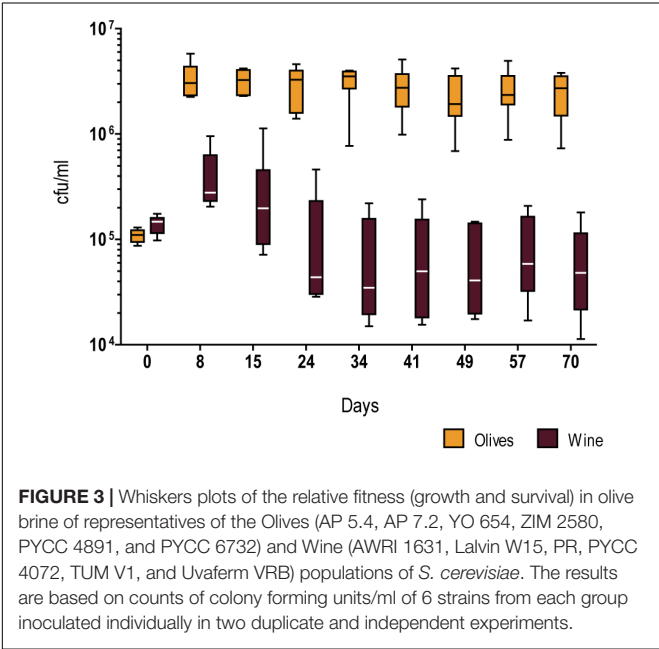
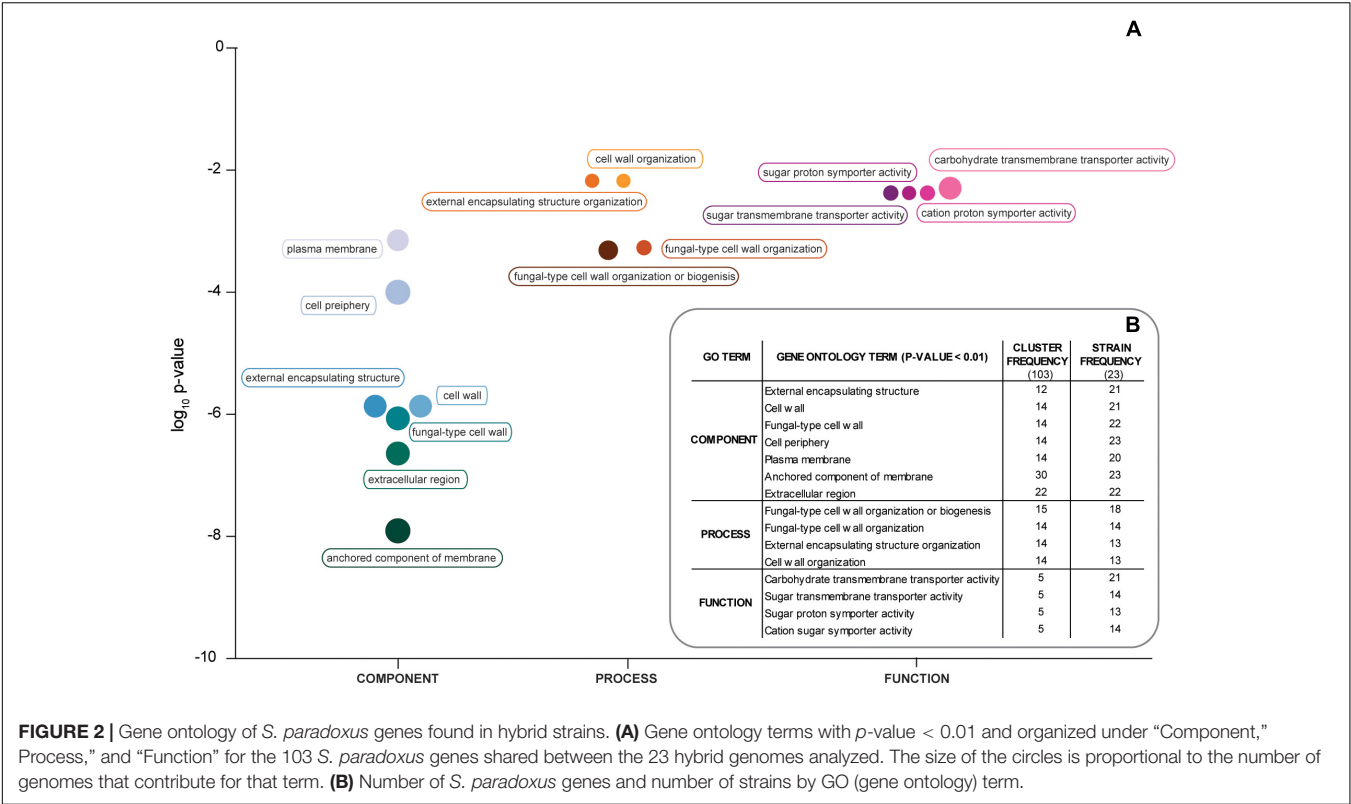
Gene Ontology analysis of the 103 ORFs shared among all the strains revealed a significant enrichment in genes encoding for proteins of the fungal cell wall and plasma membrane, like *TIP1*, *HLR1*, *DAN1*, *FCY21*, and *STL1*. However, for several strains, when an individual analysis was performed, a significant result for an enrichment in hexose transporters was also observed (**Figure 2** and **Supplementary Table S2**). Taken together, these results support the view that the hybrid strains have a similar core genomic composition, thus suggesting that they share a common (hybrid) ancestor and also that after hybridization have evolved adaptations to the processed olives niche.

Given that the hybrid strains descend from a *S. cerevisiae* wine strain (**Figure 1A**), we surveyed the hybrid genomes for the presence of typical domestication signatures of wine strains (Almeida et al., 2015; Barbosa et al., 2018). For regions A, B, and C, that encompass 39 genes potentially relevant for the winemaking process acquired by horizontal gene transfer from non-*Saccharomyces* species, at least one of these regions was present in 68.8% of the control group of 32 wine strains listed in **Supplementary Table S1**, whereas only 13% (3 out of 23) hybrid genomes shared the same characteristic. It thus appears that these regions are less prevalent in hybrid strains, either because their ancestor *S. cerevisiae* wine strain already lacked most of them and/or because there are not relevant in the olives niche and were therefore lost. With respect to the inactivation of aquaporin genes *AQY1* and *AQY2*, associated with the domestication of wine strains and with the adaptive loss of those water channels, a trait that increases fitness in sugar-rich environments (Will et al., 2010), no differences were found between the two groups and all strains had at least one aquaporin gene coding for a non-functional protein. We also investigated the variation of the number of copies of *CUP1*, a gene involved in resistance to copper toxicity in *S. cerevisiae*, especially in wine strains, due probably to their exposure to copper sulfate used in vineyards (Fay et al., 2004; Strope et al., 2015). Copy number variation (CNV) of the two paralogs of *CUP1* among reference wine and wild (oak-associated) strains is shown in **Supplementary Table S1**. Whereas among wine strains CNV of *CUP-1* could exceed 30 (in two cases), the Mediterranean oak (MO) strains did not show an enrichment in the number of *CUP-1* copies. Some of the hybrid strains showed also elevated numbers of copies of *CUP-1*, with the two most enriched genomes having 33 and 35 copies. Statistically, wine and olives strains could not be distinguished in terms of presence and expansion of *CUP-1*.

## Absolute Fitness in Olive Brine

In order to investigate whether strains of the Olive clade were adapted to thrive in the processed olives niche, we estimated absolute fitness of a set of six strains isolated from olive brine, olive oil, and alpechin, and compared it to six *S. cerevisiae* strains from the wine population. We measured absolute fitness as the number of viable cells maintained in a long-term batch culture of table olive brine, here used as a proxy for a habitat to which the *S. cerevisiae* hybrids are adapted to (**Figure 3**, **Supplementary Figure S1** and **Supplementary Table**





S3). Employing freshly collected ripe olives, we prepared a brine containing 8% (w/v) NaCl (see section “Materials and Methods” for details) which was used to test each strain separately. The absolute fitness of the olive and wine strain cohorts was inferred by measuring viable cell numbers for 70 days in two independent experiments (Figure 3, Supplementary Figure S1

and Supplementary Table S3). Although strain fitness varies within both groups, variation is much more pronounced in the wine group. In spite of the within-group differences among strains, a clear difference is observable between fitness of the olive group and the wine group ( $p < 0.0001$ , unpaired *t*-test with Welch’s correction), the former being able both to attain higher cell numbers and to sustain viability throughout the duration of the experiment (70 days). On the contrary, wine strains tended to start losing viability already during the first month of incubation.

We reasoned that one possible cause for the difference in fitness between the two groups might be related to their ability to use the nutrients available in olive brine. Contrary to what is typical of the initial stages of wine fermentation, olive brine has low concentrations of sugars. To analyze this in more detail, we identified and quantified the sugars and sugar-related compounds present in olive brine and measured their consumption by two representatives of the wine and two representatives of the olive cohorts (Table 1). While strains belonging to the Olives clade virtually exhausted the glucose and fructose present initially in the brine, the representatives of the Wine clade consume only about half of the available sugars. Mannitol was left untouched in both cases.

This finding was intriguing because some wine strains were previously found to have an impairment in high affinity hexose transport (the type of transporters expressed under the low sugar concentrations measured in olive brine), a trait that was subsequently associated to certain variants of the HXT hexose transporter genes (Luyten et al., 2002). Numerous HXT genes are present in the genomes of the species of

**TABLE 1 |** Sugar consumption in olive brine by strains of the Olives and Wine population of *S. cerevisiae*.

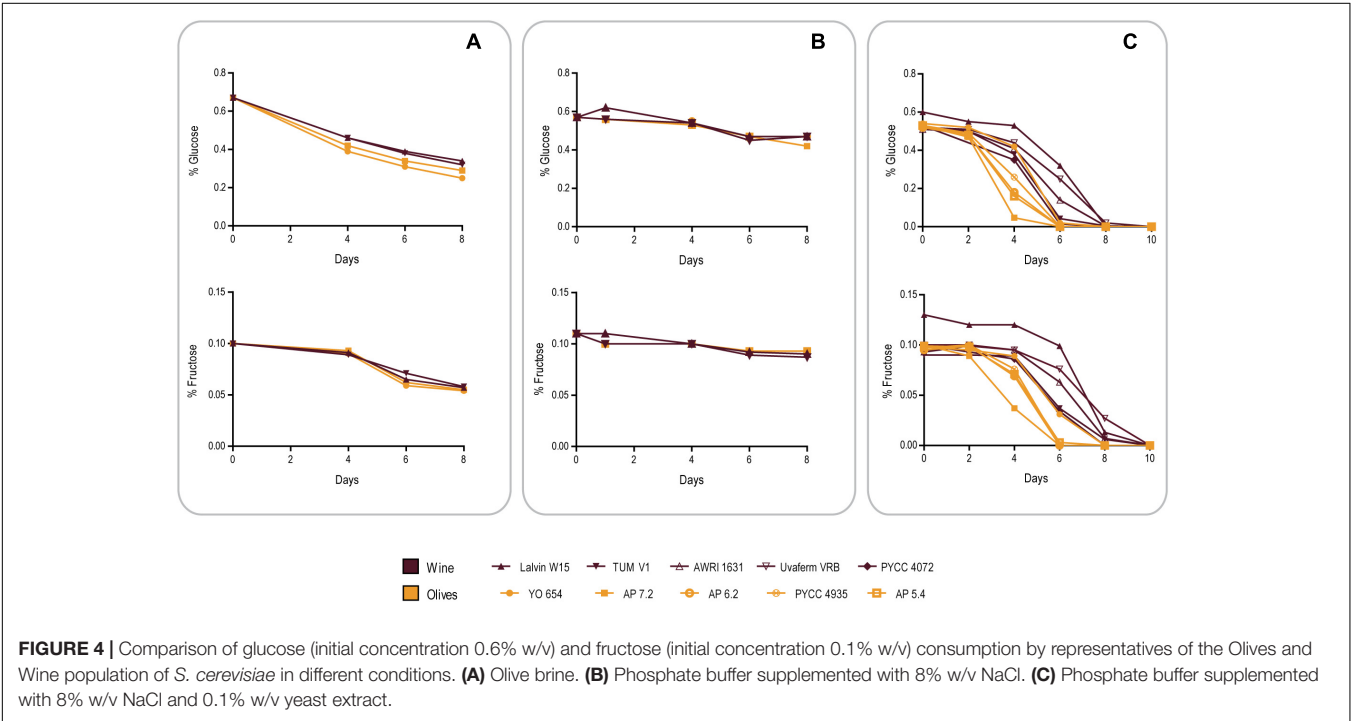
Time (days)	Population	Strain	pH	Glucose % (w/v)	Fructose % (w/v)	Mannitol % (w/v)
0			5.1	0.57	0.1	0.14
70	Olives	YO 654	5.0	0.0055	0.0023	0.13
		AP 7.2	4.9	0.0058	0.0039	0.14
	Wine	Lalvin W15	5.2	0.26	0.077	0.13
		TUM V1	5.2	0.38	0.083	0.14

Initial and final (after 70 days) pH and sugar measurements are shown as averages for each strain from two independent experiments.

the genus *Saccharomyces*, encoding transporters with different affinities for their substrates. Since hexose transporter genes encoding the main high affinity transporters (HXT6/7) were among those “replaced” in the hybrids by their *S. paradoxus* counterparts (**Supplementary Dataset S1**), we asked if these substitutions might have contributed to improve high affinity hexose transport in the hybrids.

To assess this, we compared the ability of the same strains of the Wine and Olives clades used in the previous experiment (see **Table 1**) to consume the sugars present in olive brine in the course of the first 8 days after inoculation of brine (**Figure 4A**). Surprisingly, and although in this period olive strains grew on average in the brine fitness experiments shown in **Figure 3** an order of magnitude more than wine strains, sugar consumption was very similar between wine and olive strains. Fructose consumption in particular was indistinguishable, while olive strains seemed to assimilate glucose slightly better, an observation that nevertheless does not explain the differences in growth in brine between the two sets of strains (**Figure 3**). A distinct experiment was subsequently performed in which brine was replaced by phosphate buffer supplemented with NaCl,

glucose and fructose in concentrations identical to those found in brine. This time wine strains seemed to be slightly more proficient in fructose assimilation while in glucose no clear differences were observed (**Figure 4B**). Nevertheless, when this experiment was performed without NaCl, glucose and fructose were totally consumed after 2 days by wine and olives strains. Taken together these results suggest that no considerable differences in sugar uptake capacities exist between the two groups of strains that justify the better growth of olive strains. It seems therefore that the observed difference in growth is due to a better capacity to adapt to the harsh conditions of olive brine, of which high salt concentrations stand out, resulting in better growth for olive strains during the first 8 days while consuming the same amount of sugar in the same period as wine strains. The same experiment as shown in **Figure 4B** was subsequently performed but this time adding 0.1% yeast extract to the phosphate buffer and adding more strains to both cohorts to increase representativeness (**Figure 4C**). The sugar consumption profiles of both strain cohorts were different in these conditions, with olive strains exhausting the available sugars significantly faster ( $p < 0.05$ , unpaired *t*-test), which means that in the presence of the required



nutrients olive strains are better equipped to adapt to the high salt medium. Interestingly, under these conditions wine strains were capable of consuming all the sugar available after 8–10 days while they only consumed about 50% of the available sugars during the fitness experiments in brine. This could mean that brine contains other inhibitors that affect the metabolism of wine strains more than that of olive strains, in addition to NaCl.

In summary, while *S. paradoxus* HXT genes may confer a slight advantage for glucose consumption in brine, this advantage does not explain the considerable difference in the ability of wine and olive strains to grow in brine. Instead, this difference seems to be derived from a better adaptation of olive strains to the particular conditions of brine of which the high NaCl concentration appears as a relevant factor.

## Adaptation to NaCl

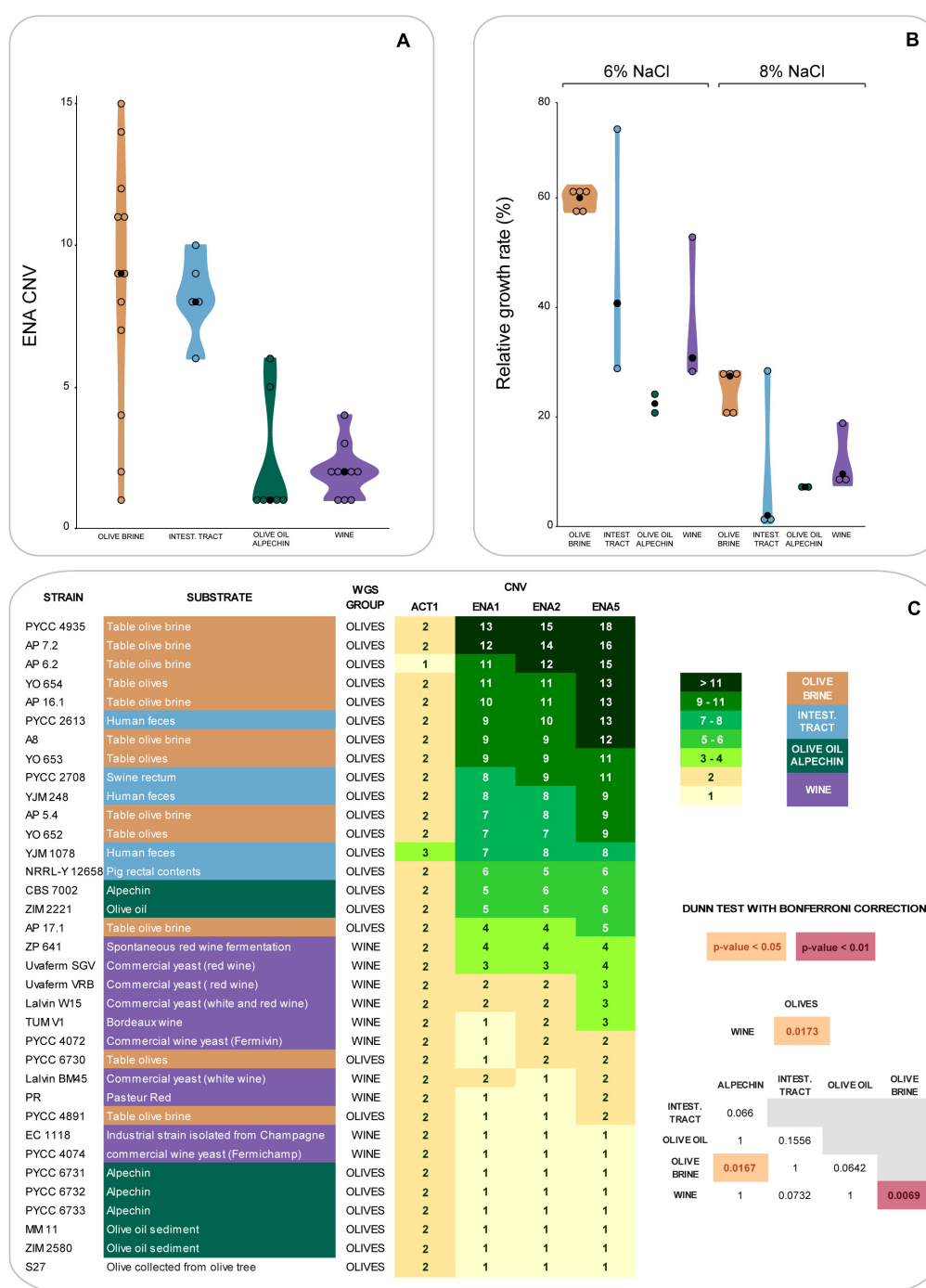
The results of the experiments shown in **Figures 3, 4** suggested that fitness and sugar consumption aptitude in brine might be related, at least partly, to salt resistance. To investigate this hypothesis, we started by determining the copy number of the three *ENA* genes found in the *S. cerevisiae* reference genome (*ENA1*, *ENA2*, and *ENA5*) in the hybrid strains of the Olives clade and compared their abundance using 10 representative strains of the Wine clade. The *ENA* proteins are sodium pumps that help the cells to cope with an excess of sodium ions in their environment (Ruiz and Ariño, 2007). *ENA* copy number variation is shown in **Figures 5A,C**. Interestingly, the highest *ENA* copy number (14–18 copies) was detected among strains isolated from olive brine (**Figure 5C**), although a marked dispersion in the number of *ENA* copies was observed in this group (**Figure 5A**). Overall, strains isolated from olive brine and the intestinal tract were more likely to have a higher number of *ENA* copies than strains isolated from olive oil, alpechin or wine (**Figure 5A**). It is possible that the strains found in the intestinal tract originate from the olive brine environment, having been subsequently ingested. This would explain their increased number of *ENA* copies. The differences between the number of *ENA* copies were found to be statistically significant between the Wine and Olive Brine populations (**Figure 5C**,  $p < 0.01$ , Dunn *post hoc* test with Bonferroni correction). The comparison of *ENA* copy numbers was also significantly distinct when all hybrid strains from the olives niche were compared with wine strains and for the comparison between the olive brine and alpechin groups ( $p < 0.05$ ).

To investigate to which extent *ENA* gene copy numbers determined fitness of the strains under study in the presence of salt, the ability of the various strains to grow in the presence of 6% and 8% NaCl was also tested (**Figure 5B**). There was, as expected, a correlation between *ENA* gene copy numbers and the ability to grow in the presence of NaCl, but this correlation was not complete. For example, while all strains isolated from olive brine performed well in the growth tests even if they had only moderately high *ENA* copy numbers (e.g., AP 17.1; 4–5 copies), strains associated with the intestinal tract behaved heterogeneously, varying between an excellent performance (PYCC 2613; 9–13 copies) and a very poor performance (e.g., YJM 1078; 7–8 copies).

## DISCUSSION

Here we analyzed in detail a *S. cerevisiae* × *S. paradoxus* hybrid lineage associated with a distinctive artificial environment, that of processed olives. Even when only *S. cerevisiae* ORFs are considered, these hybrid strains form a distinct and exclusive monophyletic lineage among those already known for *S. cerevisiae*. This suggests that the olive hybrids are genetically isolated from the other *S. cerevisiae* populations and that it is likely that they descend from a single ancestral hybridization event. A set of other additional features also suggests that olive hybrids constitute a “natural” population in an evolutionary and ecological sense. First, sexual recombination appears to be possible within members of this population and secondly the olive hybrids have a considerable dissemination both in space and in time, since in this study we analyzed representatives from the Iberian Peninsula (Portugal and Spain), Southeast Europe (Slovenia and Croatia), and also strains from the United States. Moreover, these strains were collected during a period of time that spans six decades (1957–2018). The Olives population exhibits a particular ecological preference to environments having in common the presence of processed olives or their products, but not the olive tree itself. Therefore it appears that the origin of this population is linked to human activities and to artificial substrates they create. Although some strains were found associated to the intestinal tract, these strains exhibit the characteristic expansion of the *ENA* gene copy number typical of olive brine strains, suggesting that they may have been ingested together with cured olives. The occurrence of *S. cerevisiae* hybrids in the intestinal tract parallels other reports of association of *S. cerevisiae* with humans (Angebault et al., 2013; Strobe et al., 2015) and warrants the need for investigating if these strains are better adapted to survive in the intestinal tract.

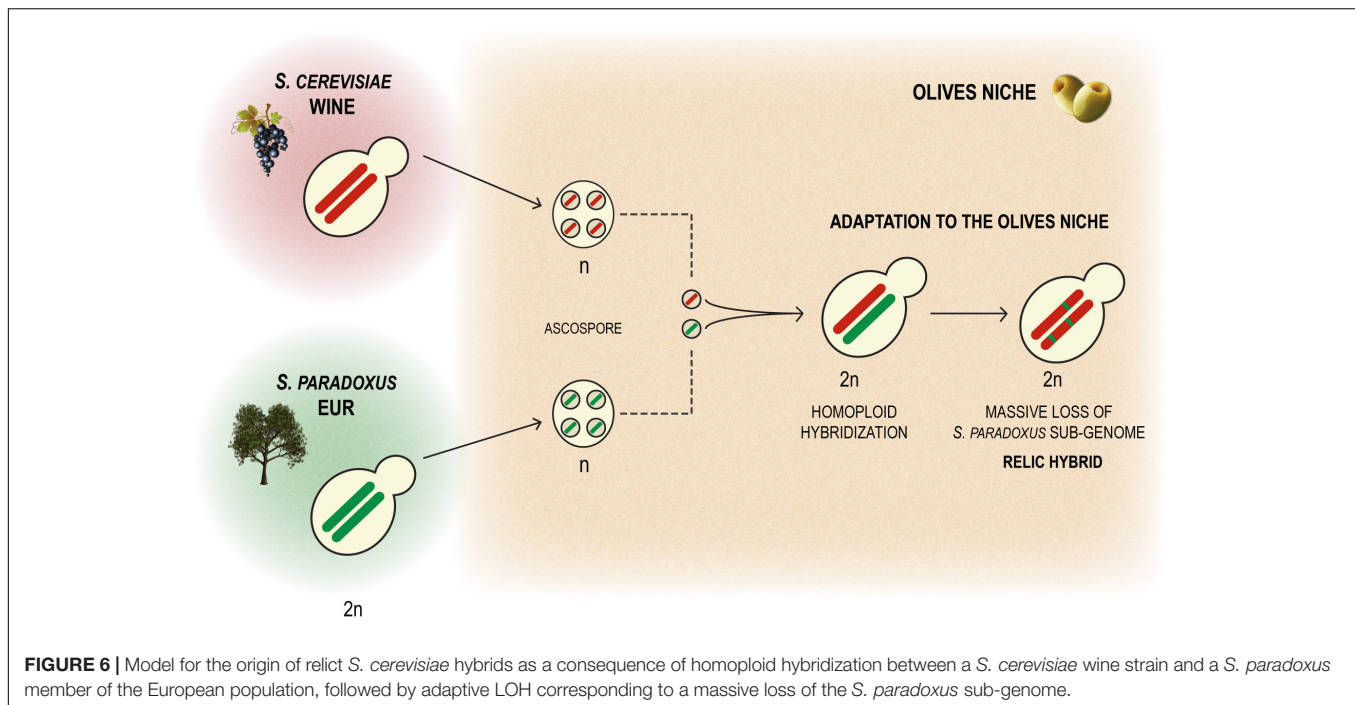
A striking feature of the genomes of the strains of the Olives population is the markedly unbalanced contribution of the two parental sub-genomes, *S. cerevisiae* being the clearly prevalent sub-genome given that *S. paradoxus* contributes only around 3.7%. A likely scenario for the origin of the hybrid that originated the Olives lineage is homoploid hybridization. This would have corresponded to the fusion of a *S. paradoxus* meiospore with a *S. cerevisiae* meiospore resulting in a “normal” diploid hybrid genome that subsequently underwent the adaptive loss of most of the *S. paradoxus* sub-genome (**Figure 6**). Therefore the observed genomic organization of the strains of the Olives lineage can be seen as reminiscent of a relic hybridization. This hybridization appears to have occurred in Europe since the *S. paradoxus* progenitor belongs to the recognized population of this species. We could determine that the *S. cerevisiae* progenitor belonged to the Wine population and that the hybrids still exhibit some of the domestication signatures of this population such as the loss of functional aquaporins, the expansion of *CUP-1* genes and the presence of region B, reminiscent of the presence of regions A, B, and C, typical in wine strains. It is noteworthy that these relic hybrids are capable of sexual reproduction, which would have facilitated the emergence of a population adapted to a newly colonized niche. The ecological barrier between the processed olives niche and the vineyards/winery



environment, even if incipient would also have promoted, together with selection, the ecological specialization of the new genotypes. Therefore, the model we propose to explain the

emergence of the Olives population is based on an original homoploid interspecies hybridization followed by a massive adaptive loss of heterozygosity (LOH) by replacement of most





*S. paradoxus* alleles by their *S. cerevisiae* orthologs, combined with intra-population gene flow through sexual recombination and evolution of new ecological adaptations, with backcrossing with the *S. cerevisiae* parent probably playing a very limited role. Similar cases of apparent reduction of the non-*cerevisiae* sub-genome have also been reported for artificially generated hybrids involving *S. kudriavzevii* (Lopandic et al., 2016) and *S. uvarum* (Antunovics et al., 2005). Most importantly, LOH following hybridization, i. e. after a dramatic gain of genetic variation through interspecies hybridization, has been revealed as a major adaptation mechanism of populations when they invade new ecological niches (Smukowski Heil et al., 2017). Contrary to previous examples known exclusively from experimental evolution studies in the laboratory (e.g., Dunn et al., 2013; Smukowski Heil et al., 2017), the Olives population illustrates the fate of a relic hybridization in real conditions. It is also relevant to mention that although the fraction of *S. paradoxus* genome is relatively small, it is still much larger than instances of introgression of *S. paradoxus* in *S. cerevisiae* reported so far (Doniger et al., 2008; Barbosa et al., 2016), excluding the cases reported by Muller and McCusker (2009), in which information on strain origin was not given but that in fact correspond to the intestinal strains studied here. The relatively high number of homozygous *S. paradoxus* ORFs (342 out of 540) found in the Olives clade suggests that the genomic contribution of this species to adaptation to the processed olives environment is likely to involve a multiplicity of cellular processes. Gene ontology analysis of the set of *S. paradoxus* genes present in homozygosity in all hybrid strains examined here, suggested that cell wall function and hexose transport were likely among the cellular processes benefiting from the *S. paradoxus* genomic contribution. Because inefficient high affinity hexose transport

was found to be associated with specific HXT alleles carried by some wine strains (Luyten et al., 2002), we investigated whether the *S. paradoxus* HXT alleles retained by the olives strains were likely to contribute to the observed superior ability of these strains to consume sugars in brine. However, our experiments indicate that sugar consumption profiles are different between members of the Olives and Wine populations solely in the presence of NaCl and only when the cells are cultivated in nutritional conditions that support adaptation of their proteome to the high salt environment. Assuming that the HXT transporters operating when cells are cultivated with or without a nitrogen source are in both instances the high affinity HXT6/7 transporters, these observations suggest that the differences perceived are due to a better overall fitness of olives strains in the presence of high NaCl concentrations, rather than to a better intrinsic ability of the *S. paradoxus* HXT6/7 versions to operate in the presence of salt. We observed that the strains isolated from olive brine had a tendency for having an increased number of copies of *ENA* genes, a feature known to increase tolerance to NaCl. However this tendency was not universal among olive brine strains and even strains with a lower number of *ENA* copies grew relatively well in the presence of NaCl, thus suggesting that other mechanisms might also be involved in the adaptation to NaCl of olive brine strains, as has been already documented (Posas et al., 2000; Dhar et al., 2011; Saito and Posas, 2012).

Interestingly, wine strains were able to exhaust glucose and fructose in the medium containing high NaCl concentrations while they failed to do so in brine even after 70 days, suggesting that inhibiting components other than NaCl are affecting the performance of wine strains in brine. Also, according to our results, these inhibitory components were

not the phenolic compounds likely present in brine, since we failed to detect differences in the sensitivity of wine and olives strains (4 strains from each group) to oleuropein (3% w/v in YPD medium, pH 5), and ferulic acid (2% w/v in YPD medium, pH 4.5).

The emergence of the relic olive hybrids from an already domesticated lineage (wine yeasts) in the artificial environment of processed olives can be seen as another instance of yeast domestication, or even a case of secondary domestication *sensu* Barbosa et al. (2018). However, contrary to wine and beer domestication, where genomic and phenotypic changes can be linked to characteristics of these beverages valued and improved over time by humans, in the present case the beneficial role of the relic hybrids has not been clearly demonstrated in olive brine fermentations and therefore their origin and prevalence in the processed olives niche can be viewed as inconsequential to humans. One illustrative example of a tangible consequence of domestication is the inactivation of *PAD1* and *FDC1* genes in beer yeasts which overcomes the phenolic off flavor defect (Gallone et al., 2016; Gonçalves et al., 2016). The phenolic aroma, due the formation of 4-vinyl guaicol, is negatively valued in most in beers but not in wine where it can be even considered as desirable. The consequence of artificial selection is that beer yeasts differ from wine and wild strains in having acquired inactivating mutations in *PAD1* and *FDC1*. Therefore, if domestication is viewed as the controlled bred of an organism that becomes genetically distinct from its wild relatives in ways making it more useful to humans (Diamond, 2002), relic olive hybrids can be seen as a case of adaptation to the human environment but without the emergence of traits that we can readily recognize as useful. In order to reflect this distinct stage of “incomplete” domestication we define these changes as a quasi-domestication event.

## AUTHOR CONTRIBUTIONS

AP, JS, and PG conceived the study, analyzed the data, and wrote the manuscript. AP and NČ performed the experiments. AP prepared the figures.

## REFERENCES

- Abyzov, A., Urban, A. E., Snyder, M., and Gerstein, M. (2011). CNVnator: an approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing. *Genome Res.* 21, 974–984. doi: 10.1101/gr.114876.110
- Almeida, P., Barbosa, R., Bensasson, D., Gonçalves, P., and Sampaio, J. P. (2017). Adaptive divergence in wine yeasts and their wild relatives suggests a prominent role for introgressions and rapid evolution at noncoding sites. *Mol. Ecol.* 26, 2167–2182. doi: 10.1111/mec.14071
- Almeida, P., Barbosa, R., Zalar, P., Imanishi, Y., Shimizu, K., Turchetti, B., et al. (2015). A population genomics insight into the mediterranean origins of wine yeast domestication. *Mol. Ecol.* 24, 5412–5427. doi: 10.1111/mec.13341
- Almeida, P., Gonçalves, C., Teixeira, S., Libkind, D., Bontrager, M., Masneuf-Pomarède, I., et al. (2014). A Gondwanan imprint on global diversity and domestication of wine and cider yeast *Saccharomyces uvarum*. *Nat. Commun.* 5:4044. doi: 10.1038/ncomms5044

## FUNDING

This work was supported by Fundação para a Ciência e a Tecnologia (Portugal) grants PTDC/BIA-MIC/30785/2017 (AP, JS, and PG), UID/Multi/04378/2013 (AP, JS, and PG), and SFRH/BD/136462/2018 (AP) and by Slovenian Research Agency Grants P4-0116 and MRIC-UL ZIM, IP-0510 (NČ).

## ACKNOWLEDGMENTS

Aimée Dudley and Justin Fay kindly provided strains YO 392, YO 652, YO 653, and YO 654.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2019.00449/full#supplementary-material>

**FIGURE S1 |** Growth and survival in olive brine of six strains of the Olives population and six strains of the Wine population. Two independent experiments were performed for each strain.

**FIGURE S2 |** The *S. paradoxus* sub-genome of hybrid strains originates in the European population of *S. paradoxus*. Phylogenetic tree constructed with the *S. paradoxus* fraction of the hybrid genomes and correspondent regions from representatives of the European, American, and Far Eastern population of *S. paradoxus*. The phylogeny was inferred from 47 sequences and 22052 SNPs using the Neighbor Joining method and the P-distance model of sequence evolution. Branch lengths correspond to the expected number of substitutions per site and black dots in tree nodes depict bootstrap support values above 90% (1000 replicates).

**TABLE S1 |** Strains and genomes used in this study and relevant information pertaining to them.

**TABLE S2 |** Gene ontology (GO) analysis of the *S. paradoxus* sub-genome of relict hybrid strains.

**TABLE S3 |** Cfu/ml counts per strain and per replicate for the growth and survival in olive brine experiment shown in **Figure 2**.

**DATASET S1 |** Gene content of the *S. paradoxus* sub-genome of relict hybrid strains.

- Angebault, C., Djossou, F., Abélanet, S., Permal, E., Soltana, M., Ben Diancourt, L., et al. (2013). *Candida albicans* is not always the preferential yeast colonizing humans: a study inwayampi amerindians. *J. Infect. Dis.* 208, 1705–1716. doi: 10.1093/infdis/jit389
- Antunovics, Z., Nguyen, H., Gaillardin, C., and Sipiczki, M. (2005). Gradual genome stabilisation by progressive reduction of the *Saccharomyces uvarum* genome in an interspecific hybrid with *Saccharomyces cerevisiae*. *FEMS Yeast Res.* 5, 1141–1150. doi: 10.1016/j.femsyr.2005.04.008
- Arroyo-López, F. N., Querol, A., Bautista-Gallego, J., and Garrido-Fernández, A. (2008). Role of yeasts in table olive production. *Int. J. Food Microbiol.* 128, 189–196. doi: 10.1016/j.ijfoodmicro.2008.08.018
- Barbosa, R., Almeida, P., Safar, S. V. B., Santos, R. O., Morais, P. B., Nielly-Thibault, L., et al. (2016). Evidence of natural hybridization in brazilian wild lineages of *Saccharomyces cerevisiae*. *Genome Biol. Evol.* 8, 317–329. doi: 10.1093/gbe/evv263
- Barbosa, R., Pontes, A., Santos, R. O., Montandon, G. G., Ponzzes-Gomes, C. M., Morais, P. B., et al. (2018). Multiple rounds of artificial selection promote

- microbe secondary domestication - the case of cachaça yeasts. *Genome Biol. Evol.* 10, 1939–1955. doi: 10.1093/gbe/evy132
- Bergström, A., Simpson, J. T., Salinas, F., Barré, B., Parts, L., Zia, A., et al. (2014). A high-definition view of functional genetic variation from natural yeast genomes. *Mol. Biol. Evol.* 31, 872–888. doi: 10.1093/molbev/msu037
- Bonatsou, S., Paramithiotis, S., and Panagou, E. Z. (2018). Evolution of yeast consortia during the fermentation of kalamata natural black olives upon two initial acidification treatments. *Front. Microbiol.* 8:2673. doi: 10.3389/fmicb.2017.02673
- Cromie, G. A., Hyma, K. E., Ludlow, C. L., Garmendia-Torres, C., Gilbert, T. L., May, P., et al. (2013). Genomic sequence diversity and population structure of *Saccharomyces cerevisiae* assessed by RAD-seq. *G3 (Bethesda)*. 3, 2163–2171. doi: 10.1534/g3.113.007492
- Dhar, R., Sägeser, R., Weikert, C., Yuan, J., and Wagner, A. (2011). Adaptation of *Saccharomyces cerevisiae* to saline stress through laboratory evolution. *J. Evol. Biol.* 24, 1135–1153. doi: 10.1111/j.1420-9101.2011.02249.x
- Diamond, J. (1997). *Guns, Germs, and Steel: The Fates of Human Societies*. New York, NY: W.W. Norton & Company.
- Diamond, J. (2002). Evolution, consequences and future of plant and animal domestication. *Nature* 418, 700–707. doi: 10.1038/nature01019
- Doebley, J. F., Gaut, B. S., and Smith, B. D. (2006). The molecular genetics of crop domestication. *Cell* 127, 1309–1321. doi: 10.1016/j.cell.2006.12.006
- Doniger, S. W., Kim, H. S., Swain, D., Corcuera, D., Williams, M., Yang, S.-P., et al. (2008). A catalog of neutral and deleterious polymorphism in yeast. *PLoS Genet.* 4:e1000183. doi: 10.1371/journal.pgen.1000183
- Duan, S.-F., Han, P.-J., Wang, Q.-M., Liu, W.-Q., Shi, J.-Y., Li, K., et al. (2018). The origin and adaptive evolution of domesticated populations of yeast from far East Asia. *Nat. Commun.* 9:2690. doi: 10.1038/s41467-018-05106-7
- Dunn, B., Paulish, T., Stanbery, A., Piotrowski, J., Koniges, G., Kroll, E., et al. (2013). Recurrent rearrangement during adaptive evolution in an interspecific yeast hybrid suggests a model for rapid introgression. *PLoS Genet.* 9:e1003366. doi: 10.1371/journal.pgen.1003366
- Enache-Angoulvant, A., and Hennequin, C. (2005). Invasive *Saccharomyces* infection: a comprehensive review. *Clin. Infect. Dis.* 41, 1559–1568. doi: 10.1086/497832
- Fay, J. C., McCullough, H. L., Sniegowski, P. D., Eisen, M. B., Schadt, E., Monks, S., et al. (2004). Population genetic variation in gene expression is associated with phenotypic variation in *Saccharomyces cerevisiae*. *Genome Biol.* 5:R26. doi: 10.1186/gb-2004-5-4-r26
- Gallone, B., Steensels, J., Prah, T., Soriaga, L., Saels, V., Herrera-Malaver, B., et al. (2016). Domestication and divergence of *Saccharomyces cerevisiae* beer yeasts. *Cell* 166, 1397–1410. doi: 10.1016/j.cell.2016.08.020
- Gonçalves, M., Pontes, A., Almeida, P., Barbosa, R., Serra, M., Libkind, D., et al. (2016). Distinct domestication trajectories in top-fermenting beer yeasts and wine yeasts. *Curr. Biol.* 26, 2750–2761. doi: 10.1016/j.cub.2016.08.040
- Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* 59, 307–321. doi: 10.1093/sysbio/syq010
- Hornsey, I. (2003). *History of Beer and Brewing*. Cambridge: Royal Society of Chemistry.
- Hutter, S., Vilella, A. J., and Rozas, J. (2006). Genome-wide DNA polymorphism analyses using VariScan. *BMC Bioinformatics* 7:409. doi: 10.1186/1471-2105-7-409
- Legras, J.-L., Galeote, V., Bigey, F., Camarasa, C., Marsit, S., Nidelet, T., et al. (2018). Adaptation of *S. cerevisiae* to fermented food environments reveals remarkable genome plasticity and the footprints of domestication. *Mol. Biol. Evol.* 35, 1712–1727. doi: 10.1093/molbev/msy066
- Letunic, I., and Bork, P. (2016). Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* 44, W242–W245. doi: 10.1093/nar/gkw290
- Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics* 25, 1754–1760. doi: 10.1093/bioinformatics/btp324
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352
- Liti, G., Ba, A. N., Blythe, M., Müller, C. A., Bergström, A., Cubillos, F. A., et al. (2013). High quality de novo sequencing and assembly of the *Saccharomyces arboricolus* genome. *BMC Genomics* 14:69. doi: 10.1186/1471-2164-14-69
- Lopandic, K., Pflieger, W. P., Tiefenbrunner, W., Gangl, H., Sipiczki, M., and Sterflinger, K. (2016). Genotypic and phenotypic evolution of yeast interspecies hybrids during high-sugar fermentation. *Appl. Microbiol. Biotechnol.* 100, 6331–6343. doi: 10.1007/s00253-016-7481-0
- Luyten, K., Riou, C., and Blondin, B. (2002). The hexose transporters of *Saccharomyces cerevisiae* play different roles during enological fermentation. *Yeast* 19, 713–726. doi: 10.1002/yea.869
- McGovern, P. E. (2009). *Uncorking the Past: The Quest for Wine, Beer, and Other Alcoholic Beverages*. Berkeley CA: University of California Press.
- Muller, L. A. H., and McCusker, J. H. (2009). A multispecies-based taxonomic microarray reveals interspecies hybridization and introgression in *Saccharomyces cerevisiae*. *FEMS Yeast Res.* 9, 143–152. doi: 10.1111/j.1567-1364.2008.00464.x
- Munoz, P., Bouza, E., Cuenca-Estrella, M., Eiros, J. M., Perez, M. J., Sanchez-Somolinos, M., et al. (2005). *Saccharomyces cerevisiae* fungemia: an emerging infectious disease. *Clin. Infect. Dis.* 40, 1625–1634. doi: 10.1086/429916
- Nguyen, L.-T., Schmidt, H. A., von Haeseler, A., and Minh, B. Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32, 268–274. doi: 10.1093/molbev/msu300
- Okonechnikov, K., Golosova, O., Fursov, M., and Ugene team (2012). Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics* 28, 1166–1167. doi: 10.1093/bioinformatics/bts091
- Peter, J., De Chiara, M., Friedrich, A., Yue, J.-X., Pflieger, D., Bergström, A., et al. (2018). Genome evolution across 1,011 *Saccharomyces cerevisiae* isolates. *Nature* 556, 339–344. doi: 10.1038/s41586-018-0030-5
- Posas, F., Chambers, J. R., Heyman, J. A., Hoeffler, J. P., de Nadal, E., and Ariño, J. (2000). The transcriptional response of yeast to saline stress. *J. Biol. Chem.* 275, 17249–17255. doi: 10.1074/jbc.M910016199
- Preiss, R., Tyrawa, C., Krogerus, K., Garshol, L. M., and Van Der Merwe, G. (2018). Traditional norwegian kveik are a genetically distinct group of domesticated *Saccharomyces cerevisiae* brewing yeasts. *Front. Microbiol.* 9:2137. doi: 10.3389/fmicb.2018.02137
- Ross-Ibarra, J., Morrell, P. L., and Gaut, B. S. (2007). Plant domestication, a unique opportunity to identify the genetic basis of adaptation. *Proc. Natl. Acad. Sci. U.S.A.* 104, 8641–8648. doi: 10.1073/pnas.0700643104
- Ruiz, A., and Ariño, J. (2007). Function and regulation of the *Saccharomyces cerevisiae* ENA sodium ATPase system. *Eukaryot. Cell* 6, 2175–2183. doi: 10.1128/EC.00337-07
- Saito, H., and Posas, F. (2012). Response to hyperosmotic stress. *Genetics* 192, 289–318. doi: 10.1534/genetics.112.140863
- Sampaio, J. P., and Gonçalves, P. (2008). Natural populations of *Saccharomyces kudriavzevii* in Portugal are associated with oak bark and are sympatric with *S. cerevisiae* and *S. paradoxus*. *Appl. Environ. Microbiol.* 74, 2144–2152. doi: 10.1128/AEM.02396-07
- Santa María, J. (1958). Ecología de las levaduras. II. Levaduras del azúcar, de la leche condensada y de la remolacha. *An. del Inst. Nac. Investig. Agronómicas* 8, 779–796. doi: 10.1016/j.riam.2017.03.001
- Santa María, J. (1962). *Saccharomyces hienipiensis*, a new melibiose-fermenting yeast, unable to assimilate raffinose. *J. Gen. Microbiol.* 28, 375–384. doi: 10.1099/00221287-28-2-375
- Scannell, D. R., Zill, O. A., Rokas, A., Payen, C., Dunham, M. J., Eisen, M. B., et al. (2011). The awesome power of yeast evolutionary genetics: new genome sequences and strain resources for the *Saccharomyces sensu stricto* genus. *G3 (Bethesda)*. 1, 11–25. doi: 10.1534/g3.111.000273
- Smukowski, H., Heil, C. S., DeSevo, C. G., Pai, D. A., Tucker, C. M., Hoang, M. L., and Dunham, M. J. (2017). Loss of heterozygosity drives adaptation in hybrid yeast. *Mol. Biol. Evol.* 34, 1596–1612. doi: 10.1093/molbev/msx098
- Strope, P. K., Skelly, D. A., Kozmin, S. G., Mahadevan, G., Stone, E. A., Magwene, P. M., et al. (2015). The 100-genomes strains, an *S. cerevisiae* resource that illuminates its natural phenotypic and genotypic variation and emergence as an opportunistic pathogen. *Genome Res.* 25, 762–774. doi: 10.1101/gr.185538.114

- Tilakaratna, V., and Bensasson, D. (2017). Habitat predicts levels of genetic admixture in *Saccharomyces cerevisiae*. *G3 (Bethesda)*. 7, 2919–2929. doi: 10.1534/g3.117.041806
- van Uden, N., and Assis-Lopes, L. (1957). Fermentation of raffinose in the absence of invertase by *Saccharomyces italicus castelli* var. *melibiosi* nov. var. *Port. Acta Biol.* 4, 323–327.
- Will, J. L., Kim, H. S., Clarke, J., Painter, J. C., Fay, J. C., and Gasch, A. P. (2010). Incipient balancing selection through adaptive loss of aquaporins in natural *saccharomyces cerevisiae* populations. *PLoS Genet.* 6:e1000893. doi: 10.1371/journal.pgen.1000893

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Pontes, Čadež, Gonçalves and Sampaio. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.





# Adaptive Response and Tolerance to Acetic Acid in *Saccharomyces cerevisiae* and *Zygosaccharomyces bailii*: A Physiological Genomics Perspective

Margarida Palma, Joana F. Guerreiro and Isabel Sá-Correia\*

Institute for Bioengineering and Biosciences, Department of Bioengineering, Instituto Superior Técnico, Universidade de Lisboa, Lisbon, Portugal

## OPEN ACCESS

### Edited by:

Jean Marie François,  
UMR5504 Laboratoire d'Ingénierie  
des Systèmes Biologiques et des  
Procédés (LISBP), France

### Reviewed by:

Sergio Giannattasio,  
Istituto di Biomembrane,  
Bioenergetica e Biotecnologie  
Molecolari (IBIOM), Italy  
Gemma Beltran,  
Universitat Rovira i Virgili, Spain

### \*Correspondence:

Isabel Sá-Correia  
isacorreia@tecnico.ulisboa.pt

### Specialty section:

This article was submitted to  
Evolutionary and Genomic  
Microbiology,  
a section of the journal  
Frontiers in Microbiology

Received: 13 December 2017

Accepted: 06 February 2018

Published: 21 February 2018

### Citation:

Palma M, Guerreiro JF and  
Sá-Correia I (2018) Adaptive  
Response and Tolerance to Acetic  
Acid in *Saccharomyces cerevisiae*  
and *Zygosaccharomyces bailii*:  
A Physiological Genomics  
Perspective. *Front. Microbiol.* 9:274.  
doi: 10.3389/fmicb.2018.00274

Acetic acid is an important microbial growth inhibitor in the food industry; it is used as a preservative in foods and beverages and is produced during normal yeast metabolism in biotechnological processes. Acetic acid is also a major inhibitory compound present in lignocellulosic hydrolysates affecting the use of this promising carbon source for sustainable bioprocesses. Although the molecular mechanisms underlying *Saccharomyces cerevisiae* response and adaptation to acetic acid have been studied for years, only recently they have been examined in more detail in *Zygosaccharomyces bailii*. However, due to its remarkable tolerance to acetic acid and other weak acids this yeast species is a major threat in the spoilage of acidic foods and beverages and considered as an interesting alternative cell factory in Biotechnology. This review paper emphasizes genome-wide strategies that are providing global insights into the molecular targets, signaling pathways and mechanisms behind *S. cerevisiae* and *Z. bailii* tolerance to acetic acid, and extends this information to other weak acids whenever relevant. Such comprehensive perspective and the knowledge gathered in these two yeast species allowed the identification of candidate molecular targets, either for the design of effective strategies to overcome yeast spoilage in acidic foods and beverages, or for the rational genome engineering to construct more robust industrial strains. Examples of successful applications are provided.

**Keywords:** *Saccharomyces cerevisiae*, *Zygosaccharomyces bailii*, weak acid food preservatives, acetic acid adaptive response, acetic acid tolerance, physiological genomics

## INTRODUCTION

The yeast *Saccharomyces cerevisiae* plays an essential role in the production of foods (e.g., bread) and alcoholic beverages (e.g., wine and beer). However, this yeast species is also a food spoilage agent, being able to overcome several harsh conditions that are employed in the food industry to maintain the microbial stability of its products and avoid undesirable changes in their organoleptic properties (James and Stratford, 2003).

Yeasts belonging to the genus *Zygosaccharomyces* are associated with a detrimental role in food and beverage industries, being considered the most problematic food spoilage yeasts. In fact, they

are able to adapt and proliferate in the presence of extremely high concentrations of weak acids (*Zygosaccharomyces bailii* and *Zygosaccharomyces lentus*), sugar and salt (*Z. rouxii*) compared to those tolerated by other spoilage yeasts (James and Stratford, 2003). Within the genus, *Z. bailii* stands out as the most problematic spoilage yeast, mainly in acidified food products, such as mayonnaise, salad dressings, fruit concentrates and various non-carbonated fruit drinks, also being frequently isolated in wine due to its tolerance to both organic acids at low pH and ethanol (Thomas and Davenport, 1985; James and Stratford, 2003). *Z. bailii* is also an emerging spoiler of new food types such as mustards and fruit-flavored carbonated soft drinks (Sá-Correia et al., 2014). The remarkable tolerance of *Z. bailii* to weak acid food preservatives allows growth to occur in food products with concentrations above those legally permitted (Sá-Correia et al., 2014). Depending on the food product, the limit concentrations approved for use of sorbic and benzoic acids as food additives mainly range from 0.5 to 2 g/L (European Commission, 2011). Concerning the use of acetic acid as a food additive, the concentration is *quantum satis* (European Commission, 2011) this meaning that acetic acid should be used in food products under conditions that do not result in consumer's deception. In the case of *Z. bailii*, the average minimum inhibitory concentration (MIC) determined for several strains is approximately 8 and 10 g/L (pH 4.0) for sorbic and benzoic acids, respectively, and around 28 g/L (pH 4.0) for acetic acid, which are much higher than the values commonly determined for *S. cerevisiae* (Stratford et al., 2013b). Those different tolerance levels are highly relevant also because, despite their widespread use and classification as "generally recognized as safe" (GRAS), weak acids may cause intolerance (Joneja, 2003; Stratford, 2006; Theron and Lues, 2010).

Acetic acid is also an important inhibitory byproduct of alcoholic fermentation carried out by *S. cerevisiae* (Garay-Arroyo et al., 2004; Graves et al., 2006) and can achieve levels that, combined with high concentrations of ethanol and other toxic metabolites, may lead to fermentation arrest or reduced ethanol productivity (Rasmussen et al., 1995; Garay-Arroyo et al., 2004; Graves et al., 2006). Moreover, acetic acid is a highly important inhibitory compound in the context of lignocellulosic hydrolysates-based bioethanol production where its presence may seriously affect fermentation performance (Jönsson et al., 2013). Concentrations of acetic acid in lignocellulosic hydrolysates strongly depend on the feedstock and on the severity of the pretreatment (Jönsson et al., 2013). Levels of 3.4 g/L (pH 5.0) can, for instance, be achieved in wheat straw hydrolysates (Olofsson et al., 2010). Although these concentrations are below *S. cerevisiae* MIC for acetic acid (around 9 g/L at pH 4.0) (Stratford et al., 2013b), it is the combined effect of acetic acid and several other compounds produced during pretreatment of lignocellulosic hydrolysates that inhibits *S. cerevisiae* fermentation performance. It is therefore essential to understand the mechanisms underlying *S. cerevisiae* tolerance to acetic acid in order to develop robust industrial strains.

Considering the importance of acetic acid as a yeast growth inhibitor in modern Biotechnology and Food Industry, this review paper provides an updated critical review of

scientific literature on the adaptive response and tolerance to this weak acid emphasizing the physiological toxicogenomics perspective. The understanding of yeast physiology exploring functional and comparative genomic strategies allows a holistic assessment of the complex adaptive responses to environmental stresses and the identification of tolerance or susceptibility determinants to these stresses at a genome-wide scale. Yeast physiological toxicogenomics is thus instrumental to guide synthetic pathway engineering and other approaches for cell robustness manipulation, either for the sustainable production of fuels and chemicals or for the control of spoiling yeasts.

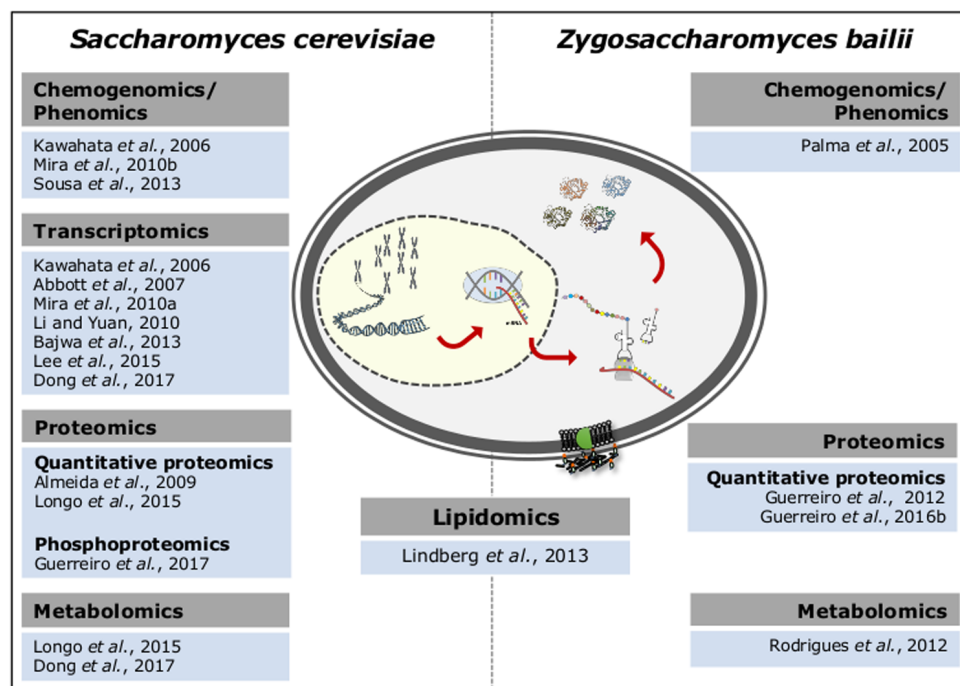
## MECHANISMS UNDERLYING THE ADAPTIVE RESPONSE AND TOLERANCE TO ACETIC ACID IN YEASTS

### The Physiological Genomic Approaches

Upon exposure to inhibitory, but sublethal, concentrations of acetic acid, yeast cells may enter a more or less extended period of growth arrest, but after this adaptation period exponential growth is resumed with a lower maximum specific growth rate (Fernandes et al., 2005; Guerreiro et al., 2012). On the other hand, lethal concentrations of acetic acid may induce regulated cell death (RCD), either by apoptosis or necrosis, depending on the severity of acetic acid stress (Ludovico et al., 2001, 2003).

After two decades of post-genomic research in *S. cerevisiae*, a more comprehensive understanding of the molecular mechanisms underlying this species response and adaptation to sublethal or lethal concentrations of acetic acid was obtained through the integration of several functional genomic approaches (**Figure 1**). When compared to *S. cerevisiae*, the number of Omic-based approaches applied to *Z. bailii* is still limited (**Figure 1**). Among other reasons, the lack of a *Z. bailii* genome sequence with suitable annotation has limited those studies for years. However, since 2013 the annotated genome sequences of two *Z. bailii* strains (Galeote et al., 2013; Palma et al., 2017b) and two hybrid strains resulting from *Z. bailii* and an unidentified *Zygosaccharomyces* species were also disclosed (Mira et al., 2014; Ortiz-Merino et al., 2017). This genome data has largely accelerated the understanding of *Z. bailii* species as a biological system with interesting genetic and physiological traits. Moreover, it has provided genomic information essential to study the mechanisms underlying *Z. bailii* tolerance to acetic acid, yet to be explored.

In *S. cerevisiae*, two chemogenomics screenings of commercial haploid yeast mutant collections containing thousands of single deletion mutants in non-essential yeast genes were useful to identify candidate molecular determinants and mechanisms of tolerance to sublethal concentrations of acetic acid (Kawahata et al., 2006; Mira et al., 2010b). These screenings, that were performed under different experimental conditions, led to the identification of a number of genes required for maximum tolerance to acetic acid, but the two datasets only share a total of 150 genes (Mira et al., 2010c). These include, for example, genes involved in intracellular pH homeostasis, vacuolar transport,



**FIGURE 1 |** Functional genomics-based approaches explored to obtain mechanistic insights into the adaptive response and tolerance to sublethal and lethal concentrations of acetic acid in *S. cerevisiae* and *Z. bailii*.

positive regulation of cellular processes, amino acid and carbohydrate metabolism (Kawahata et al., 2006; Mira et al., 2010b). Metabolic pathways that regulate RCD in *S. cerevisiae* response to lethal concentrations of acetic acid were also revealed based on a chemogenomics study, with carbohydrate metabolism emerging as an essential regulator of RCD (Sousa et al., 2013).

Given the unavailability of an equivalent single deletion mutant collection for *Z. bailii*, the identification of the determinants of tolerance to acetic acid was attempted using a genomic library of the *Z. bailii*-derived hybrid strain ISA1307 (Palma et al., 2015). This library was used to rescue the high acetic acid susceptibility phenotype of *S. cerevisiae* *haa1* deletion mutant (Palma et al., 2015). *Z. bailii* genes putatively involved in cellular transport and transport routes, protein fate, protein synthesis, amino acid metabolism and transcription were proposed as strong candidate determinants of acetic acid tolerance in *Z. bailii* (Palma et al., 2015).

The transcriptional alterations occurring in *S. cerevisiae* cells challenged with acetic acid stress were examined in several studies performed under different experimental settings, such as different cell growth phase, acetic acid concentration and medium composition and pH (Kawahata et al., 2006; Abbott et al., 2007; Li and Yuan, 2010; Mira et al., 2010a; Bajwa et al., 2013; Lee et al., 2015; Dong et al., 2017). Transcriptional profiling of the early response of *S. cerevisiae* cells to acetic acid (Kawahata et al., 2006; Li and Yuan, 2010; Mira et al., 2010a) or after adaptation to this weak acid (Kawahata et al., 2006; Bajwa et al., 2013; Lee et al., 2015) were described. Response to acetic acid of steady-state anaerobic and glucose-limited chemostat *S. cerevisiae* cultures

was also investigated (Abbott et al., 2007). Although the genes identified in different studies as up-regulated under acetic acid stress do not fully overlap, a few genes coding for plasma membrane proteins or proteins of a still unidentified function emerged in, at least, three of those works.

A quantitative proteomics analysis based on two-dimensional gel electrophoresis (2-DE) contributed to the identification of the alterations in the protein content of *Z. bailii* hybrid strain ISA1307 occurring in response to sudden exposure or during exponential growth in the presence of an inhibitory sublethal concentration of acetic acid (Guerreiro et al., 2012). The increased content of a particular set of proteins suggested that, in the presence of glucose, acetate is channeled into the tricarboxylic acid cycle, being co-consumed with glucose. These results were corroborated by a metabolomics study where the biochemical pathways associated with acetic acid utilization during co-metabolism with glucose were investigated in the same strain (Rodrigues et al., 2012). In *Z. bailii*, a quantitative proteomics study for the analysis of expression of mitochondrial proteins in cells exposed to lethal concentrations of acetic acid highlighted the importance of metabolic and energy processes, in particular of mitochondrial energetic metabolism, in acetic acid-induced RCD response (Guerreiro et al., 2016b). Cellular processes like oxidative stress response, protein translation, amino acid (in particular glutamate) and nucleotide metabolism, among others, were also found to be involved in this cellular response (Guerreiro et al., 2016b). In *S. cerevisiae*, quantitative proteomics was used to examine yeast response to a lethal concentration of acetic acid, revealing alterations in the levels

of proteins implicated in the general amino-acid control system, further shown to be associated with a severe intracellular amino-acid starvation, as well as in the Target-of-rapamycin (TOR) pathway (Almeida et al., 2009). Moreover, quantitative proteomic and metabolomic analyses of *S. cerevisiae* parental and derived deletion mutant *yca1* strains in acetic acid-induced RCD identified significant alterations in carbohydrate catabolism, lipid metabolism, proteolysis and stress-response, thus emphasizing the importance of Yca1 metacaspase in RCD caused by lethal concentrations of acetic acid (Longo et al., 2015).

The alterations occurring at the level of the membrane phosphoproteome during yeast early adaptive response to a sublethal concentration of acetic acid stress and the role played by the Hrk1 kinase in such response have recently been investigated (Guerreiro et al., 2017). Hrk1 is a protein kinase belonging to the “Npr1-family” of kinases dedicated to the regulation of plasma membrane transporters that was identified in previous Omics approaches as a determinant of acetic acid tolerance and involved in yeast response to acetic acid stress (Abbott et al., 2007; Mira et al., 2010a,b). The investigation of membrane phosphoproteome hinted toward the contribution of phosphorylation in the regulation of processes related with translation, protein folding and processing, transport, and cellular homeostasis in yeast response to acetic acid stress (Guerreiro et al., 2017).

The studies previously mentioned were exclusively dedicated to either *S. cerevisiae* or *Z. bailii* under acetic acid stress. In fact, high-throughput comparisons between *S. cerevisiae* and *Z. bailii* using similar experimental conditions are scarce in the context of acetic acid stress. One relevant exception is the lipidomic profiling of the major lipid species found in the plasma membrane of exponentially growing cells of both species under basal and acetic acid stress conditions (Lindberg et al., 2013). The correlation between the higher basal level of complex sphingolipids in *Z. bailii* when compared to *S. cerevisiae*, and consequent reduced plasma membrane permeability to acetic acid, was one of the most important findings in this study (Lindberg et al., 2013).

The exploitation of functional genomics approaches and tools, besides providing an integrative view on how yeast cells respond to a challenging environment, has strongly contributed to the understanding of the molecular players involved in such response. Nevertheless, the combined use of such global approaches with more focused molecular and cellular biology studies is essential to understand in depth the complexity of the mechanisms underlying the response and tolerance to a particular stress. In the following sections, the main molecular mechanisms underlying *S. cerevisiae* and *Z. bailii* response and tolerance to sublethal and lethal concentrations of acetic acid are reviewed based on genome-wide and high-throughput analyses complemented by more detailed molecular, biochemical and physiological studies.

## Acetic Acid Uptake and Toxicity

Acetic acid cellular uptake is dependent both on the extracellular pH and on the specific growth conditions. During *S. cerevisiae* growth in glucose-repressible conditions and at a pH below

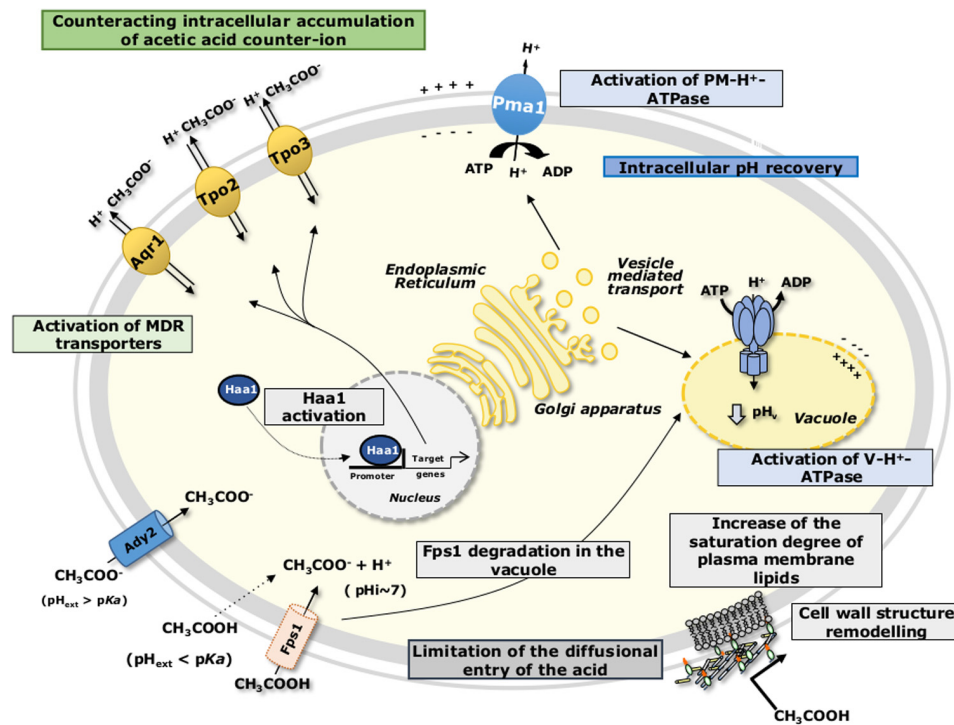
acetic acid  $pK_a$  ( $\approx 4.76$ ), acetic acid is mainly in its undissociated form,  $CH_3COOH$ , which is able to passively diffuse across the cell membrane lipid bilayer (Casal et al., 1996). Additionally, the aquaglyceroporin Fps1 was proposed to facilitate the uptake of acetic acid into the yeast cell (Mollapour and Piper, 2007) (Figure 2). Once inside the cell, at the near-neutral cytosol, acetic acid dissociates leading to the release of protons ( $H^+$ ) and of the negatively charged acetate counterion ( $CH_3COO^-$ ). During growth of derepressed *S. cerevisiae* cells or growth of cells at a pH higher than acetic acid  $pK_a$ , the dissociated form of the acid prevails, and the acid anion is transported at least by the acetate carrier Ady2 (Casal et al., 1996; Paiva et al., 1999, 2004) (Figure 2). Once inside the cell, acetate is unable to move back across the plasma membrane by simple diffusion, and accumulates in the cell interior causing increased turgor pressure and severe oxidative stress (Piper et al., 2001). On the other hand, reduction of intracellular pH ( $pH_i$ ) caused by the release of protons upon acetic acid dissociation leads to the inhibition of metabolic activity, among other deleterious effects (Pampulha and Loureiro-Dias, 1989, 1990; Orij et al., 2012).

Differently from *S. cerevisiae*, in *Z. bailii* cells cultivated in the presence of both glucose and acetic acid, simple diffusion of the undissociated form of the acid seems to have a minor contribution to the overall uptake of acetic acid (Sousa et al., 1996). Under such experimental conditions or if glucose or fructose are the sole available carbon sources, a non-glucose repressible acetic acid carrier is present and controlled by the intracellular concentration of acetate (Sousa et al., 1996, 1998) (Figure 3). When *Z. bailii* cells are cultivated in a medium containing acetic acid as the sole carbon source, acetic acid is transported by a saturable transport system, which is also able to transport propionic and formic acids (Sousa et al., 1996) (Figure 3). However, these carriers are not yet identified.

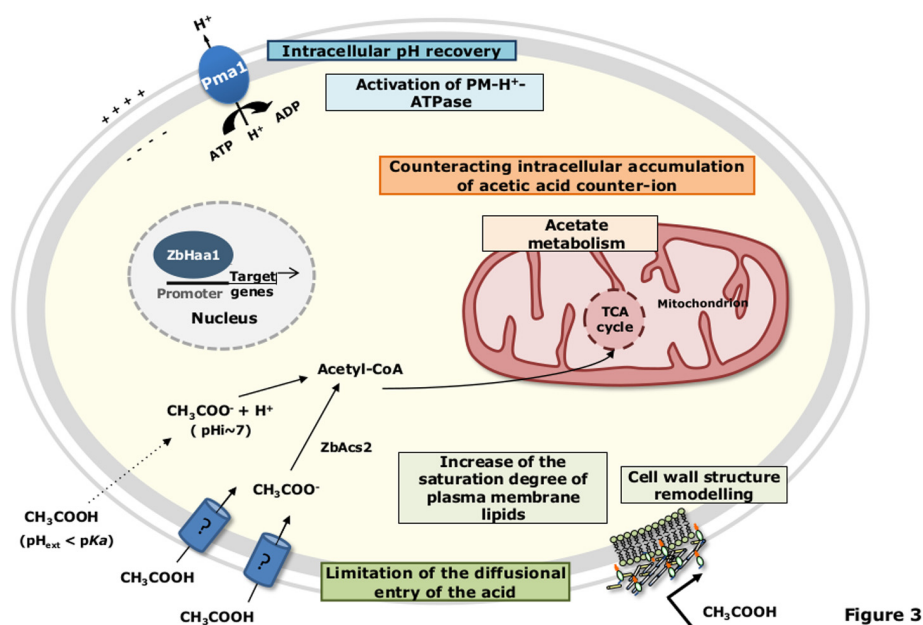
## Intracellular Acidification and pH Recovery

Acidification of the cytosol has been considered one of the major causes of *S. cerevisiae* growth inhibition upon acetic acid stress (Ullah et al., 2012; Stratford et al., 2013a). However, growth inhibition by acetic acid is related not with the initial intracellular acidification levels, but rather with cells' ability to recover a more physiological  $pH_i$  (Ullah et al., 2012). In order to counteract the dissipation of the  $H^+$  gradient across plasma membrane, the yeast cell is strongly dependent on the activation of the plasma membrane  $H^+$ -ATPase (mainly encoded by *PMA1* gene), in coordination with the activation of vacuolar membrane  $H^+$ -ATPase (Carmelo et al., 1997; Ullah et al., 2012; Stratford et al., 2013a) (Figure 2). The co-activation of these two proton pumps contributes to the active expulsion of protons accumulated in the cytosol upon acetic acid dissociation, both to the exterior of the cell and to the vacuole lumen (Carmelo et al., 1997; Stratford et al., 2013a) (Figure 2). The deletion of genes involved in  $pH_i$  homeostasis, as for example *VMA1*, *VMA2*, *VMA4-8*, and *VMA13* encoding vacuolar ATPase complex proteins confers susceptibility to *S. cerevisiae* cells exposed to sublethal concentrations of acetic acid (Kawahata et al., 2006;





**FIGURE 2** | Schematic model for the adaptive response of *S. cerevisiae* to acetic acid-induced stress as detailed in the text.  $\text{pH}_{\text{ext}}$ , external pH;  $\text{pH}_i$ , intracellular pH.



**Figure 3**

**FIGURE 3** | Schematic model for the adaptive response of *Z. bailii* to acetic acid-induced stress. Details are provided in the text.  $\text{pH}_{\text{ext}}$ , external pH;  $\text{pH}_i$ , intracellular pH; TCA, tricarboxylic acid.

Mira et al., 2010b), strongly supporting the idea of the important role of vacuolar ATPase in adaptation to acetic acid. The energy-dependent recovery of  $\text{pH}_i$  in acetic acid-stressed *S. cerevisiae* cells (Pampulha and Loureiro-Dias, 2000; Ullah et al., 2013),

together with the inhibition of glycolytic activity induced by the release of protons and subsequent cytoplasmic acidification (Pampulha and Loureiro-Dias, 1990), leads to a reduction in cellular ATP levels. Cells exposed to different levels of acetic

acid exhibit different responses in terms of intracellular ATP levels (Ullah et al., 2013). Under more severe acetic acid stress, causing complete growth inhibition, ATP levels were higher than those found in cells subjected to moderate stress, where growth was partially inhibited. This cell strategy of reducing ATP consumption upon severe stress was suggested to be advantageous to maintain energy reserves for later recovery of growth in more favorable conditions (Ullah et al., 2013).

Although the activation of plasma membrane or vacuolar membrane proton ATPases in *Z. bailii* cells exposed to acetic acid stress has never been studied, the activation of plasma membrane  $H^+$ -ATPase activity occurs under benzoic acid stress, leading to the efflux of  $H^+$  and simultaneous influx and accumulation of  $K^+$  (Macpherson et al., 2005). These ionic movements were described as well for *S. cerevisiae* in short-term preservative stress (Macpherson et al., 2005).

*Zygosaccharomyces bailii* was suggested to better tolerate short-term decrease of  $pH_i$  than *S. cerevisiae* (Arneborg et al., 2000), as well as significant  $pH_i$  drops during exponential phase of growth that are restored afterward in stationary phase (Dang et al., 2012). Apparently, *Z. bailii* employs different strategies to cope with different levels of acetic acid stress (Dang et al., 2012). For more inhibitory acetic acid concentrations,  $pH_i$  decreases and is maintained around the same value during exponential phase; recovery of  $pH_i$  to more physiological levels is registered during the stationary phase of growth (Dang et al., 2012). However, when the growth medium is supplemented with a milder inhibitory concentration of acetic acid, there is an initial moderate drop in  $pH_i$  that is maintained throughout growth, suggesting that cells adapt to the slightly lower  $pH_i$  to the point where  $pH_i$  recovery was not required (Dang et al., 2012).

In addition to the responses exhibited by the average yeast cell population, single cell-specific responses to acetic acid have also been investigated in both yeast species. Individual *Z. bailii* cells present in a population exposed to different weak acids exhibit variable tolerance to acetic, sorbic and benzoic acids (Steels et al., 2000; Stratford et al., 2013b). The most tolerant sub-population represents a small fraction of the bulk population and has a lower  $pH_i$ , which leads to reduced intracellular dissociation of any weak acid and, consequently, reduced accumulation of the counterion in the cytoplasm, thus conferring tolerance to any weak acid, but not to other type of inhibitors (Stratford et al., 2013b). This cross-tolerance phenomenon indicates that tolerance is not dependent on the specific acid structure, but rather relies on a mechanism that decreases the uptake and/or accumulation of any weak acid.

Cell-to-cell heterogeneity was also implicated in *S. cerevisiae* tolerance to acetic acid since only the fraction of cells with low initial  $pH_i$  values was able to recover  $pH_i$  and resume growth in the presence of the acid (Fernández-Niño et al., 2015). However, the initial  $pH_i$  is not the only factor influencing *S. cerevisiae* tolerance to acetic acid as shown when two strains with different tolerances to this weak acid were compared (Fernández-Niño et al., 2015).

## Acetate Detoxification Mechanisms

Specific inducible transporters are presumably involved in the active expulsion of acetate from *S. cerevisiae* cell

interior. Among them are the plasma membrane transporters of the Major Facilitator Superfamily (MFS), involved in Multidrug/Multixenobiotic Resistance (MDR/MXR) (dos Santos et al., 2014), Tpo2 and Tpo3 (Fernandes et al., 2005; Kawahata et al., 2006; Mira et al., 2010b), and Aqr1 (Tenreiro et al., 2002) (Figure 2). Remarkably, *TPO2* was found to be activated in *S. cerevisiae* cells upon acetic acid stress in several genome-wide transcriptional profiling studies (Kawahata et al., 2006; Abbott et al., 2007; Mira et al., 2010a; Bajwa et al., 2013). In contrast to *S. cerevisiae*, no acetate export system has hitherto been suggested or described in *Z. bailii*. Moreover, there are significant differences between *S. cerevisiae* and *Z. bailii* concerning acetate catabolization. In *S. cerevisiae* the use of acetate as a carbon source is in general repressed by glucose, and cells cultivated in a medium containing both glucose and acetic acid exhibit diauxic growth with acetic acid only being consumed after glucose has been exhausted from the medium (Casal et al., 1996, 1998; Vilela-Moura et al., 2011). When acetic acid is the sole carbon source present in the medium, this weak acid is consumed through its conversion to acetyl coenzyme A (acetyl-CoA), catalyzed by two acetyl-CoA synthetases, Acs1 (peroxisomal) or Acs2 (cytosolic) (dos Santos et al., 2003). The acetyl-CoA produced from acetate then enters the mitochondria to be oxidized in the tricarboxylic acid (TCA) cycle or remains outside the mitochondria to be metabolized in the glyoxylate cycle, which replenishes the cell with succinate, a crucial metabolite to produce different biosynthetic precursors (dos Santos et al., 2003). *Z. bailii* has the ability to catabolize acetate even in the presence of glucose (Sousa et al., 1998; Guerreiro et al., 2012; Rodrigues et al., 2012). This mechanism involves the regulation of both acetic acid membrane transport and acetyl-CoA synthetase activity, allowing the maintenance of an intracellular concentration of the acetate below toxic levels (Rodrigues et al., 2004; Sousa et al., 1996). Other proteins related to carbohydrate metabolism (Mdh1, Aco1, Cit1, Idh2 and Lpd1) and energy generation (Atp1 and Atp2), as well as general and oxidative stress response (Sod2, Dak2, Omp2), are also involved in acetate catabolization in *Z. bailii* (Guerreiro et al., 2012). This strengthens the concept that glucose and acetic acid are co-consumed in *Z. bailii*, with acetate being channeled into the TCA. This behavior was hypothesized to contribute to the remarkable tolerance of *Z. bailii* to acetic acid, particularly in environments that are rich in both carbon sources, such as during vinification (Sousa et al., 1998) or fermentation of lignocellulosic hydrolysates (Jönsson et al., 2013).

## Acetic Acid-Induced Alterations of the Cellular Envelope

It has been proposed that remodeling of cell wall and/or plasma membrane structure may occur and represent one of the most important adaptive mechanisms of tolerance to weak acids (Figures 2, 3) (Simões et al., 2006; Mollapour et al., 2009; Lindberg et al., 2013; Guerreiro et al., 2016a). Indeed, reducing cellular envelope permeability by altering cell wall and plasma membrane chemical structure and properties, thereby decreasing weak acid diffusion, is a much more energetically efficient method than relying on the active extrusion of protons and

acid. Consistent with this concept, multiple genes involved in the synthesis of cell wall polysaccharides, cell wall structure assembly and remodeling, and in sphingolipid and sterol biosynthetic pathways were demonstrated to be determinants of tolerance, or to be transcriptionally responsive to acetic acid in *S. cerevisiae*, in several genome-wide studies (Kawahata et al., 2006; Abbott et al., 2007; Mira et al., 2010a,b; Lee et al., 2015; Longo et al., 2015).

Among the proteins found to mediate the alteration of cell wall structure in response to weak acids in *S. cerevisiae*, is Spi1, a glycosylphosphatidylinositol-anchored cell wall protein, which is particularly important for tolerance to lipophilic acids like benzoic or octanoic (Simões et al., 2006). Although the expression of *SPI1* was considered not significant in the adaptation and tolerance to acetic acid (Simões et al., 2006), *SPI1* transcription was activated in response to this acid in several transcriptional analyses (Kawahata et al., 2006; Abbott et al., 2007; Mira et al., 2010a), thereby suggesting Spi1 as an important player in response to acetic acid. Although the specific function of *YGPI*, encoding a cell wall-related secretory glycoprotein, is still unknown, this gene is also considered a key player in *S. cerevisiae* tolerance to acetic acid, since it is activated in cells exposed to acetic acid (Kawahata et al., 2006; Abbott et al., 2007; Mira et al., 2010a) and the mutant with this gene deleted is very susceptible to this acid (Mira et al., 2010b). *Z. bailii* *YGPI* homologue was also found to be up-regulated in response to acetic acid (Palma et al., 2017a).

Major alterations occurring in the plasma membrane lipid composition are also involved in yeast tolerance to acetic acid. A lipidomic analysis revealed that, for both *S. cerevisiae* and *Z. bailii*, during aerobic growth in bioreactors, the supplementation of the growth medium with acetic acid induced significant changes in the cellular lipid content (Lindberg et al., 2013). In acetic acid-stressed *Z. bailii* cells, the total amount of glycerophospholipids (GPLs) was found to be slightly lower than in *S. cerevisiae*, while the degree of saturation of GPLs was increased under these same conditions (Lindberg et al., 2013). Increased levels of complex sphingolipids were detected for both species in the mid-exponential phase of acetic acid-adapted growth (Lindberg et al., 2013). Remarkably, the basal level of complex sphingolipids was significantly higher in *Z. bailii* than in *S. cerevisiae* leading the authors to suggest a link between high sphingolipid levels and the intrinsic tolerance of *Z. bailii* species to acetic acid (Lindberg et al., 2013). The correlation between the fraction of sphingolipids and membrane permeability to acetic acid was further investigated and confirmed based on *in silico* simulations of model membranes (Lindhöj et al., 2016). Sphingolipids are essential structural components of cellular membranes, in particular the plasma membrane, playing important roles in signaling and intracellular trafficking, as well as in the regulation of diverse processes (Dickson, 2008). The importance of the regulation of sphingolipid biosynthetic pathway in *S. cerevisiae* response and tolerance to acetic acid was more recently examined during *S. cerevisiae* early adaptive response to acetic acid (Guerreiro et al., 2016a). It was demonstrated that Ypk1 phosphorylation and activation by the membrane-localized protein kinase complex TORC2 is stimulated in response to acetic acid

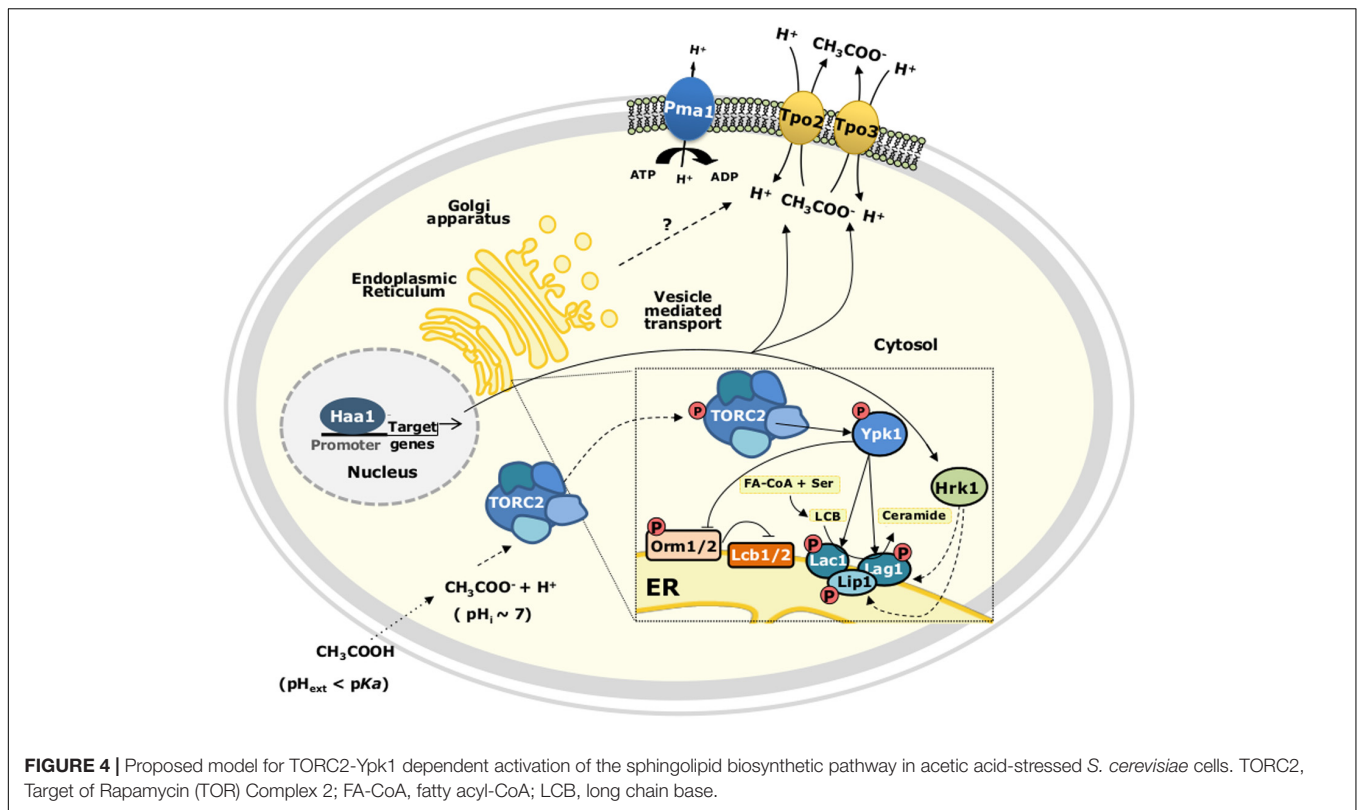
stress, consequently activating lipid synthesis (Guerreiro et al., 2016a) (Figure 4). Several plasma membrane lipid and protein homeostasis processes are regulated by the protein kinase Ypk1 (reviewed in Roelants et al., 2017). For instance, TORC2/Ypk1 signaling was proposed to inactivate by phosphorylation the endoplasmic reticulum-associated Orm1/2 protein inhibitors of L-serine:palmitoyl-CoA acyltransferase enzyme complex Lcb1/2 that catalyzes the first step of sphingolipid biosynthesis in response to compromised sphingolipid synthesis, thus restoring sphingolipid biosynthesis (Roelants et al., 2011). In *S. cerevisiae* cells exposed to acetic acid Ypk1 phosphorylates Orm1 and two functionally redundant isoforms of the ceramide synthase complex, Lac1 and Lag1 (Guerreiro et al., 2016a) (Figure 4).

In what concerns ergosterol content in acetic acid-challenged cells, the levels of this sterol decreased in *S. cerevisiae*, but not in the more tolerant species *Z. bailii* (Lindberg et al., 2013). Ergosterol is the major sterol present in the plasma membrane and has been shown to have vital functions in *S. cerevisiae* cells, affecting its membrane fluidity and permeability (Abe and Hiraki, 2009). In fact, the level of ergosterol present in the plasma membrane plays a crucial role in *S. cerevisiae* tolerance to several stresses (Swan and Watson, 1998; Higgins et al., 2003; Dupont et al., 2011; Henderson and Block, 2014), and multiple genes involved in sterols biosynthetic pathways were demonstrated to be determinants of tolerance or to be transcriptionally responsive to acetic acid (Mira et al., 2010a,b).

## TRANSCRIPTIONAL REGULATORY NETWORKS CONTROLLING ADAPTIVE RESPONSE AND TOLERANCE TO ACETIC ACID

### Genome-Wide Transcriptional Regulation Induced by Acetic Acid Stress in *S. cerevisiae*

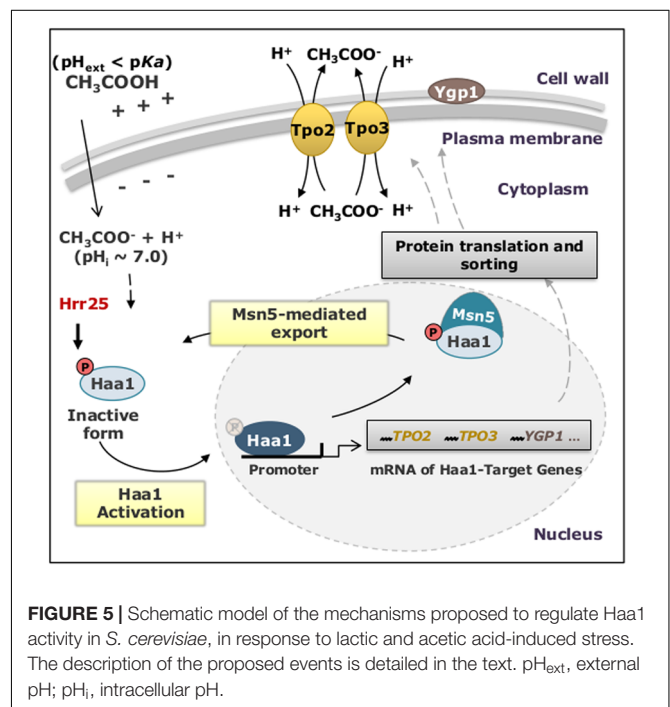
Yeast response to stressful conditions relies on the activation of general or specific regulatory pathways that determine cell fate, either to adapt and survive in the hostile environment or to die. *S. cerevisiae* cells respond to several different external insults by altering the transcription level of a particular set of genes of the generally called Environmental Stress Response (ESR), mainly controlled by Msn2 and Msn4 transcriptional activators (Gasch et al., 2000). The activation of genes from the ESR program was also registered when *S. cerevisiae* cells are exposed to several stresses relevant in Food Industry and Industrial Biotechnology, in particular in the response to acetic acid (Kawahata et al., 2006; Abbott et al., 2007; Li and Yuan, 2010; Mira et al., 2010a) and/or other weak acids (Schüller et al., 2004; Abbott et al., 2007; Mira et al., 2009). In addition to the general stress response transcription factors Msn2/Msn4, other regulators are involved in the genome-wide transcriptional response of *S. cerevisiae* to weak acid-induced stress (reviewed in Mira et al., 2010c; Teixeira et al., 2011). Specifically, War1, responsible for the induction of *PDR12* transcription and therefore being crucial for response and tolerance to propionic, benzoic and sorbic



acids (Piper et al., 2001; Schüller et al., 2004), Rim101, required for maximal tolerance to weak acid-induced stress, including acetic acid (Mira et al., 2009), and Haa1, required for adaptation and tolerance to the more hydrophilic formic, acetic, lactic and propionic acids (Fernandes et al., 2005; Abbott et al., 2007; Henriques et al., 2017). In yeast cells exposed to stresses that lead to mitochondrial dysfunction, it is the mitochondrial retrograde (RTG) signaling pathway that establishes mitochondria-to-nucleus communication, regulating the necessary alteration of nuclear gene expression (Butow and Avadhani, 2004), which is mainly controlled by the transcriptional regulators Rtg1 and Rtg3 (Sekito et al., 2000). The integration of genome-wide data from transcriptomic profiling of *S. cerevisiae* response and tolerance to several weak acids has contributed to the understanding of cellular responses to weak acid-induced stress as a dynamic system, where each transcription factor-associated network can cross-talk with others (Mira et al., 2010c). Since Haa1 is considered the primary regulator of *S. cerevisiae* transcriptional response to acetic acid (Mira et al., 2010a), the next two sections are dedicated to this transcription factor and to its *Z. bailii* ortholog ZbHaa1.

## The Haa1 Regulon as the Main Player in the Control of *S. cerevisiae* Response to Acetic Acid

Haa1 was first identified based on the homology and structural similarity with the DNA binding domain (DBD) of the copper-regulated transcription factor Cup2 (alias Ace1) (Hu et al., 1990;



Keller et al., 2001). The paralog pair Haa1 and Cup2 DBDs comprise 123 and 124 amino acid residues, respectively, at the N-terminal and include a conserved zinc module and a set of four cysteine-cysteine clusters organized in a consensus sequence that



forms the copper regulatory domain (CuRD). Such conservation at the level of the DBD led to the hypothesis that, like Cup2, Haa1 could play a role in copper homeostasis (Keller et al., 2001). However, metalloregulation and the involvement of Haa1 in *S. cerevisiae* tolerance to copper could not be associated to this transcription factor (Keller et al., 2001). Later, it was attributed for the first time a function to Haa1 as having an essential role in *S. cerevisiae* adaptation and tolerance to weak acids, especially to short-chain hydrophilic acids such as acetic acid (Fernandes et al., 2005). The alterations detected in yeast genomic expression during early response to acetic and lactic acids highlighted the involvement of Haa1 in the transcriptional reprogramming of *S. cerevisiae* cells during the adaptive response to these weak acids (Abbott et al., 2007; Mira et al., 2010a). Haa1 is required for the transcriptional activation of approximately 80% of the acetic acid-responsive genes, and thus proposed as being the main player in yeast genomic expression regulation under acetic acid stress (Mira et al., 2010a). Following Haa1-responsive element (HRE) identification (Mira et al., 2011), HRE was found to be present in the promoter region of about 55% of the genes whose expression is activated in response to acetic acid under the dependence of Haa1, suggesting that these genes are direct targets of this transcription factor (Mira et al., 2010a). The remaining Haa1-dependent acetic acid responsive genes are presumably indirectly regulated by Haa1 (Mira et al., 2010a, 2011), for example through the action of other genes encoding transcription factors directly regulated by Haa1 (Mira et al., 2010a, 2011). This is the case of Msn4, mediating the general stress response in yeast (Gasch et al., 2000), Fkh2, implicated in yeast response to oxidative stress (Postnikoff et al., 2012), and the transcriptional repressor Nrg1, also involved in yeast response to several stresses (Vyas et al., 2005; Mira et al., 2010a, 2011). These regulatory associations dependent on Haa1 upon weak acid-induced stress were highlighted recently in the upgrade of YEASTRACT database (Teixeira et al., 2017).

The biological activity of Haa1 was found to be regulated by its sub-cellular localization that, in turn, is regulated by Haa1 phosphorylation levels (Sugiyama et al., 2014). The rapid translocation of Haa1 from the cytosol to the nucleus, where it activates the transcription of its target-genes in response to lactic (Sugiyama et al., 2014) or acetic (Swinnen et al., 2017) acids, is concomitant with a decrease in Haa1 phosphorylation levels (Sugiyama et al., 2014). The casein kinase I isoform Hrr25 is an important negative regulator of Haa1, inhibiting this transcription factor's activity by phosphorylation (Collins et al., 2017). It was also demonstrated that the exportin Msn5, which preferentially exports phosphorylated cargo proteins, interacts with Haa1 being essential to its exit from the nucleus where its function as transcription factor takes place (Sugiyama et al., 2014) (Figure 5).

Among the genes of the Haa1-regulon whose expression was found to confer yeast protection against acetic acid are protein kinases, proteins involved in lipid metabolism (sphingolipids most notably ceramides, which are bioactive signaling molecules known to play a crucial role in lipid-based signaling in yeast response to stress), in nucleic acid processing, multidrug resistance transporters and proteins of unknown function.

The expression of *SAP30* (encoding a subunit of the Rpd3L histone deacetylase complex) and *HRK1* provided the strongest protective effect toward acetic acid (Mira et al., 2010a). Even though the first biological function attributed to Hrk1 was the activation of the plasma membrane proton pump Pma1 in response to glucose metabolism (Goossens et al., 2000), its mild effect on Pma1 activity suggests that this kinase could play other roles in yeast. The fact that *hrk1Δ* cells showed a marked increase in the intracellular concentration of radiolabeled acetic acid and that Hrk1 belongs to a family of kinases dedicated to the regulation of plasma membrane transporters, led to hypothesize that Hrk1 could be involved in the regulation of one or more plasma membrane acetate exporters, such as the MDR/MXR transporters Tpo3, Tpo2 and Aqr1, all known determinants of tolerance to this weak acid (Mira et al., 2010a,b). To elucidate the biological role of Hrk1, the effect of Hrk1 expression in yeast membrane associated-phosphoproteome was examined in *S. cerevisiae* parental and *hrk1Δ* cells exposed, or not, to sublethal concentrations of acetic acid (Guerreiro et al., 2017). In this study, the MDR/MXR transporters Tpo3 and Tpo4 were found to exhibit altered phosphorylation in response to acetic acid stress under the dependence of Hrk1. This evidence led to hypothesize that the Hrk1-mediated phosphorylation of Tpo3 may contribute to regulate the activity of this drug pump (Guerreiro et al., 2017), thereby lowering the accumulation of acetic acid in the parental strain when compared to *hrk1Δ* or *tpo3Δ* cells, as previously described (Fernandes et al., 2005; Mira et al., 2010a). Other important membrane proteins in the context of response and tolerance to acetic acid-induced stress are present in the published dataset and certainly deserve further studies (Guerreiro et al., 2017).

## The *Z. bailii* Haa1 (ZbHaa1) Is Required for Acetic Acid and Copper Stress Responses

The importance of ZbHaa1 in *Z. bailii* response and tolerance to acetic acid was demonstrated using the strain *Z. bailii* IST302 that proved to be susceptible to genetic engineering and has the genome fully sequenced (Palma et al., 2015, 2017a). ZbHaa1 was found to be a functional homolog of Haa1 by rescuing the acetic acid susceptibility phenotype of *S. cerevisiae* *haa1Δ*. Moreover, the disruption of *ZbHAA1* and the expression of an extra *ZbHAA1* copy in *Z. bailii* confirmed *ZbHAA1* as a determinant of acetic acid tolerance in this yeast species. The expression of *ZbHAA1* was found to be required for acetic acid stress-induced transcriptional activation of *Z. bailii* genes homologs to the demonstrated *S. cerevisiae* Haa1-target genes: *HRK1*, *TPO3*, *MSN4*, *YGP1*, *YRO2* and *HSP30* (Palma et al., 2017a). Remarkably, ZbHaa1 (the single ortholog of *S. cerevisiae* Haa1 and Cup2) was demonstrated to have a role in metalloregulation, being involved in copper tolerance and copper-induced transcriptional regulation, a role associated to *S. cerevisiae* Cup2, but not to Haa1 (Palma et al., 2017a). Phylogenetic and gene neighborhood analyses suggested the subfunctionalization of *Z. bailii* ancestral bifunctional protein Haa1/Cup2 after the whole-genome duplication event,

originating *S. cerevisiae* Haa1 and Cup2 paralogs. As found for *S. cerevisiae*, ZbHaa1 is likely a candidate molecular target for the design of new strategies to overcome *Z. bailii* spoilage in foods and beverages.

## MECHANISMS INVOLVED IN YEAST RESPONSE TO LETHAL CONCENTRATIONS OF ACETIC ACID

High concentrations of acetic acid are able to induce in *Z. bailii* either an apoptotic or a necrotic death process, depending on the acid concentration present in the medium, as observed for *S. cerevisiae* (Ludovico et al., 2001, 2003). However, since *Z. bailii* is highly resistant to acetic acid-induced cell death, this effect is observed at much higher concentrations of the acid for this yeast species (range of 320–800 mM, pH 3.0) than in *S. cerevisiae* (20–120 mM, pH 3.0) (Ludovico et al., 2001, 2002, 2003). The main known mechanisms involved in acetic acid-induced RCD were also much more extensively studied in *S. cerevisiae* than in *Z. bailii*.

In *S. cerevisiae*, acetic acid is known to induce a RCD process with an apoptotic phenotype that is dependent on mitochondria. Indeed, apart from its bioenergetic function, mitochondria have an essential role in the decision of cells' life or death (reviewed in Ždravlečić et al., 2012) and many of the hallmarks involved in yeast mitochondria-dependent RCD process with an apoptotic phenotype have been characterized, including the translocation of pro-apoptotic factors (e.g., cytochrome *c*) from the mitochondria to the cytosol, phosphatidylserine externalization to the outer layer of the cytoplasmic membrane, production and consequent build-up of mitochondrial reactive oxygen species (ROS), DNA fragmentation and chromatin condensation (reviewed in Carmona-Gutierrez et al., 2010; Guaragnella et al., 2012).

Concerning tolerance to acetic acid-induced RCD, the activation of the RTG signaling pathway was proposed to be involved when *S. cerevisiae* cells were cultivated in a non-repressible carbon source such as raffinose (Guaragnella et al., 2013) or previously adapted to low pH environments (Giannattasio et al., 2005). In yeast, the mitochondrial RTG pathway acts in parallel with the TOR and the Ras-cAMP pathways in the regulation of acetic acid-induced RCD (Phillips et al., 2006; Almeida et al., 2009; Giannattasio et al., 2013).

In *Z. bailii* the morphological changes observed during RCD induced by high concentrations of acetic acid show extensive mitochondrial ultrastructural changes during the RCD process that were not seen for *S. cerevisiae* when equivalent deleterious concentrations were used (Ludovico et al., 2003). The acetic acid-induced RCD process with an apoptotic phenotype was also characterized in *Z. bailii* by the maintenance of plasma membrane integrity, DNA fragmentation, ROS production, and cytochrome *c* translocation from the mitochondria into the cytosol (Ludovico et al., 2003; Guerreiro et al., 2016b). The global mitochondrial proteomic response to acetic acid-induced RCD in *Z. bailii* hybrid strain ISA1307 was examined by 2-DE quantitative proteomics (Guerreiro et al., 2016b). The

increase of different ROS (namely H<sub>2</sub>O<sub>2</sub> and superoxide anion) observed in *Z. bailii* cells undergoing RCD induced by acetic acid, coupled with different changes in abundance of several antioxidant enzymes observed in that same cell population, suggest that dynamic modulation of ROS might be taking place in cells exposed to acetic acid concentrations that induce RCD (Guerreiro et al., 2016b). Nevertheless, the effectors that play a role in acetic acid-induced RCD remain poorly characterized.

## PHYSIOLOGICAL GENOMICS-GUIDED STRATEGIES TO IMPROVE ACETIC ACID TOLERANCE

The topics covered in this review show that acetic acid tolerance phenotype is complex and multifactorial since it requires coordinated changes at several levels in the cell. For this reason, some of the most promising strategies being used to obtain more robust industrial strains involve the manipulation of genes that play a crucial role in the regulatory cascades controlling stress tolerance and the exploitation of genome engineering strategies, which allows the generation of diversity and subsequent selection of the strains that possess the trait of interest. The exploitation of this and other strategies are reviewed in the next sections.

### Overexpression or Deletion of Single Genes

The genetic engineering of laboratory *S. cerevisiae* strains through the overexpression of single genes has yielded strains with increased acetic acid tolerance. Specifically, the overexpression of *WHI2*, encoding a protein required for full activation of the general stress response (Chen et al., 2016b), of the genes *PEP3*, encoding a vacuolar membrane protein involved in vesicular tethering/docking/fusion, *STM1*, encoding a protein required for optimal translation under nutrient stress, *PEP5*, encoding the E3 ubiquitin protein-ligase involved in the catabolism of histones (Ding et al., 2015a), or the gene *ACS2*, encoding acetyl-coA synthetase isoform (Ding et al., 2015b), decreased the duration of lag phase of *S. cerevisiae* cells cultured with acetic acid. Improvement of *S. cerevisiae* growth and alcoholic fermentation performance in the presence of acetic acid also resulted from the overexpression of *SET5* and *PPR1*, coding for a methyltransferase for the methylation of histone H4 at Lys5, -8, and -12 and a transcription factor involved in the regulation of pyrimidine pathway, respectively (Zhang et al., 2015). The overexpression of *ASC1* (G-protein beta subunit), *GND1* (6-phosphogluconate dehydrogenase) (Lee et al., 2015), *PMA1* (Lee et al., 2017) or *COX20* (cytochrome oxidase chaperone) (Kumar et al., 2015) also increased *S. cerevisiae* tolerance to acetic acid.

*Saccharomyces cerevisiae* increased robustness to acetic acid has also been accomplished through expression of *Z. bailii* genes *ZbMSN4*, *ZbTIF3* (Palma et al., 2015), encoding the homologs of *S. cerevisiae* general stress response transcription factor and translation initiation factor, respectively, or *ZbHAA1* (Palma et al., 2017a). The overexpression of these genes in *Z. bailii* were

also found to increase its tolerance to acetic acid (Palma et al., 2015).

An increase in *S. cerevisiae* tolerance to acetic acid was also demonstrated by individual deletion of several genes. Specifically, the deletion of *HSP82* (protein chaperone), *ATO2* (putative transmembrane protein involved in export of ammonia) and *SSA3* (ATPase involved in protein folding and the response to stress) increased *S. cerevisiae* tolerance to acetic acid, presumably by contributing indirectly to enhanced proton export and diminished levels of  $H_2O_2$  within the cell upon acetic acid stress (Lee et al., 2015). The deletion of *RTT109* (histone acetyltransferase) also increased *S. cerevisiae* tolerance to acetic acid, which was suggested to be related to the activation of transcription of stress responsive genes and to increased resistance to oxidative stress (Cheng et al., 2016). In addition, the elimination of *JJJ1* (co-chaperone) enhanced acetic acid tolerance, which was proposed to be related to increased levels of long-chain fatty acids and trehalose in the cell, together with an increase in catalase activity (Wu et al., 2016). Remarkably, a single mutation in cytochrome *c* (the substitution of the highly conserved residue tryptophan 65 by a serine) that impaired electron transfer to the functional cyt *c* oxidase, was found to lead to increased cellular viability upon acetic acid stress (Guaragnella et al., 2011).

## Manipulation of Haa1-Regulon in *S. cerevisiae*

The manipulation of the Haa1-regulon, namely through *HAA1* overexpression (Tanaka et al., 2012; Inaba et al., 2013; Sakihama et al., 2015; Chen et al., 2016a) or *HAA1* mutation (Zahn and Jacobson, 2015; Swinnen et al., 2017) was shown to increase tolerance to acetic acid. In a first study, the *HAA1* gene was placed under the control of the constitutive *TDH3* promoter (*HAA1*-OP strain) (Tanaka et al., 2012). This promoter exchange led to increased *HAA1* transcript levels, as well as of four of the Haa1-regulated genes (*TPO2*, *TPO3*, *YRO2* and *YGP1*) even when cells were grown in the absence of acetic acid stress (Tanaka et al., 2012). The resulting strain, *HAA1*-OP, showed increased tolerance to acetic acid and decreased intracellular acetic acid accumulation when compared with the parental strain (Tanaka et al., 2012). A similar approach was used in an attempt to improve the industrial strain Ethanol-Red (ER). The resulting diploid ER *HAA1*-OP strain (with *HAA1* transcription being under the control of the *TDH3* promoter) showed increased *HAA1* transcription levels in non-stress conditions, and was also less susceptible to acetic acid-induced stress than the corresponding parental strain (Inaba et al., 2013). *HAA1* was also overexpressed from a multicopy plasmid in a successful attempt to increase ethanol productivity during xylose fermentation in the presence of acetic acid (Sakihama et al., 2015). Likewise, *HAA1* overexpression improved specific sugar consumption and cell growth rate in the presence of acetic acid when compared with the non-transformed strain (Chen et al., 2016a). Based on error-prone PCR of the *HAA1* coding sequence, a highly tolerant mutant allele holding two point mutations was identified as being able to improve the acetic acid tolerance of

*S. cerevisiae* (Swinnen et al., 2017). Following dissection of the individual contribution of each mutation, it was found that the major improvement in acetic acid tolerance was caused by a single amino acid exchange at position 135 (serine to phenylalanine). Remarkably, the transcriptional levels of four Haa1-target genes, selected at random, significantly increased in acetic acid-challenged cells of the strain harboring the single Haa1 mutation when compared to cells of the parental strain (Swinnen et al., 2017). In a patented work, mutagenesis of the Haa1 at the level of its transactivation domain was also found to increase the activity to this transcription factor, with a consequent increase in the yield of ethanol produced during fermentation of lignocellulose hydrolysates containing acetic acid (Zahn and Jacobson, 2015). All these studies suggest that the genetic manipulation of the Haa1 pathway is suitable to obtain more robust *S. cerevisiae* strains in an industrial context.

## *S. cerevisiae* Evolutionary Engineering and Genome Shuffling

The use of evolutionary engineering strategies, usually focusing on selection of a unique genetic trait responsible for that advantageous phenotype (Wright et al., 2011; Koppram et al., 2012), were also successfully employed for the improvement of *S. cerevisiae* tolerance to acetic acid. For example, a novel laboratory evolution strategy based on alternating cultivation cycles in the presence or absence of acetic acid was recently described as conferring a selective advantage to cells that are constitutively tolerant to acetic acid (González-Ramos et al., 2016). Mutations in four genes (*ASG1*, *ADH3*, *SKS1* and *GIS4*) were identified in this study as being implicated in the constitutive acetic acid tolerance phenotype of the evolved strains (González-Ramos et al., 2016). More recently, an evolutionary engineering study involving the cross of a strain with high acetic acid tolerance with an industrial reference strain, followed by multiple rounds of inbreeding of the resulting haploid segregants and quantitative trait loci (QTL) mapping, envisaged the study of the polygenic nature of the high acetic acid tolerance phenotype (Meijnen et al., 2016). This study allowed the identification of a mutated *HAA1* allele (serine to asparagine amino acid substitution at position 506) responsible for the superior character of the segregant strain under acetic acid stress (Meijnen et al., 2016). Among the novel genes identified in this study as contributing to high acetic acid tolerance is *DOT5*, encoding a nuclear thiol peroxidase and functioning as an alkyl-hydroperoxide reductase agent during post-diauxic growth. Remarkably, this gene also proved to play an important role in acetic acid tolerance in a study that also mapped the QTLs of segregants obtained from inbreeding of two industrial strains with distinct acetic acid tolerance phenotypes (Geng et al., 2016). Genome shuffling has also emerged as an alternative experiment strategy to improve tolerance to several stressors, including acetic acid (Zheng et al., 2011).

## *S. cerevisiae* Transcriptome Remodeling

The remodeling of the yeast transcriptome through global transcription machinery engineering has also been applied



successfully to obtain acetic acid-tolerant strains. This was achieved by re-programming the cell transcriptome through mutations in *SPT15*, encoding the TATA-binding protein, followed by screening of the mutants with improved tolerance to acetic acid (An et al., 2015). Another example involved the transformation of a yeast strain with an artificial zinc finger protein transcription factor library and subsequent selection of acetic acid-tolerant strains (Ma et al., 2015). Remodeling of transcription through introduction of point mutation in H3/H4 histones (Liu et al., 2014), as well as the generation of extensive alterations in mRNA metabolism through mutagenesis of the poly(A) binding protein encoding gene, *PAB1* (Martani et al., 2015), have also led to the development of strains with improved robustness against acetic acid stress.

## Supplementation of Growth Media with Cations

*Saccharomyces cerevisiae* tolerance to acetic acid can be alleviated by changing the composition of the growth media. The uptake of ions is essential in acetic acid tolerance, and several genes involved in ion homeostasis, for instance potassium transporters, were identified as determinants of tolerance to this weak acid (Kawahata et al., 2006; Mira et al., 2010b). This evidence led the authors to suggest and confirm that potassium supplementation of the growth medium may decrease acetic acid-induced growth inhibition (Mira et al., 2010b). Indeed, potassium is essential for many physiological functions, such as regulation of pH<sub>i</sub>, maintenance of plasma membrane potential, protein synthesis, and enzyme activation (Ariño et al., 2010), which are biological processes highlighted in this review as playing a role in response and tolerance to acetic acid. The increase of extracellular concentration of potassium was also found to be beneficial to increase the tolerance to higher alcohols and ethanol production in both commercial and laboratory strains (Lam et al., 2014). The supplementation of the growth medium with zinc sulfate also improved *S. cerevisiae* tolerance to acetic acid, with zinc presumably acting as an anti-oxidative agent (Wan et al., 2015). The addition of zinc and other metal ions (Mg<sup>2+</sup> and Ca<sup>2+</sup>) was also associated to an increase of *S. cerevisiae* tolerance toward acetic acid (Ismail et al., 2014). The comparison of the transcriptional profiling of cultures supplemented or not with each of the three metal ions and acetic acid suggested that these ions are involved in the regulation of different genes, and that the up-regulation of cell wall and membrane genes is related with the increased tolerance to acetic acid upon metal ion supplementation (Ismail et al., 2014).

## CONCLUDING REMARKS AND FUTURE PERSPECTIVES

The mechanisms underlying the adaptive response and tolerance to acetic acid in *S. cerevisiae* and *Z. bailii* have been enlightened over the past two decades based on functional and comparative genomics strategies. The knowledge gathered in the mechanisms of yeast response and tolerance to acetic acid have been mostly

explored in the development of acetic acid tolerant industrial strains, rather than in the design of novel food preservation technologies. Future manipulation of *S. cerevisiae* tolerance to acetic acid will continue to rely on increasing industrial strains robustness, either through the genetic manipulation of genes that play a crucial role in the regulatory cascades that control stress tolerance, or through genome-scale engineering, thereby allowing the generation of strain diversity and subsequent selection of the strains that possess the trait of interest. Nevertheless, caution is needed during the application of the abovementioned strategies, considering that the impact of the modifications introduced in the engineered strains must be carefully evaluated regarding the potential changes they might cause in important industrial properties of that strain (Deparis et al., 2017). Moreover, many of the successful improvement studies described so far have used low acetic acid tolerant laboratory strains, but their usefulness when applied to highly acetic acid tolerant industrial strains remains to be proved.

Understanding the mechanisms of tolerance to acetic acid in *S. cerevisiae* and *Z. bailii* has undoubtedly brought to light the interspecies diversity and complexity of those mechanisms, which rely on several molecular and physiological responses orchestrated by the expression of multiple genes. Although the model yeast *S. cerevisiae* has been at the forefront of numerous molecular, physiological and genome-wide studies on acetic acid response and tolerance, the release of *Z. bailii* genome annotated sequences and the availability of new *Z. bailii* strains more prone to genetic and laboratory manipulations and of new molecular genetic tools, such as the genome editing tool CRISPR/Cas9, is changing this paradigm. Extensive transcriptomic profiling studies are expected to emerge to characterize the transcriptional regulatory networks underlying *Z. bailii* response and adaptation to acetic acid stress. Given the described relevance of ZbHaa1, the identification and manipulation of the ZbHaa1-signaling pathway in acetic acid challenged *Z. bailii* cells is anticipated as a promising strategy to identify novel molecular targets in this food spoilage yeast species that can also be regarded as potential cell factory for the overproduction of organic acids.

The integrative perspective of the cellular processes described herein has benefited from the exploitation of a systems microbiology approach. They are granting the rational design of strategies to improve alcoholic fermentation processes when yeasts are used as microbial factories.

## AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

## FUNDING

The work performed in IS-C laboratory at iBB (Institute for Bioengineering and Biosciences) was supported by



‘Fundação para a Ciência e a Tecnologia’ (FCT) (current project contracts: PTDC/BBB-BEP/0385/2014, ERA-IB-2/0003/2015). Funding received by iBB from FCT (UID/BIO/04565/2013) and from Programa Operacional Regional de Lisboa 2020 (Project N. 007317) is also acknowledged.

## REFERENCES

- Abbott, D. A., Knijnenburg, T. A., de Poorter, L. M. I., Reinders, M. J. T., Pronk, J. T., and van Maris, A. J. A. (2007). Generic and specific transcriptional responses to different weak organic acids in anaerobic chemostat cultures of *Saccharomyces cerevisiae*. *FEMS Yeast Res.* 7, 819–833. doi: 10.1111/j.1567-1364.2007.00242.x
- Abe, F., and Hiraki, T. (2009). Mechanistic role of ergosterol in membrane rigidity and cycloheximide resistance in *Saccharomyces cerevisiae*. *Biochim. Biophys. Acta* 1788, 743–752. doi: 10.1016/j.bbame.2008.12.002
- Almeida, B., Ohlmeier, S., Almeida, A. J., Madeo, F., Leão, C., Rodrigues, F., et al. (2009). Yeast protein expression profile during acetic acid-induced apoptosis indicates causal involvement of the TOR pathway. *Proteomics* 9, 720–732. doi: 10.1002/pmic.200700816
- An, J., Kwon, H., Kim, E., Lee, Y. M., Ko, H. J., Park, H., et al. (2015). Tolerance to acetic acid is improved by mutations of the TATA-binding protein gene. *Environ. Microbiol.* 17, 656–669. doi: 10.1111/1462-2920.12489
- Ariño, J., Ramos, J., and Sychrová, H. (2010). Alkali metal cation transport and homeostasis in yeasts. *Microbiol. Mol. Biol. Rev.* 74, 95–120. doi: 10.1128/MMBR.00042-09
- Arneborg, N., Jespersen, L., and Jakobsen, M. (2000). Individual cells of *Saccharomyces cerevisiae* and *Zygosaccharomyces bailii* exhibit different short-term intracellular pH responses to acetic acid. *Arch. Microbiol.* 174, 125–128. doi: 10.1007/s002030000185
- Bajwa, P. K., Ho, C.-Y., Chan, C.-K., Martin, V. J. J., Trevors, J. T., and Lee, H. (2013). Transcriptional profiling of *Saccharomyces cerevisiae* T2 cells upon exposure to hardwood spent sulphite liquor: comparison to acetic acid, furfural and hydroxymethylfurfural. *Antonie Van Leeuwenhoek* 103, 1281–1295. doi: 10.1007/s10482-013-9909-1
- Butow, R. A., and Avadhani, N. G. (2004). Mitochondrial signaling: the retrograde response. *Mol. Cell* 14, 1–15. doi: 10.1016/S1097-2765(04)00179-0
- Carmelo, V., Santos, H., and Sá-Correia, I. (1997). Effect of extracellular acidification on the activity of plasma membrane ATPase and on the cytosolic and vacuolar pH of *Saccharomyces cerevisiae*. *Biochim. Biophys. Acta* 1325, 63–70. doi: 10.1016/S0005-2736(96)00245-3
- Carmona-Gutierrez, D., Eisenberg, T., Büttner, S., Meisinger, C., Kroemer, G., and Madeo, F. (2010). Apoptosis in yeast: triggers, pathways, subroutines. *Cell Death Differ.* 17, 763–773. doi: 10.1038/cdd.2009.219
- Casal, M., Cardoso, H., and Leão, C. (1996). Mechanisms regulating the transport of acetic acid in *Saccharomyces cerevisiae*. *Microbiology* 6, 1385–1390. doi: 10.1099/13500872-142-6-1385
- Casal, M., Cardoso, H., and Leão, C. (1998). Effects of ethanol and other alkanols on transport of acetic acid in *Saccharomyces cerevisiae*. *Appl. Environ. Microbiol.* 64, 665–668.
- Chen, Y., Sheng, J., Jiang, T., Stevens, J., Feng, X., and Wei, N. (2016a). Transcriptional profiling reveals molecular basis and novel genetic targets for improved resistance to multiple fermentation inhibitors in *Saccharomyces cerevisiae*. *Biotechnol. Biofuels* 9:9. doi: 10.1186/s13068-015-0418-5
- Chen, Y., Stabryla, L., and Wei, N. (2016b). Improved acetic acid resistance in *Saccharomyces cerevisiae* by overexpression of the *WHI2* gene identified through inverse metabolic engineering. *Appl. Environ. Microbiol.* 82, 2156–2166. doi: 10.1128/AEM.03718-15
- Cheng, C., Zhao, X., Zhang, M., and Bai, F. (2016). Absence of Rtt109p, a fungal-specific histone acetyltransferase, results in improved acetic acid tolerance of *Saccharomyces cerevisiae*. *FEMS Yeast Res.* 16:fow010. doi: 10.1093/femsyr/fow010
- Collins, M. E., Black, J. J., and Liu, Z. (2017). Casein kinase I isoform Hrr25 is a negative regulator of Haa1 in the weak acid stress response pathway in *Saccharomyces cerevisiae*. *Appl. Environ. Microbiol.* 83: e00672–17. doi: 10.1128/AEM.00672-17
- Dang, T. D. T., De Maeseneire, S. L., Zhang, B. Y., De Vos, W. H., Rajkovic, A., Vermeulen, A., et al. (2012). Monitoring the intracellular pH of *Zygosaccharomyces bailii* by green fluorescent protein. *Int. J. Food Microbiol.* 156, 290–295. doi: 10.1016/j.ijfoodmicro.2012.03.028
- Deparis, Q., Claes, A., Foulquié-Moreno, M. R., and Thevelein, J. M. (2017). Engineering tolerance to industrially relevant stress factors in yeast cell factories. *FEMS Yeast Res.* 17:fox036. doi: 10.1093/femsyr/fox036
- Dickson, R. C. (2008). Thematic review series: Sphingolipids. New insights into sphingolipid metabolism and function in budding yeast. *J. Lipid Res.* 49, 909–921. doi: 10.1194/jlr.R800003-JLR200
- Ding, J., Holzwarth, G., Bradford, C. S., Cooley, B., Yoshinaga, A. S., Patton-Vogt, J., et al. (2015a). *PEP3* overexpression shortens lag phase but does not alter growth rate in *Saccharomyces cerevisiae* exposed to acetic acid stress. *Appl. Microbiol. Biotechnol.* 99, 8667–8680. doi: 10.1007/s00253-015-6708-9
- Ding, J., Holzwarth, G., Penner, M. H., Patton-Vogt, J., and Bakalinsky, A. T. (2015b). Overexpression of acetyl-CoA synthetase in *Saccharomyces cerevisiae* increases acetic acid tolerance. *FEMS Microbiol. Lett.* 362, 1–7. doi: 10.1093/femsle/fnu042
- Dong, Y., Hu, J., Fan, L., and Chen, Q. (2017). RNA-Seq-based transcriptomic and metabolomic analysis reveal stress responses and programmed cell death induced by acetic acid in *Saccharomyces cerevisiae*. *Sci. Rep.* 7:42659. doi: 10.1038/srep42659
- dos Santos, M. M., Gombert, A. K., Christensen, B., Olsson, L., and Nielsen, J. (2003). Identification of in vivo enzyme activities in the cometabolism of glucose and acetate by *Saccharomyces cerevisiae* by using <sup>13</sup>C-labeled substrates. *Eukaryot. Cell* 2, 599–608. doi: 10.1128/EC.2.3.599-608.2003
- dos Santos, S. C., Teixeira, M. C., Dias, P. J., and Sá-Correia, I. (2014). MFS transporters required for multidrug/multixenobiotic (MD/MX) resistance in the model yeast: understanding their physiological function through post-genomic approaches. *Front. Physiol.* 5:180. doi: 10.3389/fphys.2014.00180
- Dupont, S., Beney, L., Ferreira, T., and Gervais, P. (2011). Nature of sterols affects plasma membrane behavior and yeast survival during dehydration. *Biochim. Biophys. Acta* 1808, 1520–1528. doi: 10.1016/j.bbame.2010.11.012
- European Commission (2011). *COMMISSION REGULATION (EU) No 1129/2011, amending Annex II to Regulation (EC) No 1333/2008 of the European Parliament and of the Council by Establishing a Union List of Food Additives. Official Journal of the European Union*. Brussels: European Commission.
- Fernandes, A. R., Mira, N. P., Vargas, R. C., Canelhas, I., and Sá-Correia, I. (2005). *Saccharomyces cerevisiae* adaptation to weak acids involves the transcription factor Haa1p and Haa1p-regulated genes. *Biochem. Biophys. Res. Commun.* 337, 95–103. doi: 10.1016/j.bbrc.2005.09.010
- Fernández-Niño, M., Marquina, M., Swinnen, S., Rodríguez-Porrata, B., Nevoigt, E., and Ariño, J. (2015). The Cytosolic pH of Individual *Saccharomyces cerevisiae* cells is a key factor in acetic acid tolerance. *Appl. Environ. Microbiol.* 81, 7813–7821. doi: 10.1128/AEM.02313-15
- Galeote, V., Bigey, F., Devillers, H., Neuvéglise, C., and Dequin, S. (2013). Genome sequence of the food spoilage yeast *Zygosaccharomyces bailii* CLIB 213T. *Genome Announc.* 1, e00606–13. doi: 10.1128/genomeA.00606-13
- Garay-Arroyo, A., Covarrubias, A. A., Clark, I., Nino, I., Gosset, G., and Martinez, A. (2004). Response to different environmental stress conditions of industrial and laboratory *Saccharomyces cerevisiae* strains. *Appl. Microbiol. Biotechnol.* 63, 734–741. doi: 10.1007/s00253-003-1414-4
- Gasch, A. P., Spellman, P. T., Kao, C. M., Carmel-Harel, O., Eisen, M. B., Storz, G., et al. (2000). Genomic expression programs in the response of yeast cells to environmental changes. *Mol. Biol. Cell* 11, 4241–4257. doi: 10.1091/mbc.11.12.4241
- Geng, P., Xiao, Y., Hu, Y., Sun, H., Xue, W., Zhang, L., et al. (2016). Genetic dissection of acetic acid tolerance in *Saccharomyces cerevisiae*. *World J. Microbiol. Biotechnol.* 32:145. doi: 10.1007/s11274-016-2101-9

## ACKNOWLEDGMENTS

IS-C acknowledges all those who have, over the years, contributed to the field of “Yeast adaptive response and tolerance to acetic acid” in her laboratory.

- Giannattasio, S., Guaragnella, N., Corte-Real, M., Passarella, S., and Marra, E. (2005). Acid stress adaptation protects *Saccharomyces cerevisiae* from acetic acid-induced programmed cell death. *Gene* 354, 93–98. doi: 10.1016/j.gene.2005.03.030
- Giannattasio, S., Guaragnella, N., Ždravlević, M., and Marra, E. (2013). Molecular mechanisms of *Saccharomyces cerevisiae* stress adaptation, and programmed cell death in response to acetic acid. *Front. Microbiol.* 4:33. doi: 10.3389/fmicb.2013.00033
- González-Ramos, D., Gorter de Vries, A. R., Grijseels, S. S., van Berkum, M. C., Swinnen, S., van den Broek, M., et al. (2016). A new laboratory evolution approach to select for constitutive acetic acid tolerance in *Saccharomyces cerevisiae* and identification of causal mutations. *Biotechnol. Biofuels* 9:173. doi: 10.1186/s13068-016-0583-1
- Goossens, A., de La Fuente, N., Forment, J., Serrano, R., and Portillo, F. (2000). Regulation of yeast H(+)-ATPase by protein kinases belonging to a family dedicated to activation of plasma membrane transporters. *Mol. Cell. Biol.* 20, 7654–7661. doi: 10.1128/MCB.20.20.7654-7661.2000
- Graves, T., Narendranath, N. V., Dawson, K., and Power, R. (2006). Effect of pH and lactic or acetic acid on ethanol productivity by *Saccharomyces cerevisiae* in corn mash. *J. Ind. Microbiol. Biotechnol.* 33, 469–474. doi: 10.1007/s10295-006-0091-6
- Guaragnella, N., Passarella, S., Marra, E., and Giannattasio, S. (2011). Cytochrome c Trp65Ser substitution results in inhibition of acetic acid-induced programmed cell death in *Saccharomyces cerevisiae*. *Mitochondrion* 11, 987–991. doi: 10.1016/j.mito.2011.08.007
- Guaragnella, N., Zdravlević, M., Antonacci, L., Passarella, S., Marra, E., and Giannattasio, S. (2012). The role of mitochondria in yeast programmed cell death. *Front. Oncol.* 2:70. doi: 10.3389/fonc.2012.00070
- Guaragnella, N., Zdravlević, M., Lattanzio, P., Marzulli, D., Pracheil, T., Liu, Z., et al. (2013). Yeast growth in raffinose results in resistance to acetic-acid induced programmed cell death mostly due to the activation of the mitochondrial retrograde pathway. *Biochim. Biophys. Acta* 1833, 2765–2774. doi: 10.1016/j.bbamer.2013.07.017
- Guerreiro, J. F., Mira, N. P., dos Santos, A. X., da, S., Riezman, H., and Sá-Correia, I. (2017). Membrane phosphoproteomics of yeast early response to acetic acid: role of Hrk1 kinase and lipid biosynthetic pathways, in particular sphingolipids. *Front. Microbiol.* 8:1302. doi: 10.3389/FMICB.2017.01302
- Guerreiro, J. F., Mira, N. P., and Sá-Correia, I. (2012). Adaptive response to acetic acid in the highly resistant yeast species *Zygosaccharomyces bailii* revealed by quantitative proteomics. *Proteomics* 12, 2303–2318. doi: 10.1002/pmic.201100457
- Guerreiro, J. F., Muir, A., Ramachandran, S., Thorner, J., and Sá-Correia, I. (2016a). Sphingolipid biosynthesis upregulation by TOR complex 2-Ypk1 signaling during yeast adaptive response to acetic acid stress. *Biochem. J.* 473, 4311–4325.
- Guerreiro, J. F., Sampaio-Marques, B., Soares, R., Varela Coelho, A., Leão, C., Ludovico, P., et al. (2016b). Mitochondrial proteomics of the acetic acid -induced programmed cell death response in a highly tolerant *Zygosaccharomyces bailii* -derived hybrid strain. *Microb. Cell* 3, 65–78. doi: 10.15698/mic2016.02.477
- Henderson, C. M., and Block, D. E. (2014). Examining the role of membrane lipid composition in determining the ethanol tolerance of *Saccharomyces cerevisiae*. *Appl. Environ. Microbiol.* 80, 2966–2972. doi: 10.1128/AEM.04151-13
- Henriques, S. F., Mira, N. P., and Sá-Correia, I. (2017). Genome-wide search for candidate genes for yeast robustness improvement against formic acid reveals novel susceptibility (Trk1 and positive regulators) and resistance (Haa1-regulon) determinants. *Biotechnol. Biofuels* 10:96. doi: 10.1186/s13068-017-0781-5
- Higgins, V. J., Beckhouse, A. G., Oliver, A. D., Rogers, P. J., and Dawes, I. W. (2003). Yeast genome-wide expression analysis identifies a strong ergosterol and oxidative stress response during the initial stages of an industrial lager fermentation. *Appl. Environ. Microbiol.* 69, 4777–4787. doi: 10.1128/AEM.69.8.4777-4787.2003
- Hu, S., Fürst, P., and Hamer, D. (1990). The DNA and Cu binding functions of ACE1 are interdigitated within a single domain. *New Biol.* 2, 544–555.
- Inaba, T., Watanabe, D., Yoshiyama, Y., Tanaka, K., Ogawa, J., Takagi, H., et al. (2013). An organic acid-tolerant HAA1-overexpression mutant of an industrial bioethanol strain of *Saccharomyces cerevisiae* and its application to the production of bioethanol from sugarcane molasses. *AMB Express* 3:74. doi: 10.1186/2191-0855-3-74
- Ismail, K. S. K., Sakamoto, T., Hasunuma, T., Zhao, X.-Q., and Kondo, A. (2014). Zinc, magnesium, and calcium ion supplementation confers tolerance to acetic acid stress in industrial *Saccharomyces cerevisiae* utilizing xylose. *Biotechnol. J.* 9, 1519–1525. doi: 10.1002/biot.201300553
- James, S. A., and Stratford, M. (2003). “Spoilage yeasts with emphasis on the genus *Zygosaccharomyces*,” in *Yeasts in Food*, eds T. Boekhout and V. Robert (Amsterdam: Elsevier), 171–196.
- Joneja, J. V. (2003). “Benzoate intolerance,” in *Dealing with Food Allergies*, ed. J. V. Joneja (Boulder, CO: Bull Publishing Company), 277–287.
- Jönsson, L. J., Alriksson, B., and Nilvebrant, N.-O. (2013). Bioconversion of lignocellulose: inhibitors and detoxification. *Biotechnol. Biofuels* 6:16. doi: 10.1186/1754-6834-6-16
- Kawahata, M., Masaki, K., Fujii, T., and Iefuji, H. (2006). Yeast genes involved in response to lactic acid and acetic acid: acidic conditions caused by the organic acids in *Saccharomyces cerevisiae* cultures induce expression of intracellular metal metabolism genes regulated by Aft1p. *FEMS Yeast Res.* 6, 924–936. doi: 10.1111/j.1567-1364.2006.00089.x
- Keller, G., Ray, E., Brown, P. O., and Winge, D. R. (2001). Haa1, a protein homologous to the copper-regulated transcription factor Ace1, is a novel transcriptional activator. *J. Biol. Chem.* 276, 38697–38702. doi: 10.1074/jbc.M107131200
- Koppram, R., Albers, E., and Olsson, L. (2012). Evolutionary engineering strategies to enhance tolerance of xylose utilizing recombinant yeast to inhibitors derived from spruce biomass. *Biotechnol. Biofuels* 5:32. doi: 10.1186/1754-6834-5-32
- Kumar, V., Hart, A. J., Keerthiraju, E. R., Waldron, P. R., Tucker, G. A., and Greetham, D. (2015). Expression of mitochondrial cytochrome C oxidase chaperone gene (COX20) improves tolerance to weak acid and oxidative stress during yeast fermentation. *PLoS One* 10:e0139129. doi: 10.1371/journal.pone.0139129
- Lam, F. H., Ghaderi, A., Fink, G. R., and Stephanopoulos, G. (2014). Engineering alcohol tolerance in yeast. *Science* 346, 71–75. doi: 10.1126/science.1257859
- Lee, Y., Nasution, O., Choi, E., Choi, I.-G., Kim, W., and Choi, W. (2015). Transcriptome analysis of acetic-acid-treated yeast cells identifies a large set of genes whose overexpression or deletion enhances acetic acid tolerance. *Appl. Microbiol. Biotechnol.* 99, 6391–6403. doi: 10.1007/s00253-015-6706-y
- Lee, Y., Nasution, O., Lee, Y. M., Kim, E., Choi, W., and Kim, W. (2017). Overexpression of PMA1 enhances tolerance to various types of stress and constitutively activates the SAPK pathways in *Saccharomyces cerevisiae*. *Appl. Microbiol. Biotechnol.* 101, 229–239. doi: 10.1007/s00253-016-7898-5
- Li, B.-Z., and Yuan, Y.-J. (2010). Transcriptome shifts in response to furfural and acetic acid in *Saccharomyces cerevisiae*. *Appl. Microbiol. Biotechnol.* 86, 1915–1924. doi: 10.1007/s00253-010-2518-2
- Lindahl, L., Genheden, S., Eriksson, L. A., Olsson, L., and Bettiga, M. (2016). Sphingolipids contribute to acetic acid resistance in *Zygosaccharomyces bailii*. *Biotechnol. Bioeng.* 113, 744–753. doi: 10.1002/bit.25845
- Lindberg, L., Santos, A. X., Riezman, H., Olsson, L., and Bettiga, M. (2013). Lipidomic profiling of *Saccharomyces cerevisiae* and *Zygosaccharomyces bailii* reveals critical changes in lipid composition in response to acetic acid stress. *PLoS One* 8:e73936. doi: 10.1371/journal.pone.0073936
- Liu, X., Zhang, X., and Zhang, Z. (2014). Point mutation of H3/H4 histones affects acetic acid tolerance in *Saccharomyces cerevisiae*. *J. Biotechnol.* 187, 116–123. doi: 10.1016/j.jbiotec.2014.07.445
- Longo, V., Ždravlević, M., Guaragnella, N., Giannattasio, S., Zolla, L., and Timperio, A. M. (2015). Proteome and metabolome profiling of wild-type and YCA1-knock-out yeast cells during acetic acid-induced programmed cell death. *J. Proteomics* 128, 173–188. doi: 10.1016/j.jpro.2015.08.003
- Ludovico, P., Rodrigues, F., Almeida, A., Silva, M. T., Barrientos, A., and Corte-Real, M. (2002). Cytochrome c release and mitochondria involvement in programmed cell death induced by acetic acid in *Saccharomyces cerevisiae*. *Mol. Biol. Cell* 13, 2598–2606. doi: 10.1091/mbc.E01-12-0161
- Ludovico, P., Sansonetty, F., Silva, M. T., and Corte-Real, M. (2003). Acetic acid induces a programmed cell death process in the food spoilage yeast *Zygosaccharomyces bailii*. *FEMS Yeast Res.* 3, 91–96. doi: 10.1111/j.1567-1364.2003.tb00143.x
- Ludovico, P., Sousa, M. J., Silva, M. T., Leão, C., and Corte-Real, M. (2001). *Saccharomyces cerevisiae* commits to a programmed cell death process in

- response to acetic acid. *Microbiology* 147, 2409–2415. doi: 10.1099/00221287-147-9-2409
- Ma, C., Wei, X., Sun, C., Zhang, F., Xu, J., Zhao, X., et al. (2015). Improvement of acetic acid tolerance of *Saccharomyces cerevisiae* using a zinc-finger-based artificial transcription factor and identification of novel genes involved in acetic acid tolerance. *Appl. Microbiol. Biotechnol.* 99, 2441–2449. doi: 10.1007/s00253-014-6343-x
- Macpherson, N., Shabala, L., Rooney, H., Jarman, M. G., and Davies, J. M. (2005). Plasma membrane H<sup>+</sup> and K<sup>+</sup> transporters are involved in the weak-acid preservative response of disparate food spoilage yeasts. *Microbiology* 151, 1995–2003. doi: 10.1099/mic.0.27502-0
- Martani, F., Marano, F., Bertacchi, S., Porro, D., and Branduardi, P. (2015). The *Saccharomyces cerevisiae* poly(A) binding protein Pab1 as a target for eliciting stress tolerant phenotypes. *Sci. Rep.* 5:18318. doi: 10.1038/srep18318
- Meijnen, J.-P., Randazzo, P., Foulquié-Moreno, M. R., van den Brink, J., Vandecruys, P., Stojiljkovic, M., et al. (2016). Polygenic analysis and targeted improvement of the complex trait of high acetic acid tolerance in the yeast *Saccharomyces cerevisiae*. *Biotechnol. Biofuels* 9:5. doi: 10.1186/s13068-015-0421-x
- Mira, N. P., Becker, J. D., and Sá-Correia, I. (2010a). Genomic expression program involving the Haa1p-regulon in *Saccharomyces cerevisiae* response to acetic acid. *OMICS* 14, 587–601. doi: 10.1089/omi.2010.0048
- Mira, N. P., Palma, M., Guerreiro, J. F., and Sá-Correia, I. (2010b). Genome-wide identification of *Saccharomyces cerevisiae* genes required for tolerance to acetic acid. *Microb. Cell Fact.* 9:79. doi: 10.1186/1475-2859-9-79
- Mira, N. P., Teixeira, M. C., and Sá-Correia, I. (2010c). Adaptive response and tolerance to weak acids in *Saccharomyces cerevisiae*: a genome-wide view. *OMICS* 14, 525–540. doi: 10.1089/omi.2010.0072
- Mira, N. P., Henriques, S. F., Keller, G., Teixeira, M. C., Matos, R. G., Arraiano, C. M., et al. (2011). Identification of a DNA-binding site for the transcription factor Haa1, required for *Saccharomyces cerevisiae* response to acetic acid stress. *Nucleic Acids Res.* 39, 6896–6907. doi: 10.1093/nar/gkr228
- Mira, N. P., Lourenço, A. B., Fernandes, A. R., Becker, J. D., and Sá-Correia, I. (2009). The *RIM101* pathway has a role in *Saccharomyces cerevisiae* adaptive response and resistance to propionic acid and other weak acids. *FEMS Yeast Res.* 9, 202–216. doi: 10.1111/j.1567-1364.2008.00473.x
- Mira, N. P., Munsterkötter, M., Dias-Valada, F., Santos, J., Palma, M., Roque, F. C., et al. (2014). The genome sequence of the highly acetic acid-tolerant *Zygosaccharomyces bailii*-derived interspecies hybrid strain ISA1307, isolated from a sparkling wine plant. *DNA Res.* 21, 299–313. doi: 10.1093/dnares/dst058
- Mollapour, M., and Piper, P. W. (2007). Hog1 mitogen-activated protein kinase phosphorylation targets the yeast Fps1 aquaglyceroporin for endocytosis, thereby rendering cells resistant to acetic acid. *Mol. Cell. Biol.* 27, 6446–6456. doi: 10.1128/MCB.02205-06
- Mollapour, M., Shepherd, A., and Piper, P. W. (2009). Presence of the Fps1p aquaglyceroporin channel is essential for Hog1p activation, but suppresses Slt2(Mpk1)p activation, with acetic acid stress of yeast. *Microbiology* 155, 3304–3311. doi: 10.1099/mic.0.030502-0
- Olofsson, K., Palmqvist, B., and Lidén, G. (2010). Improving simultaneous saccharification and co-fermentation of pretreated wheat straw using both enzyme and substrate feeding. *Biotechnol. Biofuels* 3:17. doi: 10.1186/1754-6834-3-17
- Orij, R., Urbanus, M. L., Vizeacoumar, F. J., Giaever, G., Boone, C., Nislow, C., et al. (2012). Genome-wide analysis of intracellular pH reveals quantitative control of cell division rate by pHc in *Saccharomyces cerevisiae*. *Genome Biol.* 13:R80. doi: 10.1186/gb-2012-13-9-r80
- Ortiz-Merino, R. A., Kuanyshv, N., Braun-Galleani, S., Byrne, K. P., Porro, D., Branduardi, P., et al. (2017). Evolutionary restoration of fertility in an interspecies hybrid yeast, by whole-genome duplication after a failed mating-type switch. *PLoS Biol.* 15:e2002128. doi: 10.1371/journal.pbio.2002128
- Paiva, S., Althoff, S., Casal, M., and Leão, C. (1999). Transport of acetate in mutants of *Saccharomyces cerevisiae* defective in monocarboxylate permeases. *FEMS Microbiol. Lett.* 170, 301–306. doi: 10.1111/j.1574-6968.1999.tb13387.x
- Paiva, S., Devaux, F., Barbosa, S., Jacq, C., and Casal, M. (2004). Ady2p is essential for the acetate permease activity in the yeast *Saccharomyces cerevisiae*. *Yeast* 21, 201–210. doi: 10.1002/yea.1056
- Palma, M., Dias, P. J., Roque, F. C., Luzia, L., Guerreiro, J. F., and Sá-Correia, I. (2017a). The *Zygosaccharomyces bailii* transcription factor Haa1 is required for acetic acid and copper stress responses suggesting subfunctionalization of the ancestral bifunctional protein Haa1/Cup2. *BMC Genomics* 18:75. doi: 10.1186/s12864-016-3443-2
- Palma, M., Münsterkötter, M., Peça, J., Güldener, U., and Sá-Correia, I. (2017b). Genome sequence of the highly weak-acid-tolerant *Zygosaccharomyces bailii* IST302, amenable to genetic manipulations and physiological studies. *FEMS Yeast Res.* 17:fox025. doi: 10.1093/femsyr/fox025
- Palma, M., Roque, F., de, C., Guerreiro, J. F., Mira, N. P., Queiroz, L., et al. (2015). Search for genes responsible for the remarkably high acetic acid tolerance of a *Zygosaccharomyces bailii*-derived interspecies hybrid strain. *BMC Genomics* 16:1070. doi: 10.1186/s12864-015-2278-6
- Pampulha, M. E., and Loureiro-Dias, M. C. (1989). Combined effect of acetic acid, pH and ethanol on intracellular pH of fermenting yeast. *Appl. Microbiol. Biotechnol.* 31, 547–550. doi: 10.1007/BF00270792
- Pampulha, M. E., and Loureiro-Dias, M. C. (1990). Activity of glycolytic enzymes of *Saccharomyces cerevisiae* in the presence of acetic acid. *Appl. Microbiol. Biotechnol.* 34, 375–380. doi: 10.1007/BF00170063
- Pampulha, M. E., and Loureiro-Dias, M. C. (2000). Energetics of the effect of acetic acid on growth of *Saccharomyces cerevisiae*. *FEMS Microbiol. Lett.* 184, 69–72. doi: 10.1111/j.1574-6968.2000.tb08992.x
- Phillips, A. J., Crowe, J. D., and Ramsdale, M. (2006). Ras pathway signaling accelerates programmed cell death in the pathogenic fungus *Candida albicans*. *Proc. Natl. Acad. Sci. U.S.A.* 103, 726–731. doi: 10.1073/pnas.0506405103
- Piper, P., Calderon, C. O., Hatzixanthis, K., and Mollapour, M. (2001). Weak acid adaptation: the stress response that confers yeasts with resistance to organic acid food preservatives. *Microbiology* 147, 2635–2642. doi: 10.1099/00221287-147-10-2635
- Postnikoff, S. D. L., Malo, M. E., Wong, B., and Harkness, T. A. A. (2012). The yeast forkhead transcription factors Fkh1 and Fkh2 regulate lifespan and stress response together with the anaphase-promoting complex. *PLoS Genet.* 8:e1002583. doi: 10.1371/journal.pgen.1002583
- Rasmussen, J. E., Schultz, E., Snyder, R. E., Jones, R. S., and Smith, C. R. (1995). Acetic acid as a causative agent in producing stuck fermentations. *Am. J. Enol. Vitic.* 46, 278–280.
- Rodrigues, F., Sousa, M. J., Ludovico, P., Santos, H., Côrte-Real, M., and Leão, C. (2012). The fate of acetic acid during glucose co-metabolism by the spoilage yeast *Zygosaccharomyces bailii*. *PLoS One* 7:e52402. doi: 10.1371/journal.pone.0052402
- Rodrigues, F., Zeeman, A. M., Cardoso, H., Sousa, M. J., Steensma, H. Y., Corte-Real, M., et al. (2004). Isolation of an acetyl-CoA synthetase gene (*ZbACS2*) from *Zygosaccharomyces bailii*. *Yeast* 21, 325–331. doi: 10.1002/yea.1081
- Roelants, F., Leskoske, K., Martinez Marshall, M. N., Locke, M., and Thorner, J. (2017). The TORC2-dependent signaling network in the yeast *Saccharomyces cerevisiae*. *Biomolecules* 7:66. doi: 10.3390/biom7030066
- Roelants, F. M., Breslow, D. K., Muir, A., Weissman, J. S., and Thorner, J. (2011). Protein kinase Ypk1 phosphorylates regulatory proteins Orm1 and Orm2 to control sphingolipid homeostasis in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. U.S.A.* 108, 19222–19227. doi: 10.1073/pnas.1116948108
- Sá-Correia, I., Guerreiro, J. F., Loureiro-Dias, M. C., Leão, C., and Côrte-Real, M. (2014). “*Zygosaccharomyces*,” in *Encyclopedia of Food Microbiology*, eds C. A. Batt and M. L. Tortorello (Cambridge, MA: Elsevier Ltd.), 849–855. doi: 10.1016/B978-0-12-384730-0.00364-5
- Sakihama, Y., Hasunuma, T., and Kondo, A. (2015). Improved ethanol production from xylose in the presence of acetic acid by the overexpression of the *HAA1* gene in *Saccharomyces cerevisiae*. *J. Biosci. Bioeng.* 119, 297–302. doi: 10.1016/j.jbiosc.2014.09.004
- Schüller, C., Mammun, Y. M., Mollapour, M., Krapf, G., Schuster, M., Bauer, B. E., et al. (2004). Global phenotypic analysis and transcriptional profiling defines the weak acid stress response regulon in *Saccharomyces cerevisiae*. *Mol. Biol. Cell* 15, 706–720. doi: 10.1091/mbc.E03-05-0322
- Sekito, T., Thornton, J., and Butow, R. A. (2000). Mitochondria-to-nuclear signaling is regulated by the subcellular localization of the transcription factors Rtg1p, and Rtg3p. *Mol. Biol. Cell* 11, 2103–2115. doi: 10.1091/mbc.11.6.2103
- Simões, T., Mira, N. P., Fernandes, A. R., and Sá-Correia, I. (2006). The *SPH1* gene, encoding a glycosylphosphatidylinositol-anchored cell wall protein, plays a prominent role in the development of yeast resistance to lipophilic weak-acid food preservatives. *Appl. Environ. Microbiol.* 72, 7168–7175. doi: 10.1128/AEM.01476-06



- Sousa, M., Duarte, A., Fernandes, T. R., Chaves, S. R., Pacheco, A., Leão, C., et al. (2013). Genome-wide identification of genes involved in the positive and negative regulation of acetic acid-induced programmed cell death in *Saccharomyces cerevisiae*. *BMC Genomics* 14:838. doi: 10.1186/1471-2164-14-838
- Sousa, M. J., Miranda, L., Corte-Real, M., and Leão, C. (1996). Transport of acetic acid in *Zygosaccharomyces bailii*: effects of ethanol and their implications on the resistance of the yeast to acidic environments. *Appl. Environ. Microbiol.* 62, 3152–3157.
- Sousa, M. J., Rodrigues, F., Corte-Real, M., and Leão, C. (1998). Mechanisms underlying the transport and intracellular metabolism of acetic acid in the presence of glucose in the yeast *Zygosaccharomyces bailii*. *Microbiology* 144, 665–670. doi: 10.1099/00221287-144-3-665
- Steels, H., James, S. A., Roberts, I. N., and Stratford, M. (2000). Sorbic acid resistance: the inoculum effect. *Yeast* 16, 1173–1183. doi: 10.1002/1097-0061(20000930)16:13<1173::AID-YEA617>3.0.CO;2-8
- Stratford, M. (2006). "Food and beverage spoilage yeasts," in *Yeasts in Food and Beverages*, eds A. Querol and G. Fleet (New York, NY: Springer-Verlag), 335–379. doi: 10.1007/978-3-540-28398-0\_11
- Stratford, M., Nebe-von-Caron, G., Steels, H., Novodvorska, M., Ueckert, J., and Archer, D. B. (2013a). Weak-acid preservatives: pH and proton movements in the yeast *Saccharomyces cerevisiae*. *Int. J. Food Microbiol.* 161, 164–171. doi: 10.1016/j.ijfoodmicro.2012.12.013
- Stratford, M., Steels, H., Nebe-von-Caron, G., Novodvorska, M., Hayer, K., and Archer, D. B. (2013b). Extreme resistance to weak-acid preservatives in the spoilage yeast *Zygosaccharomyces bailii*. *Int. J. Food Microbiol.* 166, 126–134. doi: 10.1016/j.ijfoodmicro.2013.06.025
- Sugiyama, M., Akase, S.-P., Nakanishi, R., Horie, H., Kaneko, Y., and Harashima, S. (2014). Nuclear localization of Haa1, which is linked to its phosphorylation status, mediates lactic acid tolerance in *Saccharomyces cerevisiae*. *Appl. Environ. Microbiol.* 80, 3488–3495. doi: 10.1128/AEM.04241-13
- Swan, T. M., and Watson, K. (1998). Stress tolerance in a yeast sterol auxotroph: role of ergosterol, heat shock proteins and trehalose. *FEMS Microbiol. Lett.* 169, 191–197. doi: 10.1111/j.1574-6968.1998.tb13317.x
- Swinnen, S., Henriques, S. F., Shrestha, R., Ho, P.-W., Sá-Correia, I., and Nevoigt, E. (2017). Improvement of yeast tolerance to acetic acid through Haa1 transcription factor engineering: towards the underlying mechanisms. *Microb. Cell Fact.* 16:7. doi: 10.1186/s12934-016-0621-5
- Tanaka, K., Ishii, Y., Ogawa, J., and Shima, J. (2012). Enhancement of acetic acid tolerance in *Saccharomyces cerevisiae* by overexpression of the HAA1 gene, encoding a transcriptional activator. *Appl. Environ. Microbiol.* 78, 8161–8163. doi: 10.1128/AEM.02356-12
- Teixeira, M. C., Mira, N. P., and Sá-Correia, I. (2011). A genome-wide perspective on the response and tolerance to food-relevant stresses in *Saccharomyces cerevisiae*. *Curr. Opin. Biotechnol.* 22, 150–156. doi: 10.1016/j.copbio.2010.10.011
- Teixeira, M. C., Monteiro, P. T., Palma, M., Costa, C., Godinho, C. P., Pais, P., et al. (2017). YEASTRACT: an upgraded database for the analysis of transcription regulatory networks in *Saccharomyces cerevisiae*. *Nucleic Acids Res.* 46, D348–D353. doi: 10.1093/nar/gkx842
- Tenreiro, S., Nunes, P. A., Viegas, C. A., Neves, M. S., Teixeira, M. C., Cabral, M. G., et al. (2002). AQR1 gene (ORF YNL065w) encodes a plasma membrane transporter of the major facilitator superfamily that confers resistance to short-chain monocarboxylic acids and quinidine in *Saccharomyces cerevisiae*. *Biochem. Biophys. Res. Commun.* 292, 741–748. doi: 10.1006/bbrc.2002.6703
- Theron, M., and Lues, J. (eds) (2010). "Application of organic acids in food preservation," in *Organic Acids and Food Preservation*, (Florida: CRC Press), 51–96.
- Thomas, D. S., and Davenport, R. R. (1985). *Zygosaccharomyces bailii* — a profile of characteristics and spoilage activities. *Food Microbiol.* 2, 157–169. doi: 10.1016/S0740-0020(85)80008-3
- Ullah, A., Chandrasekaran, G., Brul, S., and Smits, G. J. (2013). Yeast adaptation to weak acids prevents futile energy expenditure. *Front. Microbiol.* 4:142. doi: 10.3389/fmicb.2013.00142
- Ullah, A., Orij, R., Brul, S., and Smits, G. J. (2012). Quantitative analysis of the modes of growth inhibition by weak organic acids in *Saccharomyces cerevisiae*. *Appl. Environ. Microbiol.* 78, 8377–8387. doi: 10.1128/AEM.02126-12
- Vilela-Moura, A., Schuller, D., Mendes-Faia, A., Silva, R. D., Chaves, S. R., Sousa, M. J., et al. (2011). The impact of acetate metabolism on yeast fermentative performance and wine quality: reduction of volatile acidity of grape musts and wines. *Appl. Microbiol. Biotechnol.* 89, 271–280. doi: 10.1007/s00253-010-2898-3
- Vyas, V. K., Berkey, C. D., Miyao, T., and Carlson, M. (2005). Repressors Nrg1 and Nrg2 regulate a set of stress-responsive genes in *Saccharomyces cerevisiae*. *Eukaryot. Cell* 4, 1882–1891. doi: 10.1128/EC.4.11.1882-1891.2005
- Wan, C., Zhang, M., Fang, Q., Xiong, L., Zhao, X., Hasunuma, T., et al. (2015). The impact of zinc sulfate addition on the dynamic metabolic profiling of *Saccharomyces cerevisiae* subjected to long term acetic acid stress treatment and identification of key metabolites involved in the antioxidant effect of zinc. *Metallomics* 7, 322–332. doi: 10.1039/c4mt00275j
- Wright, J., Bellissimi, E., de Hulster, E., Wagner, A., Pronk, J. T., and van Maris, A. J. A. (2011). Batch and continuous culture-based selection strategies for acetic acid tolerance in xylose-fermenting *Saccharomyces cerevisiae*. *FEMS Yeast Res.* 11, 299–306. doi: 10.1111/j.1567-1364.2011.00719.x
- Wu, X., Zhang, L., Jin, X., Fang, Y., Zhang, K., Qi, L., et al. (2016). Deletion of JJJ1 improves acetic acid tolerance and bioethanol fermentation performance of *Saccharomyces cerevisiae* strains. *Biotechnol. Lett.* 38, 1097–1106. doi: 10.1007/s10529-016-2085-4
- Zahn, K., and Jacobson, S. (2015). Acetate resistance in yeast based on introduction of a mutant HAA1 allele. U.S. Patent No. US9085781 B2.
- Ždraljević, M., Guaragnella, N., Antonacci, L., Marra, E., and Giannattasio, S. (2012). Yeast as a tool to study signaling pathways in mitochondrial stress response, and Cytoprotection. *Sci. World J.* 2012, 1–10. doi: 10.1100/2012/912147
- Zhang, M.-M., Zhao, X.-Q., Cheng, C., and Bai, F.-W. (2015). Improved growth and ethanol fermentation of *Saccharomyces cerevisiae* in the presence of acetic acid by overexpression of SET5 and PPR1. *Biotechnol. J.* 10, 1903–1911. doi: 10.1002/biot.201500508
- Zheng, D.-Q., Wu, X.-C., Wang, P.-M., Chi, X.-Q., Tao, X.-L., Li, P., et al. (2011). Drug resistance marker-aided genome shuffling to improve acetic acid tolerance in *Saccharomyces cerevisiae*. *J. Ind. Microbiol. Biotechnol.* 38, 415–422. doi: 10.1007/s10295-010-0784-8

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Palma, Guerreiro and Sá-Correia. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.





# Analysis of the NCR Mechanisms in *Hanseniaspora vineae* and *Saccharomyces cerevisiae* During Winemaking

Jessica Lleixà<sup>1</sup>, Valentina Martín<sup>2</sup>, Facundo Giorello<sup>2</sup>, Maria C. Portillo<sup>1</sup>, Francisco Carrau<sup>2</sup>, Gemma Beltran<sup>1\*</sup> and Albert Mas<sup>1</sup>

<sup>1</sup> Departament de Bioquímica i Biotecnologia, Facultat d'Enologia, Universitat Rovira i Virgili, Tarragona, Spain, <sup>2</sup> Sección Enología, Food Science and Technology Department, Facultad de Química, Universidad de la República (UdelaR), Montevideo, Uruguay

## OPEN ACCESS

### Edited by:

Isabel Sá-Correia,  
University of Lisbon, Portugal

### Reviewed by:

Maria João Sousa,  
University of Minho, Portugal  
Francisco Salinas,  
Universidad de Santiago de Chile,  
Chile  
Claudio Martínez,  
Universidad de Santiago de Chile,  
Chile

### \*Correspondence:

Gemma Beltran  
gemma.beltran@urv.cat

### Specialty section:

This article was submitted to  
Evolutionary and Genomic  
Microbiology,  
a section of the journal  
Frontiers in Genetics

**Received:** 03 October 2018

**Accepted:** 31 December 2018

**Published:** 11 January 2019

### Citation:

Lleixà J, Martín V, Giorello F,  
Portillo MC, Carrau F, Beltran G and  
Mas A (2019) Analysis of the NCR  
Mechanisms in *Hanseniaspora vineae*  
and *Saccharomyces cerevisiae* During  
Winemaking. *Front. Genet.* 9:747.  
doi: 10.3389/fgene.2018.00747

There is increasing interest in the use of non-*Saccharomyces* yeasts in winemaking due to their positive attributes. The non-*Saccharomyces* yeast *Hanseniaspora vineae* is an apiculate yeast that has been associated with the production of wine with good fermentation capacity and an increase in aromatic properties. However, this yeast represents a concern in mixed culture fermentation because of its nutrient consumption, especially nitrogen, as its mechanisms of regulation and consumption are still unknown. In this study, we analyzed the nitrogen consumption, as well as the nitrogen catabolism repression (NCR) mechanism, in two genome-sequenced *H. vineae* strains, using synthetic must fermentations. The use of synthetic must with an established nitrogen content allowed us to study the NCR mechanism in *H. vineae*, following the amino acid and ammonia consumption, and the expression of genes known to be regulated by the NCR mechanism in *S. cerevisiae*, *AGP1*, *GAP1*, *MEP2*, and *PUT2*. *H. vineae* exhibited a similar amino acid consumption and gene expression profile to *S. cerevisiae*. However, the wine strain of *S. cerevisiae* QA23 consumed ammonia and valine more quickly and, in contrast, tyrosine and tryptophan more slowly, than the *H. vineae* strains. Our results showed a similar behavior of nitrogen regulation in *H. vineae* and *S. cerevisiae*, indicating the presence of the NCR mechanism in this *Hanseniaspora* yeast differentiated before the whole genome duplication event of the *Saccharomyces* complex. Future study will elucidate if the NCR mechanism is the only strategy used by *H. vineae* to optimize nitrogen consumption.

**Keywords:** non-*Saccharomyces*, yeast assimilable nitrogen, nitrogen consumption, alcoholic fermentation, amino acids

## INTRODUCTION

For many years, the microbiological process of winemaking has been focused on the use of starter cultures of *Saccharomyces cerevisiae*. The inoculation of commercial strains of *S. cerevisiae* is a common practice in wineries to ensure the completion of the fermentation and the quality of the final product. However, the elaboration of uniformed wines is not always desired, and winemakers

are becoming more interested in obtaining characteristic and differential wines. Considering this fact, in recent years, much effort has been focused on the use of non-*Saccharomyces* yeasts to obtain wine with new organoleptic characteristics (Fleet, 2008; Jolly et al., 2014; Carrau et al., 2015). Non-*Saccharomyces* yeasts are naturally present on grape surfaces, and they can start spontaneous fermentations that can lead in incomplete fermentations or result in wines with unpleasant properties. Despite this fact, many of these non-*Saccharomyces* yeasts have proven to produce enzymatic activities and release metabolites that improve some oenological processes and the wine flavor (Jolly et al., 2014; Padilla et al., 2016; Varela, 2016). For this reason, the interest in the use of co-fermented or sequential mixed cultures of non-*Saccharomyces* and *S. cerevisiae* has increased to take advantage of both trends during the winemaking process.

*Hanseniaspora vineae* is one species of yeast that belongs to the non-*Saccharomyces* yeasts of oenological interest (Martin et al., 2018). The primary positive contributions of this yeast during the winemaking process are basically related to the aroma profile in the final wine. *H. vineae* has been demonstrated to increase fruity aromas and produce high amounts of acetate esters, primarily 2-phenylethyl acetate and benzenoids, in wines elaborated in either synthetic (Martin et al., 2016) or natural musts inoculated with *H. vineae* (Lleixà et al., 2016) or by sequential fermentation with *S. cerevisiae* (Viana et al., 2011; Medina et al., 2013). The higher ester content produced by this non-*Saccharomyces* yeast can be explained by its prominent  $\beta$ -glucosidase activity that enables it to release these compounds into the media (Barquet et al., 2012; López et al., 2015).

The development of non-*Saccharomyces* yeasts can affect the growth of the primary wine yeast *S. cerevisiae* and the fermentation progress as a consequence of the consumption of important nutrients, such as nitrogen and vitamins (Medina et al., 2012). Some studies confirmed the effect of non-*Saccharomyces* yeasts on nutrient availability in mixed cultures. Andorrà et al. (2010) observed that mixed cultures with *Candida zemplinina* and *Hanseniaspora uvarum* had a higher amino acid consumption than pure cultures of these yeasts. Indeed, pure and mixed cultures showed a preferential uptake of some amino acid groups related with the synthesis of aroma compounds that might be strain-dependent as was shown for *S. cerevisiae* (Bisson, 1991). This higher nitrogen consumption also happened in mixed cultures with *H. vineae* (Medina et al., 2012). The moment of inoculation, simultaneous or sequential, and the inoculum size in mixed cultures determine the progress of the fermentation because of the nutrient competition between the *Saccharomyces* and non-*Saccharomyces* yeasts. Some researchers have demonstrated that a sequential fermentation resulted in sluggish or stuck fermentations as a consequence of the nutrient consumption of the non-*Saccharomyces* strain, which reduced the nutrient availability to the *Saccharomyces* strain (Medina et al., 2012; Taillandier et al., 2014). In *S. cerevisiae* has been observed to activate the genes responsible for nitrogen and glucose metabolism to prevent this situation when it was co-cultivated with different non-*Saccharomyces* yeast to decrease the

nutrients available to the non-*Saccharomyces* yeast (Curiel et al., 2017). In addition, recent studies focused on the specific use of ammonia and amino acids by the different non-*Saccharomyces* species and its implication on *S. cerevisiae* performance in sequential fermentations. Non-*Saccharomyces* yeast have been shown to exhibit a specific amino acid consumption profile depending on the yeast species, which interferes with *S. cerevisiae* development and generates changes in the volatile profile during sequential fermentations (Gobert et al., 2017; Rollero et al., 2018). In summary, the specific nutrient addition of amino acids, ammonia or vitamins has to be evaluated to ensure a good fermentation performance under sequential yeast inoculation (Medina et al., 2012; Gobert et al., 2017; Rollero et al., 2018).

Specifically, in grape must, we can find different nitrogen compounds, but only some of them can be consumed by *S. cerevisiae* to produce biomass and encourage the fermentation process. These compounds, known as Yeast Assimilable Nitrogen (YAN), are comprised of the ammonia and the amino acids present in the grape juice (Bell and Henschke, 2005). From this YAN, we can differentiate the preferred nitrogen sources, such as ammonia, asparagine, and glutamine, which promote *S. cerevisiae* growth, and the non-preferred nitrogen sources, such as urea, that result in a low growth rate when it grows only with those nitrogen sources (Ter Schure et al., 2000; Magasanik and Kaiser, 2002).

Therefore, *S. cerevisiae* has developed a mechanism called Nitrogen Catabolism Repression (NCR) that selects the best nitrogen sources for growth. The NCR mechanism consists in the reduction of proteins responsible for utilization and uptake of non-preferred nitrogen sources in the presence of preferred nitrogen sources. This mechanism acts at two levels to assure the consumption of preferred nitrogen sources. The first consists in the inactivation and degradation of the existing non-preferred nitrogen source permeases, and the second consists in the repression of genes encoding for non-preferred nitrogen source permeases (Ter Schure et al., 2000; Magasanik and Kaiser, 2002).

From the 19 amino acid permeases that *S. cerevisiae* contains, there are three high-capacity permeases that are nitrogen-regulated, including *AGP1* (high-Affinity Glutamine Permease), *GAP1* (General Amino acid Permease) and *PUT4* (Proline UTilization). In addition, other non-permeases proteins like *PUT2* (Delta-1-pyrroline-5-carboxylate dehydrogenase), which is a key enzyme for the conversion of proline into glutamate in the mitochondria once it has entered the cell through *PUT4*, are also nitrogen-regulated (Hofman-Bang, 1999). The amino acid permeases *GAP1* and *PUT4* together with the dehydrogenase *PUT2* are active during growth in non-preferred nitrogen content and repressed in the presence of a preferred nitrogen source, such as ammonium (Forsberg and Ljungdahl, 2001). Alternatively, *AGP1* is active in the presence of a preferred nitrogen source and repressed when this nitrogen is consumed (Regenberg et al., 1999).

In the case of ammonium, three permeases are responsible of its uptake, namely *MEP1*, *MEP2*, and *MEP3*. When the

concentration of ammonia in the medium is low, these permeases become active. However, in a non-preferred nitrogen source, the expression of *MEP2* is much higher than those of *MEP1* and *MEP3*, since it is the one with the highest affinity for ammonium (Ter Schure et al., 2000). Previous studies have reported that the expression of those nitrogen-regulated proteins can be used as a biomarker for nitrogen deficiency in wine fermentations (Beltran et al., 2005; Deed et al., 2011; Gutiérrez et al., 2013). The study of the expression of these proteins can be also an indirect evidence of the existence of NCR mechanism. In fact, NCR genes are regulated by several transcription factors, amongst others Gln3 and Ntl1, and also by their regulator Ure2p. Under nitrogen limitation, Gln3 dissociates from Ure2p, the dephosphorylated Gln3 goes to the nucleus and increases the transcription of genes containing UAS<sub>NTR</sub> sequence (Upstream activating sequence), like *GAP1*, *PUT4*, *PUT2*, and *MEP2* genes (Ter Schure et al., 2000; Tesnière et al., 2015).

Nitrogen metabolism and the NCR mechanism have been deeply studied in *S. cerevisiae*, both in laboratory and wild strains, showing the multiple mechanisms used by this species under nitrogen-limited conditions (Beltran et al., 2004; Godard et al., 2007; Gutiérrez et al., 2013; Tesnière et al., 2015). However, very little is known about nitrogen preferences and regulation in non-*Saccharomyces* species. A better understanding of nitrogen utilization among the different yeast species is important to increase the efficiency, predictability and quality of wine production, as well as of other biotechnological uses of yeast. The great variability on respiro-fermentative metabolism observed in non-*Saccharomyces* yeasts (Gonzalez et al., 2013) is an example of the possible divergences in nitrogen metabolism between *Saccharomyces* and non-*Saccharomyces* species. One of the limitations for performing molecular studies on non-*Saccharomyces* yeasts has been the lack of genomic data. Fortunately, in the last decade, the genomes of a large number of wine yeast species have been sequenced (Masneuf-Pomarede et al., 2016), and these sequences are available for molecular or genetic studies, such as those of the wine yeast *H. vineae* (Giorello et al., 2014).

In summary, the use of non-*Saccharomyces* yeasts is increasing to produce new wine styles taking advantage of their potential abilities. The nitrogen availability is important for yeast for its growth, as well as for the production of volatile compounds during the fermentation process. The mechanism used by *S. cerevisiae* to select the best nitrogen source is well known and documented, while it has not been studied in non-*Saccharomyces* yeasts.

The aim of this study was to evaluate the presence of the nitrogen catabolite repression (NCR) mechanism in *H. vineae*. We performed laboratory-scale fermentations of *H. vineae* and *S. cerevisiae* using a synthetic must with a defined nitrogen content. We followed the expression of the ortholog NCR-sensitive genes in *H. vineae* and the amino acid and ammonium consumption during the fermentation. Finally, we compared the results of *H. vineae* fermentations with the fermentations performed using a commercial *S. cerevisiae* strain.

## MATERIALS AND METHODS

### Yeast Strains

The commercial wine yeast strain used in this study was *Saccharomyces cerevisiae* QA23 (Lallemand®, Canada). The apiculate yeast strains used, *Hanseniaspora vineae* T02/5AF and *Hanseniaspora vineae* T02/19AF, were both isolated from Uruguayan vineyards (Barquet et al., 2012). The use of two strains of *H. vineae* responds to the need of validating the results in this specie since the strains chosen have shown differences in aroma production which could be related with nitrogen metabolism (Martin et al., 2016).

Yeast strain *S. cerevisiae* QA23 was in active dry yeast (ADY) form. The rehydration process was performed according to the manufacturer's instructions (Lallemand®, Canada). Both strains of *H. vineae*, T02/5AF and T02/19AF, were in fresh paste form, and both were prepared in the same way as QA23 using warm water.

### Fermentation Conditions

To determine the uptake and metabolism of nitrogen, yeast strains were grown at 28°C during 24 h in a solid yeast extract-peptone dextrose (YPD) medium (1% yeast extract, 2% peptone, 2% glucose, and 1.7% agar). A colony from the yeast culture was inoculated in 50 mL liquid YPD media for 24 h in Erlenmeyer flasks at 120 rpm and 28°C. A population of  $1 \times 10^6$  cells/mL of the yeast strain was inoculated into an Erlenmeyer flask with 100 mL of yeast nitrogen base (YNB) media without amino acids (Difco™) with 150 mg/L of (NH<sub>4</sub>)SO<sub>4</sub> and 20 g/L of glucose (AppliChem Panreac®) for 24 h at 120 rpm and 28°C. The YNB medium was used to exhaust the yeast nitrogen reserves.

After a microscopic counting of the cells using a Neubauer chamber, 1,500 mL of synthetic must was inoculated to a final concentration of  $2 \times 10^6$  cells/mL. The cells were washed and resuspended with synthetic must before inoculation to remove the nitrogen residues.

The fermentations were performed in synthetic must (Supplementary Table S1) with a nitrogen content of 140 mg YAN/L (Supplementary Table S2) since this concentration has been established as the ideal one to achieve a complete fermentation without residual or excess nitrogen (Bely et al., 1990).

The fermentations were conducted in triplicate in laboratory-scale fermenters, i.e., 500 mL bottles filled with 440 mL of synthetic must and covered with a cap with two tubes that allowed sampling and the exit of carbon dioxide. The fermenters were maintained on a rotating shaker at 120 rpm at room temperature (22–23°C). The fermentation activity was assayed by the juice density every day using a portable density meter (Mettler Toledo).

### Cell Growth Measurements

In the laboratory-scale fermentations, cell population monitoring was established by measuring the absorbance at 600 nm. The samples were measured every 4 h during the first 48 h after inoculation and once a day from 48 h to the end of the fermentation.

## Determination of Relative Gene Expression

The evaluation of the gene expression affected by the nitrogen catabolite repression (NCR) was performed during the first hours on the synthetic must fermentation. Sampling every 4 h during the first 24 h and every 6 h from 24 to 36 h was followed by centrifugation (16,000 rpm, 5 min and 4°C) and removal of the supernatant. The pellet was washed with cold sterile MilliQ water (Millipore Q-POD™ Advantage A10), centrifuged (16,000 rpm, 5 min, and 4°C) and after removal of the supernatant, it was frozen in liquid nitrogen and stored at −80°C.

The RNA was extracted from these samples using an RNeasy® Mini kit (QIAGEN®) and RNase-Free DNase Set (QIAGEN®) according to the manufacturer's instructions. The RNA obtained was then measured using a Nano Drop (NanoDrop 1000 Thermo® Scientific) and diluted to a final concentration of 320 ng/μL in a total volume of 11 μL. The cDNA synthesis of each sample was performed using the corresponding RNA, 1 μL of oligo-dT primer (Invitrogen™), 1 μL of dNTPs (10 mM) and 1 μL of transcriptase (SuperScript® II Reverse Transcriptase-Invitrogen™) and amplified using a 2720 Thermal Cycler (Applied Biosystems) according to the manufacturer's instructions.

The genes evaluated in this experiment considering their role in the NCR mechanism were *AGP1*, *GAP1*, *MEP2*, and *PUT2* and their orthologous in *H. vineae*. Annotation of putative orthologous was based on BLASTx searches using *H. vineae* predicted CDS and the proteome of *S. cerevisiae*. A hit was considered significant if: (i) e-value threshold was less than 1e-10 (ii) the alignment length covered more than 90% of the length of both sequences, and (iii) both sequences presented the same pfam domain. In case of multiple hits we selected the *H. vineae* prediction with higher percentage of amino acid identity (**Supplementary Table S3**). Primer design for each gene was performed using Primer Express software (Primer Express 3.0 Applied Biosystems) (**Table 1**).

**TABLE 1** | Primers used for the analysis of the expression of NCR-related genes in *H. vineae*.

Gene	Name	Oligonucleotide sequence (5'–3' end)
<i>ACT1</i>	ACT-F	GGCTTCCTTGACCACTTTCCAA
	ACT-R	GATGGACCACTTTCGTCGTATTC
<i>AGP1</i>	AGP1-F	ATTGCTGGGTGACGGTTCCT
	AGP1-R	TGACATTGGTAGCGGCAATAAC
<i>GAP1</i>	GAP1-F	CAAAGGTTTGCCATCTGTGCATC
	GAP1-R	TGCAGAGTTACCGACAGACAACA
<i>MEP2</i>	MEP2-F	TCGATGACGGGTTGGATGTT
	MEP2-R	CATAACGGCACCACTTAAACCA
<i>PUT2</i>	PUT2-F	GATACGACATGTTGGCAGCAA
	PUT2-R	TTTCGCCTTGGAAACGTTT

The primer design for the *H. vineae* genes was performed using Primer Express software (Primer Express 3.0 Applied Biosystems). The primers used for *S. cerevisiae* have been previously described by Beltran et al. (2004) and Gutiérrez et al. (2013).

The housekeeping genes encoding Actin (*ACT1*) and Inorganic Pyrophosphatase 1 (*IPPI*) from *S. cerevisiae* and *H. vineae* were used to normalize the amplification curves of the selected genes considering their stability (Ståhlberg et al., 2008). All samples from each fermentation replicate were analyzed in duplicate.

In all the samples, the Real-Time Quantitative PCR reaction was performed using 10 μL of SYBR Green [SYBR® Premix Ex Taq II (Tli RNaseH Plus)], 0.4 μL of ROX Reference Dye (SYBR® Premix Ex Taq II), 0.8 μL of each specific primer (10 μM) and 6 μL of sterile MilliQ water (Millipore Q-POD™ Advantage A10). The amplification process was conducted using a 7300 Real Time PCR System (Applied Biosystems) as follows: 50°C for 2 min, 95°C for 10 min and 40 cycles at 95°C for 15 s, 60°C for 2 min and 72°C for 30 s.

The relative gene expression was determined using the  $2^{-\Delta\Delta C_t}$  method (Beltran et al., 2005), where the  $C_t$  value corresponds to the number of cycles needed to achieve the background fluorescence. This method is used to compare the  $C_t$  values of the gene of interest, and the  $C_t$  values of the reference genes (*ACT1* and *IPPI*) ( $\Delta C_t$ ); and  $-\Delta\Delta C_t$  consists of the difference of  $\Delta C_t$  from the samples of each time point, and the  $\Delta C_t$  of the reference time (4 h after inoculation). Results were expressed as the mean Log10 relative gene expression. All samples were analyzed in triplicate, and the resulting Log10  $2^{-\Delta\Delta C_t}$  values were statistically analyzed using ANOVA and Tukey's post-test.

## Nitrogen Content Analysis of Laboratory Fermentation

The individual amino acid and ammonium contents of each sample were determined using high-performance liquid chromatography (HPLC) (Agilent 1100 Series HPLC) (Gómez-Alonso et al., 2007). The sample (400 μL) was mixed with borate buffer (700 μL), methanol (300 μL), diethyl ethoxymethylenemalonate (DEEM) (15 μL) and L-amino adipic acid (internal control) (10 μL). After 2 h at 80°C, 50 μL of each sample was directly injected into the HPLC, which consists of a low pressure gradient quaternary pump, a thermostatted autosampler, a DAD ultraviolet detector and a fluorescence detector (Agilent Technologies, Germany). The separation process of the sample was performed using a 4.6 × 250 mm × 5 μm Hypersil ODS column (Agilent Technologies, Germany).

The solvent system was as follows: A solvent (mobile phase) [4.1 g of sodium acetate anhydrous diluted in 250 mL of MilliQ water, adjusted to pH 5.8 with glacial acetic acid and 0.4 g of sodium azide brought to a final volume of 2 L with MilliQ water (Millipore Q-POD™ Advantage A10)] and B solvent (stationary phase) [80% acetonitrile and 20% methanol]. The analytical temperature was 20°C, and the flow rate was 0.9 mL/min. The concentration of each amino acid and ammonia was calculated using an external calibration curve of each component and expressed as mg N/L. The software used for the integration was Agilent ChemStation Plus (Agilent Technologies, Germany).



## Statistical Analysis

Statistical analysis of the gene expression data was performed using an ANOVA and indicated by the Tukey's post-test (all pair comparisons) using XLSTAT Software. The results were considered statistically significant at a  $p$ -value less than 0.05.

## RESULTS

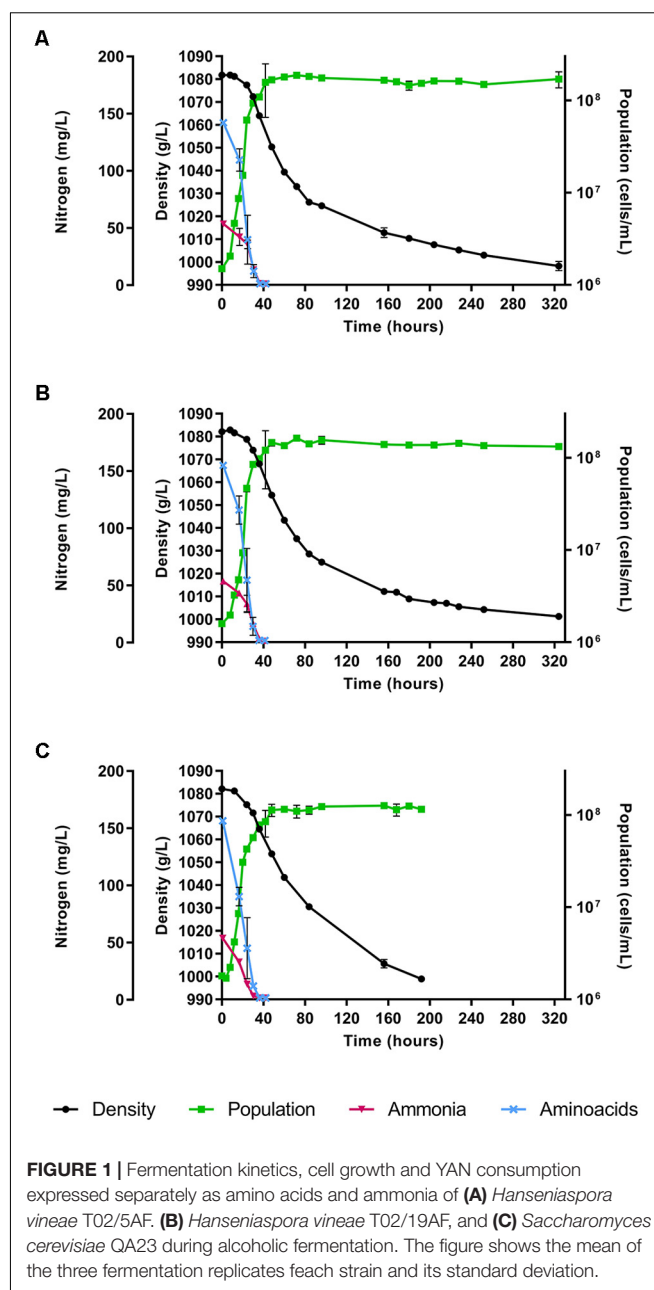
### Fermentation Kinetics and YAN Consumption

The fermentations were performed using a synthetic must with a nitrogen content of 140 mg YAN/L (corresponding to 190 mg N/L). Two strains of *H. vineae*, T02/5AF and T02/19AF, were evaluated, and *S. cerevisiae* strain QA23 was used as a control. Media density, cell growth and nitrogen content were assessed along with alcoholic fermentation. Both *H. vineae* strains showed a similar behavior in fermentation kinetics, cell growth and YAN consumption (Figure 1). These strains achieved a must density below 1000 g/L in approximately 13 days (324 h), while the *S. cerevisiae* strain was faster and reached this point in 8 days (192 h). The YAN was completely consumed by all the strains during the exponential growth phase that coincides with the initial stages of the fermentation (Figure 1). Ammonia and amino acids were consumed in 36 h by the *H. vineae* strains. Even though *S. cerevisiae* also consumed all the amino acids in 36 h, it exhausted the ammonia earlier, specifically before 30 h.

The consumption of each amino acid and ammonia was measured during the first 36 h of the different strain fermentations. Figure 2 and Table 2 show the evolution of their consumption at different time points. In general, a similar consumption pattern of amino acids occurred in both *H. vineae* strains and *S. cerevisiae*. In all cases, lysine, glutamic acid, cysteine, isoleucine, leucine, and phenylalanine were completely assimilated during the first 24 h. As for the previous amino acids, histidine was also exhausted during this period solely by the *H. vineae* strain T02/5AF. The slowest consumed amino acids, arginine, and valine, were still available in very small amounts after 30 h in both *S. cerevisiae* and *H. vineae*. The remaining amino acids were consumed between 24 and 36 h in every case.

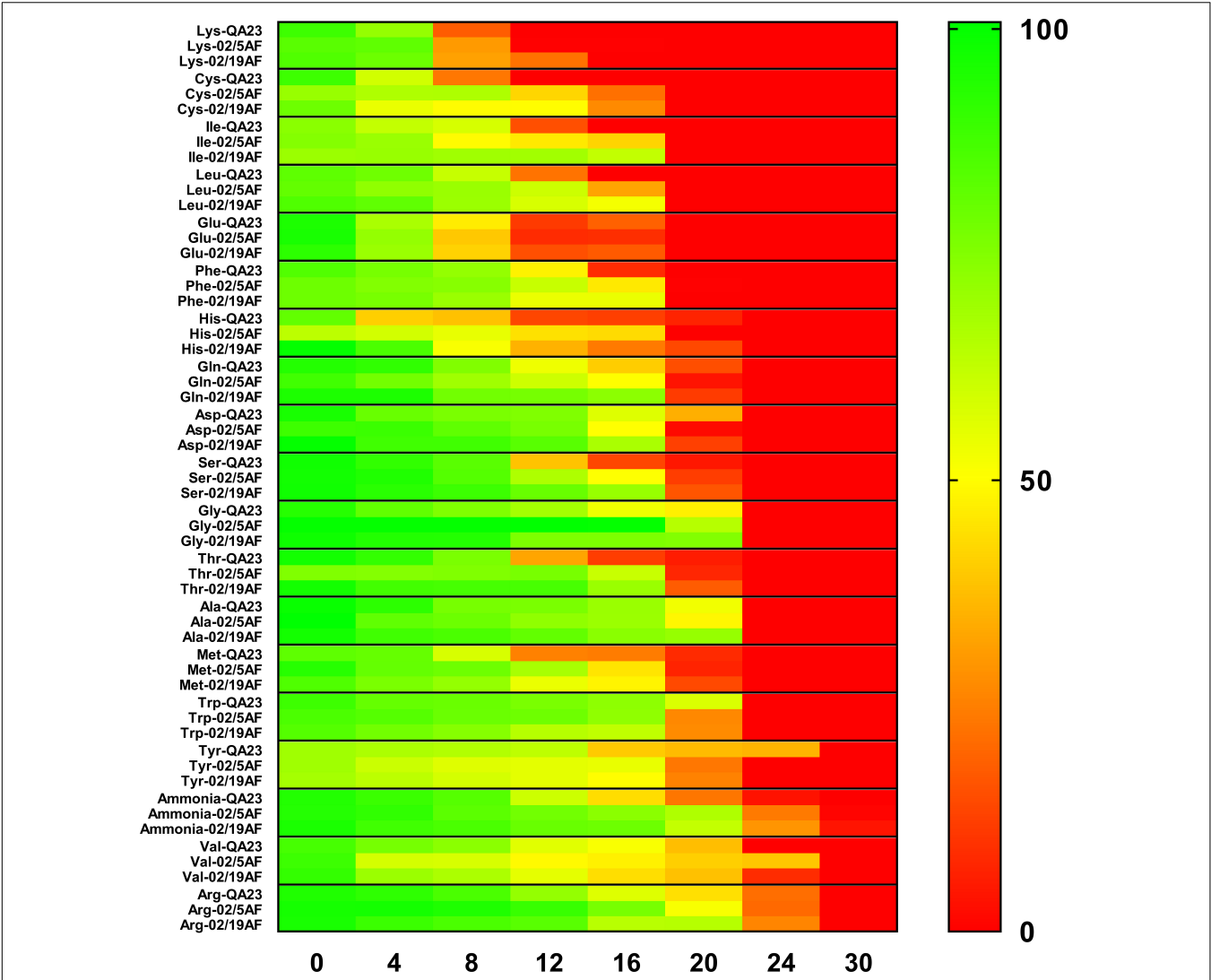
### Expression of NCR-Regulated Genes

Different genes related to NCR mechanism were evaluated for their homology in *H. vineae* including three permeases (*AGP1*, *GAP1*, and *MEP2*), one dehydrogenase (*PUT2*) and four transcriptional factors (*GAT1*, *GLN3*, *GZF3*, and *DAL80*). Except from *DAL80*, all the other genes had their homologous in *H. vineae* suggesting the presence of this nitrogen regulation mechanism in this yeast (Supplementary Table S3). To further check that this species displays this regulation, four genes related to nitrogen transport into the cell (*AGP1*, *GAP1*, and *MEP2*) and one gene related to proline utilization (*PUT2*) were selected to analyze their expression pattern during the first fermentation hours. These genes have been described and used in *S. cerevisiae* as markers for nitrogen limitation (Beltran et al.,



2004, 2007; Gutiérrez et al., 2013) and also as an indirect marker of transcriptional factors activity.

Figure 3 and Supplementary Table S4 show the expression evolution of the different genes during the first 48 h for each strain. Gene expression at 4 h was considered to be a reference, since the expression at 0 h corresponds to the inoculum that was nitrogen-depleted. The pattern of gene expression was similar for all the strains. *AGP1* was the only gene that was down-regulated during the fermentation, compared to the expression obtained at 4 h. Thus, the expression of *AGP1* is higher at the beginning of the fermentation (e.g., 4 h, our reference time), when the amino acid concentration is also higher, and decreases as the amino acids are consumed.



**FIGURE 2 |** Heatmap of the available percentage of each amino acid in the must during the first 30 h of *H. vineae* T02/5AF, *H. vineae* T02/19AF and *S. cerevisiae* QA23 fermentations. Green color corresponds to 100% of the total amino acid content available, and red corresponds to 0% of amino acid content available in the media. Standard deviations were always lower than 10% and have been avoided in the figure for clarity.

The other three genes started to be up-regulated at different time points along the fermentation depending on the yeast species. Therefore, both *H. vineae* strains activated the *GAP1* and *MEP2* expression after 24 h, and *PUT2* after 16 h of fermentation. Finally, *S. cerevisiae* QA23 expressed *GAP1*, *MEP2*, and *PUT2* after 16, 20 and 24 h of fermentation, respectively. Despite the differences between the yeast strains, the gradual activation or repression of the different genes coincided with

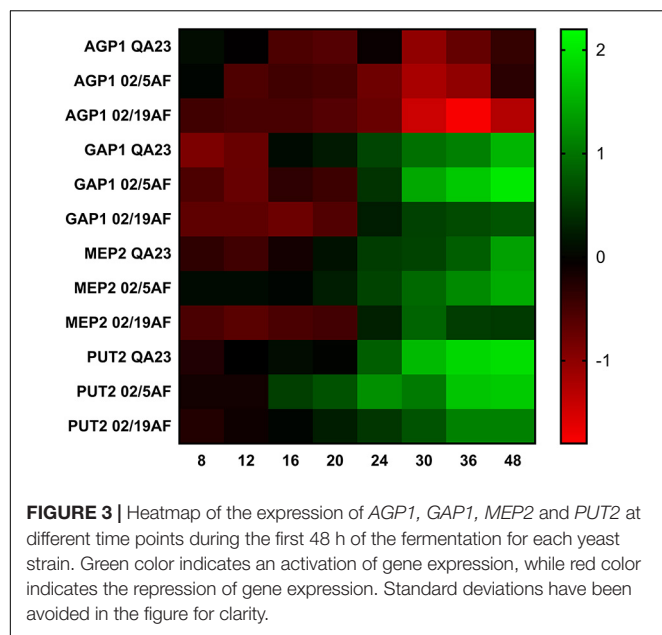
the progressive consumption of the amino acids and ammonia during fermentation.

DISCUSSION

In this study, we aimed to determine if *H. vineae*, a non-*Saccharomyces* yeast of oenological interest, displays the NCR

**TABLE 2 |** Time (h) required for each yeast strain to exhaust the different nitrogen compounds of the synthetic must.

	Consumed in 16–20 h	Consumed in 20–24 h	Consumed in 24–30 h	Consumed in 30–36 h
<i>H. vineae</i> T02/5AF	Lys	Glu, His, Cys, Ile, Leu, Phe	Asp, Ser, Gln, Gly, Thr, Ala, Met, Trp, Tyr	Arg, Val, NH <sub>4</sub> <sup>+</sup>
<i>H. vineae</i> T02/19AF	Lys	Glu, Cys, Ile, Leu, Phe	Asp, Ser, Gln, His, Gly, Thr, Ala, Trp, Tyr, Met	Arg, Val, NH <sub>4</sub> <sup>+</sup>
<i>S. cerevisiae</i> QA23	Lys, Cys, Ile, Leu	Glu, Phe	Asp, Ser, Gln, His, Gly, Thr, Ala, Val, Met Trp, NH <sub>4</sub> <sup>+</sup>	Arg, Tyr



mechanism under fermentation conditions. This metabolism has been thoroughly studied in *S. cerevisiae* during alcoholic fermentation (Beltran et al., 2004; Tesnière et al., 2015), and it was considered to be a reference in this study. Our results suggest that *H. vineae* exhibits an NCR mechanism similar to that of *S. cerevisiae*.

Fermentations using synthetic must with 140 mg YAN/L allowed us to evaluate ammonia and amino acid consumption together with the analysis of the expression of NCR-regulated genes during the first hours of fermentation. First, the nitrogen content of the synthetic must used in this study was not limiting, and it is considered to be the minimum concentration needed for yeasts to complete the alcoholic fermentation (Ribéreau-Gayon et al., 2006). In fact, all the strains tested in this work were able to complete the fermentation process (Figure 1) which agrees on previous reports (Ribéreau-Gayon et al., 2006). However, as we expected, *S. cerevisiae* finished the fermentation more quickly, because of its oenological abilities to resist the fermentation conditions. In addition, *H. vineae*, as well as *S. cerevisiae*, exhausted all the available YAN in 36 h even though *S. cerevisiae* consumed all the ammonia in 30 h, 6 h sooner than *H. vineae* (Figure 2 and Table 2). Medina et al. (2012) showed a competition for nutrients, especially nitrogen, in mixed fermentations of *S. cerevisiae* and *H. vineae*. The similar consumption of the nitrogen of these two yeasts observed in this study would explain the competition for this nutrient noted by Medina et al. (2012) in mixed fermentations.

In the same way as previous studies performed in *S. cerevisiae*, the strains evaluated exhausted all the YAN during the growth phase demonstrating how nitrogen availability plays a role as a limiting fermentation factor (Beltran et al., 2004; Crépin et al., 2012). In addition, the kinetic consumption of different nitrogen compounds has been evaluated simulating oenological conditions in different *S. cerevisiae* strains and conditions

in different studies. Beltran et al. (2007) demonstrated how temperature affects the amino acid intake, which affects yeast growth and metabolism. In our case, the fermentations proceeded at 22–23°C, and the consumption pattern of the nitrogen compounds was similar to that reported by Crépin et al. (2012). In fact, Crépin et al. (2012) classified nitrogen compounds in three groups according to their order of use by different *S. cerevisiae* strains: prematurely consumed, early consumed and late consumed. We classified the nitrogen compounds considering the time it took for them to be completely exhausted by each yeast strain (Table 2). However, we can observe that lysine is the fastest to be consumed by all the strains, and it is the one classified as prematurely consumed or that arginine, valine, tyrosine, and  $\text{NH}_4^+$  are the later ones to be completely exhausted, which belong to the late consumed compounds group established by Crépin et al. (2012). Considering these aspects, we observed that *H. vineae* has a similar behavior to *S. cerevisiae* in nitrogen uptake, and the variability of nitrogen compound preferences in *H. vineae* appear to also depend on the strain coinciding with previous studies on different *S. cerevisiae* strains (Crépin et al., 2014).

Interestingly, arginine was the slowest amino acid to be consumed in all cases. This amino acid is known as a non-preferred nitrogen source, since its support to yeast growth is very poor (Cooper, 1982), and it is the most stored amino acid in the vacuole during the growth phase (Crépin et al., 2014). In addition, the evaluation of arginase activity was proposed to be an indicator of the available nitrogen in fermentation (Carrasco et al., 2003), because as nitrogen becomes limiting, yeasts start to metabolize the stored nitrogen for additional growth (Crépin et al., 2014). In addition, Beltran et al. (2004) observed that the activation of arginase activity coincides with the mobilization of arginine, the ammonium depletion and the activation of *GAP1*. In this study, the highest arginine consumption coincided with ammonia depletion in *S. cerevisiae*. However, in *H. vineae*, arginine intake is simultaneous to that of ammonium, which may indicate that this yeast species uses a different way to store or consume this amino acid. The lower preference of *H. vineae* for ammonium is consistent with the reported poor effect of ammonium addition to agave juice fermentations compared to other nitrogen sources (Díaz-Montañó et al., 2010). In addition, *H. vineae* strains produced significantly lower levels of isobutyl alcohol derived from valine (Martín, 2016), which could be related to the slower consumption of this amino acid exhibited by this yeast species in this study.

The gene expression of *GAP1*, *MEP2*, and *PUT2* evolved from nitrogen-repressed to nitrogen-activated conditions as nitrogen was consumed in all cases. Between 16 and 30 h after inoculation with the different yeast strains, the gene expression of *GAP1*, *MEP2*, and *PUT2* began to be significantly activated. On the other hand, *AGP1* began to be repressed after 8 and 12 h of fermentations. As described before, *AGP1* acts as a sensor for amino acids, and its expression is induced by extracellular amino acids via SPS system, and down-regulated when the amino acids are consumed (Regenberg et al., 1999; Godard et al., 2007), which is consistent with our results. In the case of *GAP1* and *PUT2*, the transcription of these genes is known to

be activated under limiting nitrogen conditions (Forsberg and Ljungdahl, 2001; Gutiérrez et al., 2013), and this fact would explain their upregulation once the most preferred nitrogen compounds are consumed. Finally, we analyzed the ammonium permease *MEP2* expression, which is notably higher than other ammonium permeases (Ter Schure et al., 2000). Previous studies in *S. cerevisiae* have observed the activation of both *GAP1* and *MEP2* when ammonium is depleted (Beltran et al., 2004, 2005). However, in our study, the three strains tested showed a gradual activation of these two genes as ammonium and preferred amino acids were being consumed. From these results, we can deduce the activation of the transcriptional factors responsible for the expression of NCR genes.

The homology found on the NCR related proteins between *H. vineae* and *S. cerevisiae*, as well as the similarity in nitrogen consumption and the regulation of NCR genes, suggested the presence of the NCR mechanism in this non-*Saccharomyces* yeast. In addition, similarly to what has been described in *S. cerevisiae* (Beltran et al., 2004; Tesnière et al., 2015), the *H. vineae* wine yeasts evaluated entered the stationary phase coinciding with the exhaustion of nitrogen and consequently, the upregulation of the NCR genes. However, further research would be necessary to fully understand nitrogen metabolism in *H. vineae* and to elucidate if other mechanisms not regulated by NCR are responsible for nitrogen transport in this yeast.

Finally, the aim of this study was to determine if *H. vineae*, a non-*Saccharomyces* yeast of oenological interest, exhibits the NCR mechanism. Since nitrogen is one of the most limiting factors during alcoholic fermentation, knowing how it is metabolized gains importance. For that reason, we performed fermentations using synthetic must with an established nitrogen content, and we analyzed the nitrogen consumption and the expression of the NCR-regulated genes. The observed pattern of gene expression and nitrogen intake for the *H. vineae* strains and *S. cerevisiae* was similar, suggesting the presence of this regulatory mechanism in *H. vineae*. This study contributes to a better understanding of nitrogen

metabolism in the most active species in terms of the fermentation capacity of the genus *Hanseniaspora*, yeasts differentiated before the whole genome duplication event of the *Saccharomyces* group. In addition to our results, more studies are needed to completely understand nitrogen metabolism in this species.

## AUTHOR CONTRIBUTIONS

JL performed and designed the experiments, wrote the manuscript, and discussed and analyzed the results. VM, FG, and FC analyzed and discussed the results. MP analyzed and discussed the results, and wrote the manuscript. GB and AM designed the experiments, analyzed and discussed the results, and wrote the manuscript.

## FUNDING

This study was supported by the project from the Spanish Government AGL2015-73273-JIN. JL was financially supported by a Martí Franquès Fellowship from the Universitat Rovira i Virgili (URV) (2016PVF-PIPF-8).

## ACKNOWLEDGMENTS

The authors would like to thank Nicolàs Rozès for his help with HPLC analysis.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2018.00747/full#supplementary-material>

## REFERENCES

- Andorrà, I., Berradre, M., Rozès, N., Mas, A., Guillamón, J. M., and Esteve-Zarzoso, B. (2010). Effect of pure and mixed cultures of the main wine yeast species on grape must fermentations. *Eur. Food Res. Technol.* 231, 215–224. doi: 10.1007/s00217-010-1272-0
- Barquet, M., Martín, V., Medina, K., Pérez, G., Carrau, F., and Gaggero, C. (2012). Tandem repeat-tRNA (TRtRNA) PCR method for the molecular typing of non-*Saccharomyces* subspecies. *Appl. Microbiol. Biotechnol.* 93, 807–814. doi: 10.1007/s00253-011-3714-4
- Bell, S.-J., and Henschke, P. A. (2005). Implications of nitrogen nutrition for grapes, fermentation and wine. *Aust. J. Grape Wine Res.* 11, 242–295. doi: 10.1111/j.1755-0238.2005.tb00028.x
- Beltran, G., Esteve-Zarzoso, B., Rozès, N., Mas, A., and Guillamón, J. M. (2005). Influence of the timing of nitrogen additions during synthetic grape must fermentations on fermentation kinetics and nitrogen consumption. *J. Agric. Food Chem.* 53, 996–1002. doi: 10.1021/jf0487001
- Beltran, G., Novo, M., Rozès, N., Mas, A., and Guillamón, J. M. (2004). Nitrogen catabolite repression in *Saccharomyces cerevisiae* during wine fermentations. *FEMS Yeast Res.* 4, 625–632. doi: 10.1016/j.femsyr.2003.12.004
- Beltran, G., Rozès, N., Mas, A., and Guillamón, J. M. (2007). Effect of low-temperature fermentation on yeast nitrogen metabolism. *World J. Microbiol. Biotechnol.* 23, 809–815. doi: 10.1007/s11274-006-9302-6
- Bely, M., Sablayrolles, J. M., and Barre, P. (1990). Automatic detection of assimilable nitrogen deficiencies during alcoholic fermentations in enological conditions. *J. Ferment. Bioeng.* 70, 246–252. doi: 10.1016/0922-338X(90)90057-4
- Bisson, L. F. (1991). “Influence of nitrogen on yeast and fermentation of grapes,” in *Proceedings of the International Symposium on Nitrogen in Grapes and Wine*, Seattle, 78–79.
- Carrasco, P., Pérez-Ortín, J. E., and Del Olmo, M. (2003). Arginase activity is a useful marker of nitrogen limitation during alcoholic fermentations. *Syst. Appl. Microbiol.* 26, 471–479. doi: 10.1078/072320203322497518
- Carrau, F., Gaggero, C., and Aguilar, P. S. (2015). Yeast diversity and native vigor for flavor phenotypes. *Trends Biotechnol.* 33, 148–154. doi: 10.1016/j.tibtech.2014.12.009
- Cooper, T. G. (1982). “Nitrogen metabolism in *Saccharomyces cerevisiae*,” in *The Molecular Biology of the Yeast Saccharomyces: Metabolism and Gene Expression*, ed. J. N. Strathern (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press), 39–99.



- Crépin, L., Nidelet, T., Sanchez, I., Dequin, S., and Camarasa, C. (2012). Sequential use of nitrogen compounds by *Saccharomyces cerevisiae* during wine fermentation: a model based on kinetic and regulation characteristics of nitrogen permeases. *Appl. Environ. Microbiol.* 78, 8102–8111. doi: 10.1128/AEM.02294-12
- Crépin, L., Sanchez, I., Nidelet, T., Dequin, S., and Camarasa, C. (2014). Efficient ammonium uptake and mobilization of vacuolar arginine by *Saccharomyces cerevisiae* wine strains during wine fermentation. *Microb. Cell Fact.* 13, 1–13. doi: 10.1186/s12934-014-0109-0
- Curiel, J. A., Morales, P., Gonzalez, R., and Tronchoni, J. (2017). Different non-*Saccharomyces* yeast species stimulate nutrient consumption in *S. cerevisiae* mixed cultures. *Front. Microbiol.* 8:2121. doi: 10.3389/fmicb.2017.02121
- Deed, N. K., Van Vuuren, H. J. J., and Gardner, R. C. (2011). Effects of nitrogen catabolite repression and di-ammonium phosphate addition during wine fermentation by a commercial strain of *S. cerevisiae*. *Appl. Microbiol. Biotechnol.* 89, 1537–1549. doi: 10.1007/s00253-011-3084-y
- Díaz-Montaño, D. M., Favela-Torres, E., and Córdova, J. (2010). Improvement of growth, fermentative efficiency and ethanol tolerance of *Kloeckera africana* during the fermentation of agave tequilana juice by addition of yeast extract. *J. Sci. Food Agric.* 90, 321–328. doi: 10.1002/jsfa.3820
- Fleet, G. H. (2008). Wine yeasts for the future. *FEMS Yeast Res.* 8, 979–995. doi: 10.1111/j.1567-1364.2008.00427.x
- Forsberg, H., and Ljungdahl, P. (2001). Sensors of extracellular nutrients in *Saccharomyces cerevisiae*. *Curr. Genet.* 40, 91–109. doi: 10.1007/s002940100244
- Giorello, F. M., Berna, L., Greif, G., Camesasca, L., Salzman, V., Medina, K., et al. (2014). genome sequence of the native apiculate wine yeast *Hanseniaspora vineae* T02/19AF. *Genome Announc.* 2:e00530-14. doi: 10.1128/genomeA.00530-14
- Gobert, A., Tourdot-Maréchal, R., Morge, C., Sparrow, C., Liu, Y., Quintanilla-Casas, B., et al. (2017). Non-*Saccharomyces* yeasts nitrogen source preferences: impact on sequential fermentation and wine volatile compounds profile. *Front. Microbiol.* 8:2175. doi: 10.3389/fmicb.2017.02175
- Godard, P., Urrestarazu, A., Vissers, S., Kontos, K., Bontempi, G., van Helden, J., et al. (2007). Effect of 21 different nitrogen sources on global gene expression in the yeast *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* 27, 3065–3086. doi: 10.1128/MCB.01084-06
- Gómez-Alonso, S., Hermosín-Gutiérrez, I., and García-Romero, E. (2007). Simultaneous HPLC analysis of biogenic amines, amino acids, and ammonium ion as aminoenone derivatives in wine and beer samples. *J. Agric. Food Chem.* 55, 608–613. doi: 10.1021/jf062820m
- Gonzalez, R., Quirós, M., and Morales, P. (2013). Yeast respiration of sugars by non-*Saccharomyces* yeast species: a promising and barely explored approach to lowering alcohol content of wines. *Trends Food Sci. Technol.* 29, 55–61. doi: 10.1016/j.tifs.2012.06.015
- Gutiérrez, A., Chiva, R., Beltran, G., Mas, A., and Guillaumon, J. M. (2013). Biomarkers for detecting nitrogen deficiency during alcoholic fermentation in different commercial wine yeast strains. *Food Microbiol.* 34, 1–11. doi: 10.1016/j.fm.2012.12.004
- Hofman-Bang, J. (1999). Nitrogen catabolite repression in *Saccharomyces cerevisiae*. *Mol. Biotechnol.* 12, 35–73. doi: 10.1385/MB:12:1:35
- Jolly, N. P., Varela, C., and Pretorius, I. S. (2014). Not your ordinary yeast: Non-*Saccharomyces* yeasts in wine production uncovered. *FEMS Yeast Res.* 14, 215–237. doi: 10.1111/1567-1364.12111
- Lleixà, J., Martín, V., Portillo Mdel, C., Carrau, F., Beltran, G., and Mas, A. (2016). Comparison of fermentation and wines produced by inoculation of *Hanseniaspora vineae* and *Saccharomyces cerevisiae*. *Front. Microbiol.* 7:338. doi: 10.3389/fmicb.2016.00338
- López, M. C., Mateo, J. J., and Maicas, S. (2015). Screening of  $\beta$ -glucosidase and  $\beta$ -xylosidase activities in four non-*Saccharomyces* yeast isolates. *J. Food Sci.* 80, C1696–C1704. doi: 10.1111/1750-3841.12954
- Magasanik, B., and Kaiser, C. A. (2002). Nitrogen regulation in *Saccharomyces cerevisiae*. *Geno* 290, 1–18. doi: 10.1016/S0378-1119(02)00558-9
- Martín, V. (2016). *Hanseniaspora vineae*: Caracterización y su uso en la Vinificación. Ph.D. thesis, Universidades de la República, Montevideo.
- Martin, V., Giorello, F., Fariña, L., Minteguiaga, M., Salzman, V., Boido, E., et al. (2016). De novo synthesis of benzenoid compounds by the yeast *Hanseniaspora vineae* increases the flavor diversity of wines. *J. Agric. Food Chem.* 64, 4574–4583. doi: 10.1021/acs.jafc.5b05442
- Martin, V., Valera, M. J., Medina, K., Boido, E., and Carrau, F. (2018). Oenological Impact of the *Hanseniaspora/Kloeckera* yeast genus on wines—a review. *Fermentation* 4:76. doi: 10.3390/fermentation4030076
- Masneuf-Pomarede, I., Bely, M., Marullo, P., and Albertin, W. (2016). The genetics of non-conventional wine yeasts: current knowledge and future challenges. *Front. Microbiol.* 6:1563. doi: 10.3389/fmicb.2015.01563
- Medina, K., Boido, E., Dellacassa, E., and Carrau, F. (2012). Growth of non-*Saccharomyces* yeasts affects nutrient availability for *Saccharomyces cerevisiae* during wine fermentation. *Int. J. Food Microbiol.* 157, 245–250. doi: 10.1016/j.ijfoodmicro.2012.05.012
- Medina, K., Boido, E., Fariña, L., Gioia, O., Gomez, M. E., Barquet, M., et al. (2013). Increased flavour diversity of chardonnay wines by spontaneous fermentation and co-fermentation with *Hanseniaspora vineae*. *Food Chem.* 141, 2513–2521. doi: 10.1016/j.foodchem.2013.04.056
- Padilla, B., Gil, J. V., and Manzanares, P. (2016). Past and future of non-*Saccharomyces* yeasts: from spoilage microorganisms to biotechnological tools for improving wine aroma complexity. *Front. Microbiol.* 7:411. doi: 10.3389/fmicb.2016.00411
- Regenberg, B., Düring-Olsen, L., Kielland-Brandt, M. C., and Holmberg, S. (1999). Substrate specificity and gene expression of the amino-acid permeases in *Saccharomyces cerevisiae*. *Curr. Genet.* 36, 317–328. doi: 10.1007/s002940050506
- Ribéreau-Gayon, P., Glories, Y., Maujean, A., and Dubourdieu, D. (2006). *Handbook of Enology: The Microbiology of Wine and Vinifications*. Hoboken, NJ: Wiley. doi: 10.1002/0470010398
- Rollero, S., Bloem, A., Ortiz-Julien, A., Camarasa, C., and Divol, B. (2018). Altered fermentation performances, growth, and metabolic footprints reveal competition for nutrients between yeast species inoculated in synthetic grape juice-like medium. *Front. Microbiol.* 9:196. doi: 10.3389/fmicb.2018.00196
- Ståhlberg, A., Elbing, K., Andrade-Garda, J. M., Sjögreen, B., Forootan, A., and Kubista, M. (2008). Multiway real-time PCR gene expression profiling in yeast *Saccharomyces cerevisiae* reveals altered transcriptional response of ADH -genes to glucose stimuli. *BMC Genomics* 9:170. doi: 10.1186/1471-2164-9-170
- Taillandier, P., Lai, Q. P., Julien-Ortiz, A., and Brandam, C. (2014). Interactions between *Torulaspora delbrueckii* and *Saccharomyces cerevisiae* in wine fermentation: influence of inoculation and nitrogen content. *World J. Microbiol. Biotechnol.* 30, 1959–1967. doi: 10.1007/s11274-014-1618-z
- Ter Schure, E. G., Van Riel, N. A. W., and Verrips, C. T. (2000). The role of ammonia metabolism in nitrogen catabolite repression in *Saccharomyces cerevisiae*. *FEMS Microbiol. Rev.* 24, 67–83. doi: 10.1016/S0168-6445(99)00030-3
- Tesnière, C., Brice, C., and Blondin, B. (2015). Responses of *Saccharomyces cerevisiae* to nitrogen starvation in wine alcoholic fermentation. *Appl. Microbiol. Biotechnol.* 99, 7025–7034. doi: 10.1007/s00253-015-6810-z
- Varela, C. (2016). The impact of non-*Saccharomyces* yeasts in the production of alcoholic beverages. *Appl. Microbiol. Biotechnol.* 100, 9861–9874. doi: 10.1007/s00253-016-7941-6
- Viana, F., Belloch, C., Vallés, S., and Manzanares, P. (2011). Monitoring a mixed starter of *Hanseniaspora vineae*-*Saccharomyces cerevisiae* in natural must: impact on 2-phenylethyl acetate production. *Int. J. Food Microbiol.* 151, 235–240. doi: 10.1016/j.ijfoodmicro.2011.09.005

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Lleixà, Martín, Giorello, Portillo, Carrau, Beltran and Mas. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Expansion of a Telomeric *FLO/ALS*-Like Sequence Gene Family in *Saccharomycopsis fermentans*

Beatrice Bernardi, Yeseren Kayacan and Jürgen Wendland\*

Department of Bioengineering Sciences, Research Group of Microbiology, Functional Yeast Genomics, Faculty of Sciences and Bioengineering Sciences, Vrije Universiteit Brussel, Brussels, Belgium

## OPEN ACCESS

### Edited by:

Ed Louis,  
University of Leicester,  
United Kingdom

### Reviewed by:

Alexander DeLuna,  
Centro de Investigación y de Estudios  
Avanzados del Instituto Politécnico  
Nacional (CINVESTAV), Mexico  
Marti Aldea,  
Instituto de Biología Molecular  
de Barcelona (IBMB), Spain

### \*Correspondence:

Jürgen Wendland  
jurgen.wendland@vub.be

### Specialty section:

This article was submitted to  
Evolutionary and Genomic  
Microbiology,  
a section of the journal  
Frontiers in Genetics

Received: 11 July 2018

Accepted: 23 October 2018

Published: 13 November 2018

### Citation:

Bernardi B, Kayacan Y and  
Wendland J (2018) Expansion of a  
Telomeric *FLO/ALS*-Like Sequence  
Gene Family in *Saccharomycopsis*  
*fermentans*. *Front. Genet.* 9:536.  
doi: 10.3389/fgene.2018.00536

Non-*Saccharomyces* species have been recognized for their beneficial contribution to fermented food and beverages based on their volatile compound formation and their ability to ferment glucose into ethanol. At the end of fermentation brewer's yeast flocculate which provides an easy means of separation of yeasts from green beer. Flocculation in *Saccharomyces cerevisiae* requires a set of flocculation genes. These *FLO*-genes, *FLO1*, *FLO5*, *FLO9*, *FLO10*, and *FLO11*, are located at telomeres and transcription of these adhesins is regulated by Flo8 and Mss11. Here, we show that *Saccharomycopsis fermentans*, an ascomycete yeast distantly related to *S. cerevisiae*, possesses a very large *FLO/ALS*-like Sequence (*FAS*) family encompassing 34 genes. *Fas* proteins are variable in size and divergent in sequence and show similarity to the Flo1/5/9 family. *Fas* proteins show the general build with a signal peptide, an N-terminal carbohydrate binding PA14 domain, a central region differing by the number of repeats and a C-terminus with a consensus sequence for GPI-anchor attachment. Like *FLO* genes in *S. cerevisiae*, *FAS* genes are mostly telomeric with several paralogs at each telomere. We term such genes that share evolutionary conserved telomere localization “telologs” and provide several other examples. Adhesin expression in *S. cerevisiae* and filamentation in *Candida albicans* is regulated by Flo8 and Mss11. In *Saccharomycopsis* we identified only a single protein with similarity to Flo8 based on sequence similarity and the presence of a LisH domain.

**Keywords:** flocculation, comparative genomics, telomere, gene family, non-conventional yeast, *FLO8*, *MSS11*

## INTRODUCTION

Non-conventional yeasts have long been studied based on their identification in spontaneous fermentations around the world. Their additional value for fermentation is due to the more complex aroma profiles they produce compared to *Saccharomyces cerevisiae*. This was shown, e.g., for the use of *Torulaspora delbrueckii* and *Saccharomyces bayanus* in bread fermentations (Aslankoochi et al., 2016). The change of the previously dominating view of non-*Saccharomyces* yeasts as spoilage organisms to valuable and increasingly successful contributors to industrial fermentations may best be seen by the introduction of *Brettanomyces* (*Dekkera bruxellensis*) into beer and wine productions (Schiffedercker et al., 2014; Blomqvist and Passoth, 2015;

Steensels et al., 2015). Yet, other genera including *Candida*, *Hanseniaspora*, *Kluyveromyces*, *Metschnikowia*, and *Pichia* have been repeatedly identified in biodiversity studies analyzing spontaneous fermentations (Urso et al., 2008; Capozzi et al., 2015). Besides flavor contributions interest also stems from the ability of non-conventional yeasts to reduce the final alcohol content of fermented beverages (Ciani et al., 2016; Rossouw and Bauer, 2016).

In natural fermentations running over weeks diverse successions in the microbial diversity have been observed in coffee fermentations and in the fermentation of Belgian beers (lambic beers or red-brown ales) by autochthonous microorganisms (Silva et al., 2008; Wu et al., 2015; Snauwaert et al., 2016). Improved sequencing technology allows for large scale phylogenomics to explore the biodiversity and successions in such fermentations or to understand the evolution of yeasts in general and of a large variety of *Saccharomyces* strains in particular also with regard to human selection (Illegheems et al., 2012; De Filippis et al., 2017; Sternes et al., 2017; Peter et al., 2018). Taming this biodiversity, i.e., making use of different yeasts in mixed-fermentations, still proves challenging. To be commercially successful, non-conventional yeasts will have to compete with *S. cerevisiae* in terms of fermentation ability, alcohol tolerance, aroma compound formation, and compliance with current process technology.

One central aspect in the processing of fermentations is the ability of brewer's yeast strains to flocculate at the end of fermentation. This provides a cost-effective means to remove large quantities of yeast from the alcoholic beverage and reuse these yeasts by re-pitching them in a new brew (Stewart, 2018). Flocculation is a reversible aggregation of yeast cells forming large aggregates, flocs, that rapidly sediment – in contrast to settling of cells by gravitational sedimentation – to the bottom of the fermentation medium in bottom fermenting yeasts or float at the top in top fermenting yeasts. Aggregation is cation, mainly  $\text{Ca}^{2+}$ , dependent and flocs can be dispersed by the addition of EDTA (ethylenediaminetetraacetic acid) in *S. cerevisiae* and related strains used in fermentations (Stratford, 1989). Initiation of flocculation occurs at the end of fermentation when carbon or nitrogen sources are depleted and requires the synthesis of Flo proteins. In *S. cerevisiae* flocculation I positively regulated by the cAMP-protein kinase A pathway, MAP kinase signaling via Kss1 and repressed by Ssn6/Tup1 (Verstrepen et al., 2003; Verstrepen and Klis, 2006). The main transcription factor inducing the expression of *FLO* genes, however, is Flo8, which can dimerize with another protein of similar domain structure, Mss11. Laboratory yeast strains derived from S288C are non-flocculent as they harbor a nonsense mutation in *FLO8* at codon 142 converting a tryptophan codon into a stop codon. Flo8 activates the expression of *FLO* genes including the *FLO1/5/9* adhesin family as well as *FLO11*. *FLO*-genes of *S. cerevisiae* are located near the telomeres of different chromosomes and tend to show genetic instability by changes in size, mainly of the repeat length, e.g., induced by recombination between different telomeres (Carro et al., 2003).

Adhesins are ubiquitous cell surface proteins which facilitate cell–cell adhesion or cell-surface adherence. Not surprisingly, use of adhesion molecules serves different lifestyles, e.g., cell–cell attachment and biofilm formation on the one hand, but virulence and infection on the other. Fimbriae (thin appendages) of gram-negative bacteria act as adhesins, e.g., by themselves or by expressing a minor adhesin component located at the fimbrial tip. Fimbriae bind to carbohydrate residues, e.g., the adhesin FimH to D-mannose (Klemm and Schembri, 2000). Carbohydrate binding proteins, i.e., lectins, are ubiquitous in nature and occur in plants, animals, and fungi (and even viruses) (Boyd and Shapleigh, 1954; Sharon and Lis, 2004; Hassan et al., 2015; Hirabayashi et al., 2015). With genomics discovery of many more lectin genes a protein family-based classification of lectins – instead of a carbohydrate-binding specificity-based classification – was introduced (reviewed by Finn et al., 2014). In the homobasidiomycete *Coprinopsis cinerea* three galectins (encoded by *CGL1-3*) are expressed in the fruit bodies (Boulianne et al., 2000; Wälti et al., 2008). Mutations in the carbohydrate binding site of lectins may alter their carbohydrate binding specificity (Hassan et al., 2015).

In ascomycetes the adhesins from *Candida albicans* (*ALS* genes) and *S. cerevisiae* (*FLO* genes) are by far the best characterized (for details on other ascomycete adhesins see Lipke, 2018). Yet, most adhesins have a similar domain structure: N-terminal secretion signal peptide, a conserved N-term with ligand binding domain which is crucial for the functional diversity of adhesins, a central repeat domain that may be highly N- and O-glycosylated and a C-terminal domain that allows the addition of a GPI-anchor and thus the covalent attachment to the cell wall (Lipke, 2018). This offers facile ways to use bioinformatics data mining for the discovery novel adhesins. Adhesins differ by the length and sequence of their internal repeats. Generation of length polymorphisms may occur via DNA-replication errors or due to unequal sister chromatid exchanges (Verstrepen and Fink, 2009). In *C. albicans* and *C. glabrata* adhesins are termed agglutinin-like sequence (ALS) and epithelial adhesins (EPA), respectively (Gabaldon et al., 2013; Hoyer and Cota, 2016). For *C. albicans* other adhesin families have been described, namely the HWP and HYR families (De Groot et al., 2013) and in *S. cerevisiae* Aga1 and Fig2 are adhesins involved in mating and biofilm formation (Brückner and Mösch, 2012).

Fungal adhesins mediate contact interactions of cells with the environment. This can be for “social” behavior, e.g., in mating, colony and/or fruit body formation and biofilm formation or for “aggressive” behavior mediating host–pathogen interactions as seen in the human pathogens *C. glabrata* and *C. albicans* (Dranginis et al., 2007). *Saccharomycopsis* species have been described as fungal necrotrophs that kill other fungi via penetration pegs (Lachance and Pang, 1997). They may therefore have a dual use for their adhesins as they could employ them for flocculation at the end of fermentation and/or for attaching to fungal prey cells at the onset of their attack.

Non-conventional yeasts may introduce new flavors to alcoholic beverage fermentation but should conform with current process technology. *Saccharomycopsis* yeasts, more closely related



to *Wickerhamomyces* and *Ascoidea* species than to *S. cerevisiae*, have previously been found in spontaneous fermentations. *S. fibuligera* and *S. fermentans*, for example, were found in rice wine or palm wine fermentations (Ouoba et al., 2012; Kurtzman and Robnett, 2013; Carroll et al., 2017; Farh et al., 2017). This demonstrates that this genus harbors suitable and experienced strains for alcoholic beverage fermentation and thus warrants further analysis. For *S. fibuligera* a whole genome sequence analysis has been published indicating a gene repertoire for starch degradation; additionally, a hybridization event between two closely related species has been discovered (Choo et al., 2016). We recently presented a draft genome sequence of *S. fermentans*, which we now analyze in more detail focussing on the *FAS* gene family (Hesselbart et al., 2018).

Interestingly, we found an amplification of the *FAS* gene family at *S. fermentans* telomeres in a similar manner previously observed in *S. cerevisiae*. We termed those orthologous or paralogous genes that share evolutionarily conserved positions at telomeres “telologs” and identified several additional telologs between *S. cerevisiae* and *S. fermentans*. The *Fas* family of *S. fermentans* was compared to *S. cerevisiae* Flo proteins and *C. albicans* Als proteins. Additionally, a gene with similarity to the *FLO8/MSS11* transcription factors was identified that may be instrumental in regulating this gene set for flocculation at the end of fermentation. Furthermore, cell–cell adhesion may be a key element in initiating necrotrophic mycoparasitism, the ability of *S. fermentans* to penetrate and kill prey fungi to acquire their nutrients.

## MATERIALS AND METHODS

### Strains and Culture Conditions

*Saccharomycopsis fermentans* (CBS 7830, wild type, heterothallic) and the lager yeast strain Weihenstephan WS34/70 (allotetraploid) were grown in rich medium (YPD; 1% yeast extract, 2% bacto peptone, 2–20% glucose) at 30°C. Mat formation was assayed on low agar YPD plates containing 0.3% agar as described previously (Reynolds and Fink, 2001; Cullen, 2015). The culture for the flocculation assay was prepared by inoculating 50 mL YPD with 500 µL of either *S. fermentans* or WS34/70 from a water stock in a 250 mL Erlenmeyer flask. The cultures were incubated at 30°C in a rotary shaker with 150 rpm for 48 h. The flocculation test was done in the following way: 5 mL of each yeast culture was diluted with 5 mL of YPD and placed in a glass tube. The samples were rigorously vortexed and then placed vertically in a stand to monitor flocculation. To test if flocculation was calcium dependent EDTA (ethylenediaminetetraacetic acid, 50 mM final concentration) was added.

### Draft Genome Sequencing and Assembly

Draft genome sequencing and the assembly strategy of the *S. fermentans* genome was recently published (Hesselbart et al., 2018). The genome sequence is available at GenBank under accession number JNFW00000000. The *FAS* genes are listed in **Supplementary Table S1**.

## Gene and Protein Bioinformatic Analyses

A more detailed comparative genomic analysis of the *S. fermentans* genome will be published elsewhere. The scaffolds of the *Saccharomycopsis* genome were compared against the *Saccharomyces* Genome Database<sup>1</sup> using BLAST tools (available at <http://blast.ncbi.nlm.nih.gov>). This identified the set of *Fas* proteins based on their sequence similarity to *S. cerevisiae* the *FLO1/5/9* family. *Fas* proteins were further analyzed using several webtools with default settings as follows: for the presence of signal peptides the SignalP 4.1. server at <http://www.cbs.dtu.dk/services/SignalP/> was used. The PA14 domain was identified using the NCBI conserved domain tool available at <https://www.ncbi.nlm.nih.gov/Structure/bwrpsb/bwrpsb.cgi>. GPI anchors and omega site predictions were done at <http://gpcr.biocomp.unibo.it/predgpi/>. Repeats sequences of central domains of *Fas* proteins were identified using RADAR<sup>2</sup>. The individual repeats of all *Fas* genes were extracted (488 in total) and aligned using MegAlign of the DNA Lasergene 15 software package. The multiple sequence alignment generated with MegAlign was used as input sequence alignment for weblogo at <https://weblogo.berkeley.edu/examples.html>. This generated a consensus sequence for the internal repeats of the *Fas* proteins. The Flo proteins of *S. cerevisiae* were retrieved from SGD and the Als proteins of *C. albicans* from the *Candida* Genome Database<sup>3</sup>. Consensus sequences for the internal repeats of Flo proteins and Als proteins were processed in the same way as described above for *Fas* proteins. A sequence distance table comparing protein sequence identities of fungal adhesins was generated by with MegAlign (**Supplementary Figure S1**). To generate a dendrogram of fungal adhesins full length adhesins from *S. fermentans*, *S. cerevisiae*, and *C. albicans* were aligned using MegAlign. Bootstrapping was performed on the alignment using standard settings and 1000 replicas.

### FLO8 Identification and Alignment

The *S. fermentans* genome provides a weak hit against either *S. cerevisiae* Flo8 or Mss11 (sequences obtained from the *Saccharomyces* genome database; see text footnote 1). Sequences of other yeast species similar to ScFlo8 and ScMss11 (see text footnote 1) were retrieved from NCBI and used in additional genome wide blast searches<sup>4</sup> or in alignments (done with MegAlign using default settings). These include *Ashbya gossypii* (AFL194W and AGL300C<sup>5</sup>); *C. albicans* SC5314 (Flo8, C6\_04350cp\_a; Mss11, CR\_04840C\_A); *Cyberlindnera jadinii* (CEP23573); *Eremothecium cymbalariae* DBVPG#7215 (Flo8, XP\_003647814 and Mss11, XP\_003647183); *Lachancea thermotolerans* CBS6340 (Flo8, XP\_002552501; Mss11, XP\_002554632); *Saccharomycopsis crataegensis* CBS 6447 (JNFX000000000), *S. fodiens* CBS 8332

<sup>1</sup><https://www.yeastgenome.org/>

<sup>2</sup><https://www.ebi.ac.uk/Tools/pfa/radar/>

<sup>3</sup><http://www.candidagenome.org/>

<sup>4</sup><https://blast.ncbi.nlm.nih.gov/Blast.cgi>

<sup>5</sup><http://agd.unibas.ch>



(JNFV00000000); *Wickerhamomyces anomalus* NRRL Y-366-8 (Flo8, XP\_019036814.1; Mss11 XP\_019038767); and *Wickerhamomyces ciferrii* (Flo8, XP\_011273448; Mss11, XP\_011275532).

Dendrograms are based on full length protein alignments and bootstrapping was done with 1000 replicas. For the identification of LisH domain containing proteins in *S. fermentans* all non-overlapping translated ORF sequences were searched at NCBI using the pfam database<sup>6</sup> with default settings.

## RESULTS

### Mat Formation and Ca<sup>2+</sup>-Dependent Flocculation in *Saccharomycopsis fermentans*

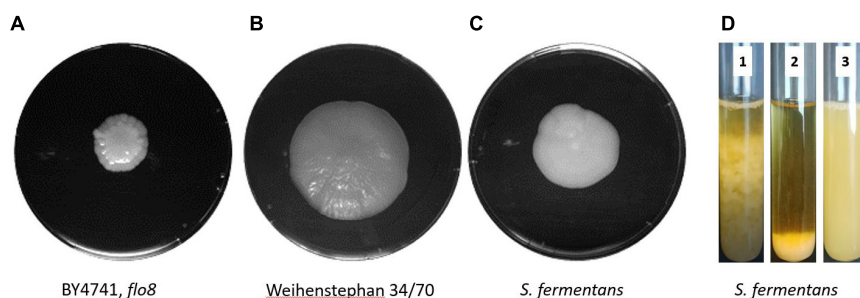
We observed flocculation in *S. fermentans* cultures at the end of the growth phase. To compare the ability to form biofilms we used a mat formation assay as described previously (Reynolds and Fink, 2001; Cullen, 2015). We compared growth of *S. fermentans* with the non-flocculating laboratory yeast strain BY4741 and the lager yeast *Weihenstephan* 34/70. On low rigidity agar plates (0.3%) *S. fermentans* mat formation was not covering an area as large as that of the lager yeast strain but was substantially more spread out than that of BY4741 (Figures 1A–C). *S. fermentans* cultures were grown into stationary phase and flocculation was monitored over a short time interval. In *S. fermentans* flocculation is much faster than sedimentation of cells by gravity with the result that after 1 min cells formed a pellet at the bottom of the test tube (Figure 1D). One of the hall marks of flocculation is its dependency on Ca<sup>2+</sup> cations. Flocculation can thus be inhibited by the addition of ion chelating molecules such as EDTA (Soares, 2011). Also, in *S. fermentans* flocculation is abolished in the presence of EDTA indicating that *S. fermentans* employs a similar mechanism to generate yeast flocs as brewer's yeasts (Figure 1D).

<sup>6</sup><https://www.ncbi.nlm.nih.gov/Structure/bwrpsb/bwrpsb.cgi>

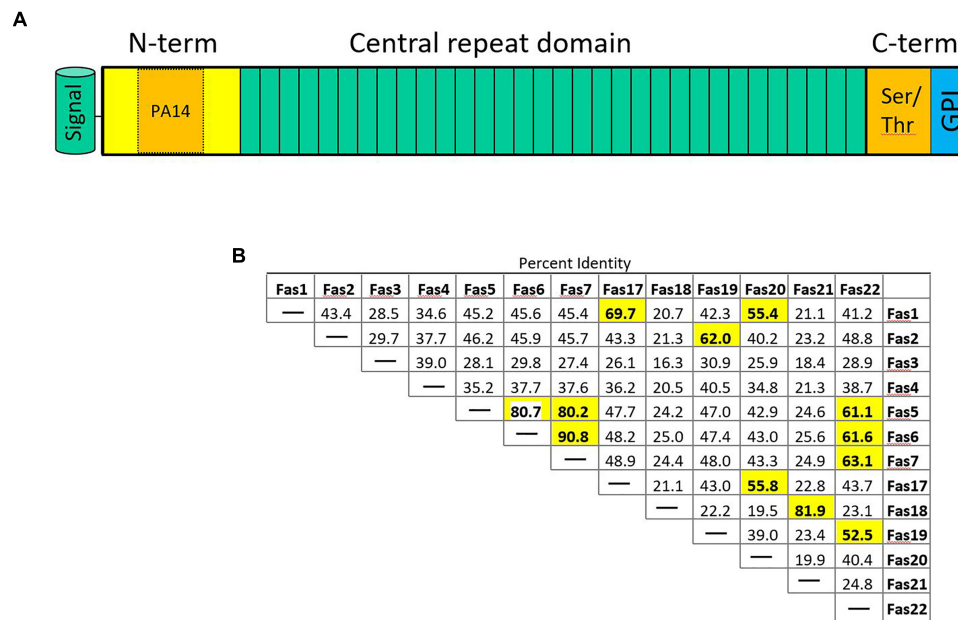
### Identification of the *S. fermentans* FAS Gene Family

Flocculation genes in *S. cerevisiae* belong to several classes. Flocculation itself is defined as the asexual aggregation of cells into flocs (Stratford, 1989; Bony et al., 1997). Thus, sexual agglutinins, e.g., encoded by *AGA1* and *SAG1*, that in *S. cerevisiae* promote cell to cell adhesion during mating will not be further discussed here. In *S. cerevisiae* the *FLO1/5/9* gene family is distinct from two other flocculins encoded by *FLO10* and *FLO11*. Yet, all *S. cerevisiae* *FLO* genes show the canonical domain architecture (Figure 2A). This includes an N-terminal signal peptide is followed by a PA14 domain – the name of this domain is derived from “protective antigen” a bacterial toxin found in *Bacillus anthracis* (Rigden et al., 2004). This domain is not only found in adhesins but also in glycosidases and glycosyltransferases consistent with a function in carbohydrate binding (Goossens and Willaert, 2010). In adhesins the PA14 motif makes up a part of the N-terminal domain, which is followed by a central domain of various length consisting of similar sized repeats. The C-terminal domain may bear sites for O- and N-glycosylation and harbors a recognition site for the addition of a glycosyl-phosphatidylinositol (GPI) anchor with which adhesins are inserted into the cell wall.

*Saccharomycopsis fermentans* FAS genes were identified by blast searches showing highest sequence similarities in their N-termini with *S. cerevisiae* *FLO* genes. In total 34 FAS genes were identified (Supplementary Table S1). While the *FLO1/5/9* family members show more than 85% identity on the amino acid sequence level Fas proteins identified in *S. fermentans* are much more divergent. Only a few Fas protein pairs show more than 90% sequence identity over the entire lengths of their proteins: this includes Fas5/Fas10, Fas6/Fas7, Fas8/Fas28, Fas 14/Fas29, and Fas15/Fas30 (Figure 2B and Supplementary Figure S1). These gene pairs may have evolved more “recently” by gene duplication. *FAS6* and *FAS7* are directly adjacent, while the other gene pairs may subsequently have been transferred to other loci, e.g., by reciprocal translocations (Nag et al., 2004). Several protein pairs show a high sequence identity over their N-term. This adds Fas21/Fas23, Fas9/pseudoFas31, and Fas22/Fas33 to previous set. A large size variation can be found with the smallest



**FIGURE 1 |** Mat formation in yeast. *S. cerevisiae* BY4741, *flo8* (A), the lager yeast *Weihenstephan* 34/70 (B), and *S. fermentans* (C) were grown on low-rigidity 0.3% agar YPD plates for 1 week at 25°C prior to photography. (D) Flocculation of *S. fermentans* in liquid media. The same culture was used to indicate speed of flocculation at time point zero, when stirring stopped (1, left) and after 60 s (2, middle). Then after the addition of EDTA to a final concentration of 50 mM after 60 s (3, right) indicating a block in flocculation.



**FIGURE 2 | (A)** Schematic representation of fungal adhesins. The N-terminal signal peptide is followed by an N-terminal domain that contains the PA14 domain promoting protein–carbohydrate interaction. The central domain contains a variable number of repeats (3–36 in *S. fermentans*). The C-terminal domain is rich in Ser/Thr residues and harbors an omega site for GPI-anchor attachment. **(B)** Sequence distance table indicating pair-wise protein sequence identities of the indicated Fas proteins whose coding genes are located on *S. fermentans* chromosome 1 and 4. Sequence identities >50% are highlighted in bold print.

FAS gene, FAS27, coding for a 591 aa protein and the largest, FAS12, encoding a protein of 1723 aa. There are variable length C-termini in Fas proteins that allow for differential glycosylation patterns. Amongst the 34 FAS genes three may be classified as pseudogenes. These are FAS3, FAS18, and FAS31. This is due to either lack of a signal peptide or a lacking start codon in the latter. The other FAS genes conform to the standard domain structure of adhesins with a ~20 aa signal peptide, the N-terminal domain including the PA14 domain, the central repeat rich domain and the C-terminal domain with omega sites required for GPI-anchoring. The repetitive sequences within FAS genes may cause recombination events between closely related sequences, i.e., either within the same gene, between FAS genes or between a FAS gene and a pseudogene, and thus generate size variations leading to an increase or decrease in gene sizes as described for *S. cerevisiae* (Verstrepen et al., 2005).

## Analysis of the Repeat Structure of *S. fermentans* Fas Proteins

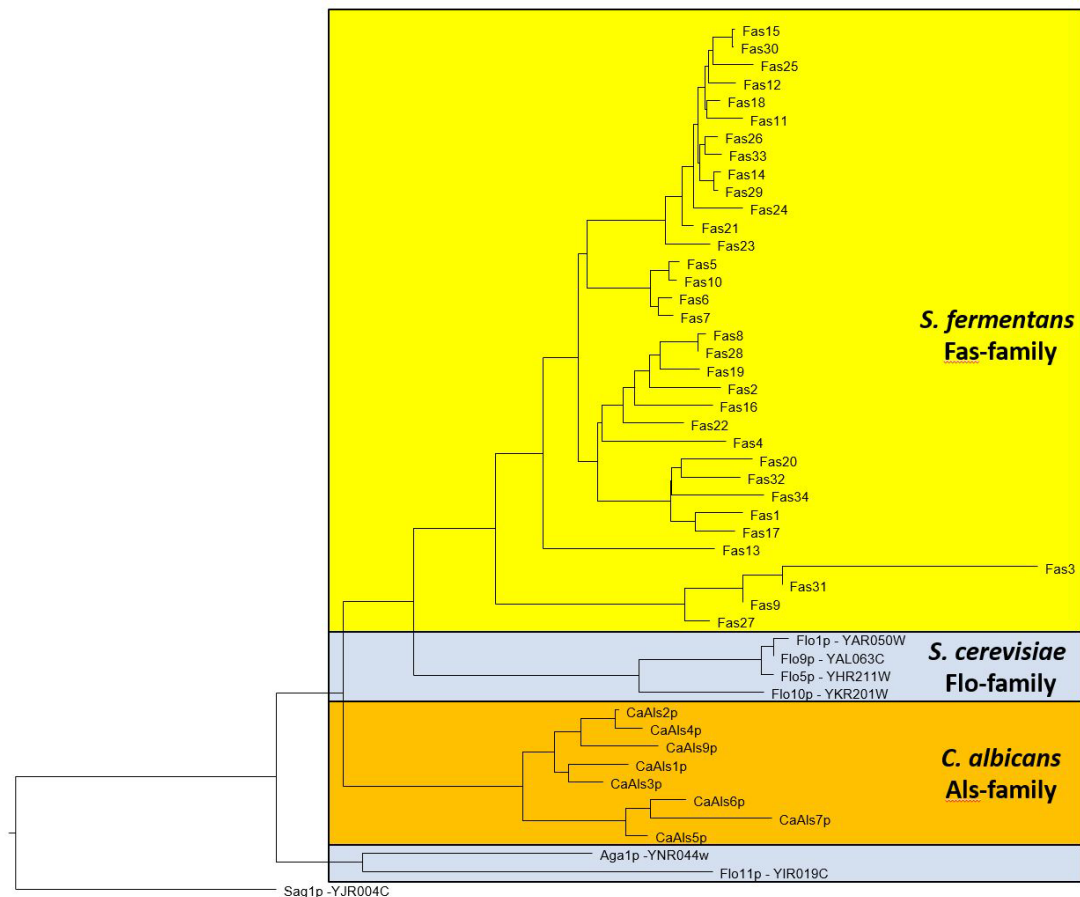
The large size differences in Fas proteins is, mainly due to the various number of repeats in the central region. For example, Fas27 has only 3 repeats while Fas12 encodes 36 repeats while on average 14 repeats can be found. We aligned the different adhesins and a dendrogram indicates that adhesins from *S. fermentans*, *C. albicans*, and *S. cerevisiae* form distinct groups (Figure 3). We then went on and examined the repeat sequences of *S. fermentans* Fas proteins and compared them with the repeats found in *S. cerevisiae* Flo proteins and *C. albicans* Als proteins. In total we identified 488 repeats in Fas proteins

(see Supplementary Table S1), 39 repeats in Flo1, Flo5, Flo9; 10 repeats in Flo10; 41 repeats in Flo11 and 140 in the Als proteins, respectively. These repeats were trimmed, aligned and consensus sequences were established using weblogo (see section “Materials and Methods”). Flo10 and Flo11 harbor distinct repeats different from each other and from the Flo1/5/9 family. In Flo10 there are two variants a 27 aa and an extended 36 aa repeat adding an invariant 9 aa sequence of AAANYTSSF to 4 of the repeats. Similarly, Flo11 has a basic 12 aa repeat which, in half of the repeats, is extended by the tripeptide PTP (Figure 4).

In Als proteins the repeat length is uniformly 36 aa with a large degree of sequence identity. The repeats within the Flo1/5/9 family are also highly conserved and are 45 aa in length. Repeats in the Fas proteins are on average 36 aa long. However, there is substantial divergence and all repeats align into a 46 aa consensus sequence. Comparison of the consensus sequences shows a high content of Ser/Thr residues and also conserved Pro residues with the repeats (Figure 4).

## *S. fermentans* FAS Genes Constitute a Telomeric Gene Family

In *S. cerevisiae* several gene families are located at subtelomeric regions including the FLO genes (Teunissen and Steensma, 1995; Luo and Van Vuuren, 2009). The FLO genes can be found at five different telomeres belonging to four chromosomes. Also, pseudogenes of FLO-like sequences were found, e.g., on chromosome I (Bussey et al., 1995; Figure 5A). We have analyzed the location of FAS genes in *S. fermentans*. Two chromosomal scaffolds corresponding to *S. fermentans* chromosome 1 and



**FIGURE 3 |** Dendrogram showing relationships between *S. fermentans*, *S. cerevisiae*, and *C. albicans* adhesins. Proteins were aligned using ClustalW and a tree based on this alignment is presented.

chromosome 4 were analyzed in detail (**Figure 5B**). At *TEL1R* and *TEL4R* a sequence repeat corresponding to  $TG_3(GA)_{2-4}$  was found and at *TEL1L* a transposon marks the end of the scaffold sequence. Interestingly, each of these chromosomes harbor several *FAS* genes at their telomeric ends. Sequence identity between these Fas proteins is much lower than in the *S. cerevisiae* and *C. albicans* adhesins except for the right arm of chromosome 1. Here Fas5 has 80% identity with Fas6 and Fas7, while Fas6 and Fas7 even share 90% aa sequence identity (see **Figure 2B**). Most of the *FAS* genes are apparently located at (sub)-telomeric regions, with at least one exception: *FAS4* was found to be at an internal position on chromosome 1.

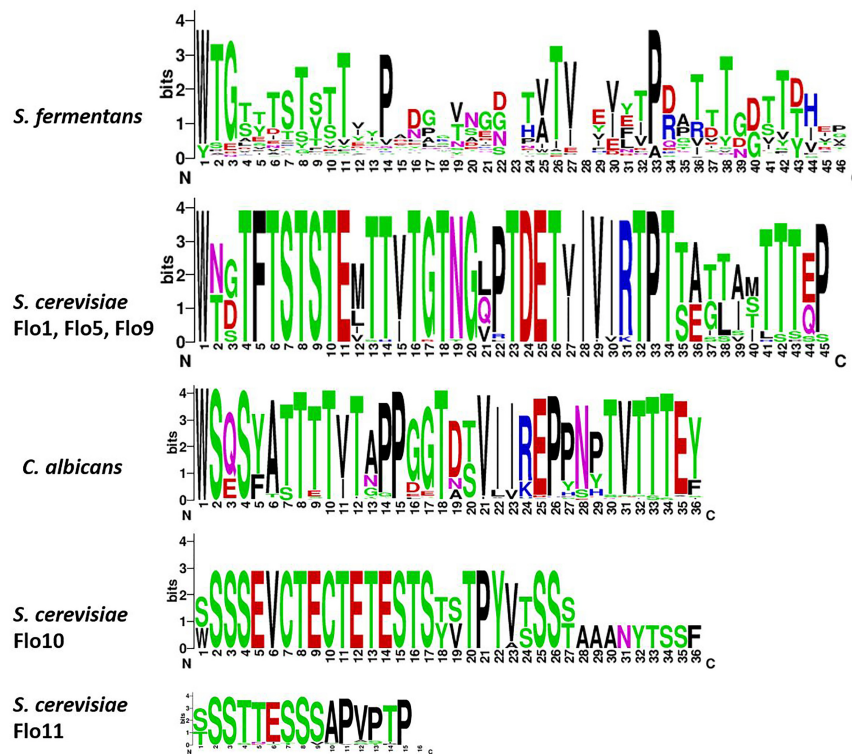
The *S. fermentans* genome was found to contain several gene families, including the *FAS* genes but also proteases and chitinases and transporters (Hesselbart et al., 2018). Within these expanded gene families, we could identify 26 aspartic proteases encoded by homologs of either YLR120C/*YPS1* (5 genes) or YLR121C/*YPS3* (21 genes) in *S. cerevisiae*, 22 homologs of a *S. cerevisiae* chitinase encoded by YLR186C/*CTS1* and, for example, 10 homologs of the YGR260W/*TNA1* high affinity nicotinic acid permease. Here we found that protease paralogs of *S. cerevisiae* *YPS3/YLR121C* and genes coding for nicotinic acid permeases, paralogs of

*TNA1/YGR260W*, are also located at telomere ends. *YPS3* genes were present at all four telomeres and *TNA1* genes were found at two of these telomeres (**Figure 5**).

Additionally, at these *S. fermentans* telomeres several homologs of *S. cerevisiae* genes were found that in *S. cerevisiae* are also located in sub-telomeric regions. These include *YIR042C* of unknown function; *OPT1/YJL212C*, an oligopeptide transporter; *OXPI/YKL215C*, a 5-oxoprolinase; *AIP1/YNR074C*, a homolog of the mammalian Apoptosis-Inducing Factor; and *ARR3/YPR201W*, a transporter required for resistance to arsenic compounds (**Figure 5**). These genes represent telologs, i.e., genes sharing a conserved telomer localization – yet not necessarily at the ancestral locus. Due to the evolutionary distance of the genera *Saccharomycopsis* and *Saccharomyces* it may be expected that these genes kept their telomeric position also in other genera and thus may be useful in genome assemblies.

### Identification of *FLO8* in *Saccharomycopsis fermentans*

Flo8 is a transcription factor that regulates the expression of *S. cerevisiae* *FLO* genes (Goossens and Willaert, 2010). *C. albicans*



**FIGURE 4 |** Sequence enrichment plots of adhesin repeats. For the *S. fermentans* and *C. albicans* all repeats of Fas and Als proteins, respectively, were combined. For *S. cerevisiae* the Flo1, Flo5, and Flo9 repeats were combined and Flo10 and Flo11 repeats were treated separately. Repeats were trimmed and aligned using ClustalW and then sequence enrichment plots were calculated using weblogo.

*FLO8* is essential for hyphal morphogenesis and is required for the expression of *ALS1* (Cao et al., 2006). In *S. cerevisiae* and *C. albicans* a second gene, *MSS11*, is also involved in expression of adhesins, together with *FLO8* (Bester et al., 2006; Su et al., 2009). Flo8 and Mss11 contain N-terminal LisH domains (pfam08513) and CaMss11 was shown to interact with CaFlo8 via this domain (Su et al., 2009). We performed BLASTp searches querying the translated ORF-datasets of *S. fermentans*, *S. crataegensis*, and *S. fodiens* for Flo8/Mss11 sequences. The best hit was derived from *Wickerhamomyces ciferrii* (4.3e-023) Flo8p as queries. We aligned Flo8/Mss11 sequences of the three *Saccharomycopsis* species with Flo8 and Mss11 orthologs from other yeasts (Figure 6). The deduced tree indicates a well supported separation between Mss11 and Flo8 proteins and a thus a placement of the *Saccharomycopsis* protein sequences with fungal Flo8 proteins. However, the protein sequences are highly divergent, and similarities are confined to a small N-terminal region.

The key domain of *S. cerevisiae* and *C. albicans* Flo8 and Mss11 proteins is a LisH domain. This allowed an independent search approach from blastp searches. Therefore, we used all non-overlapping translated ORFs from the draft genome sequence of *S. fermentans* and searched the conserved domain pfam database for LisH-domain containing proteins. Only two hits were retrieved, one to Flo8 (*e*-value 5.0e-05, pfam08513) and a second one to a *S. cerevisiae* homolog of Sif2 (*e*-value 2.2e-06,

pfam08513), which is known to harbor a LisH domain also in *S. cerevisiae*. This suggests that in *S. fermentans* only one protein corresponding to Flo8/Mss11 is present, which we named *FLO8*. Similarly, only one gene coding for Flo8 was found in *S. fibuligera*, *S. fodiens*, and *S. crataegensis* suggesting that this finding may not be due to the incompleteness of the draft genome sequences even though the closely related *Wickerhamomyces* genus harbors *FLO8* and *MSS11* genes.

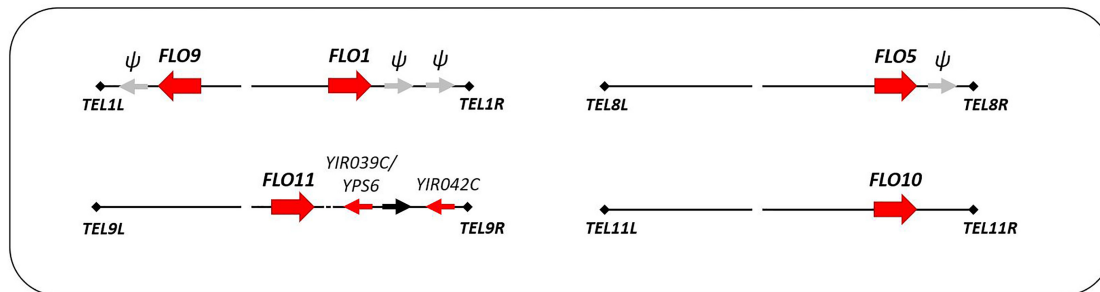
An alignment of N-terminal sequences of Flo8 and Mss11 proteins shows the similarity of these proteins in the region encompassing the LisH-domain (Figure 7). In Flo8 and Mss11 glutamine rich regions can be found. These are mostly internal. However, *S. fermentans* Flo8 shows an extended N-terminal region with an enlarged poly-Q-repeat of 89 residues (Figure 7).

## DISCUSSION

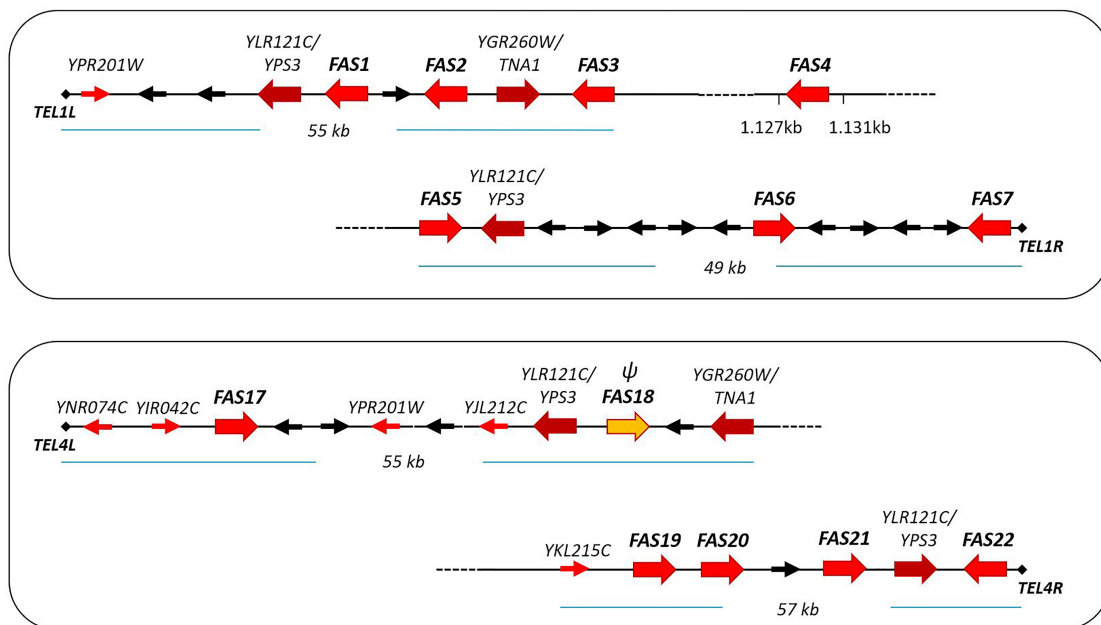
Alcoholic beverages such as wine and beer have been produced for several millennia and today the wine and beer sectors constitute key industries in world-wide beverage production. Brewer's yeasts have been the workhorses for these industries and besides their ability for alcoholic fermentation and flavor production their flocculation at the end of fermentation is most convenient to separate yeast slurries from the produced beverage



A

*Saccharomyces cerevisiae*

B

*Saccharomycopsis fermentans*

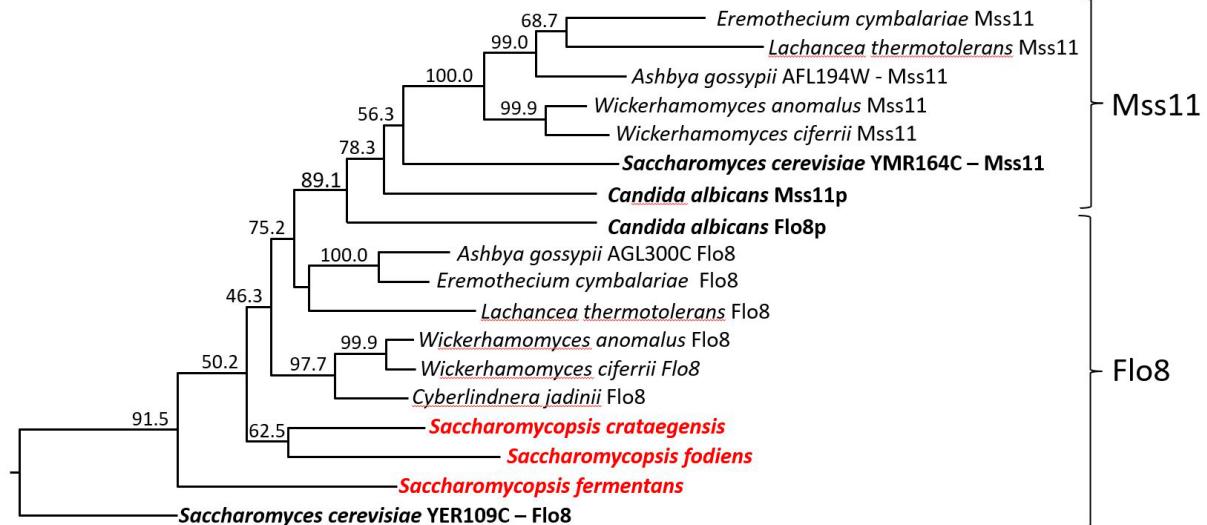
**FIGURE 5 |** Schematic presentation of adhesin gene localization at telomeric ends of *S. cerevisiae* chromosomes (A) and *S. fermentans* chromosomes (B). Adhesin genes are highlighted with large red arrows. Other *S. fermentans* genes belonging to gene families (YPS and TNA) are also represented by large arrows. Genes present at telomere loci in both *S. cerevisiae* and *S. fermentans*, so called telologs, are shown as small red arrows. Additional genes are drawn as black arrows and gray arrows with “ψ” mark *FLO*-like pseudogenes and *FAS18*.

(Verstrepen et al., 2003). With the craft beer movement came further challenges and innovations in the beverage producing sectors. One is the search for non-conventional, i.e., non-*Saccharomyces*, yeasts to generate more diversity and richness in flavor production. The other is the requirement for strains to be compatible with *S. cerevisiae*, e.g., in co-fermentations, but also with existing brewing technology. Here flocculation plays a major role.

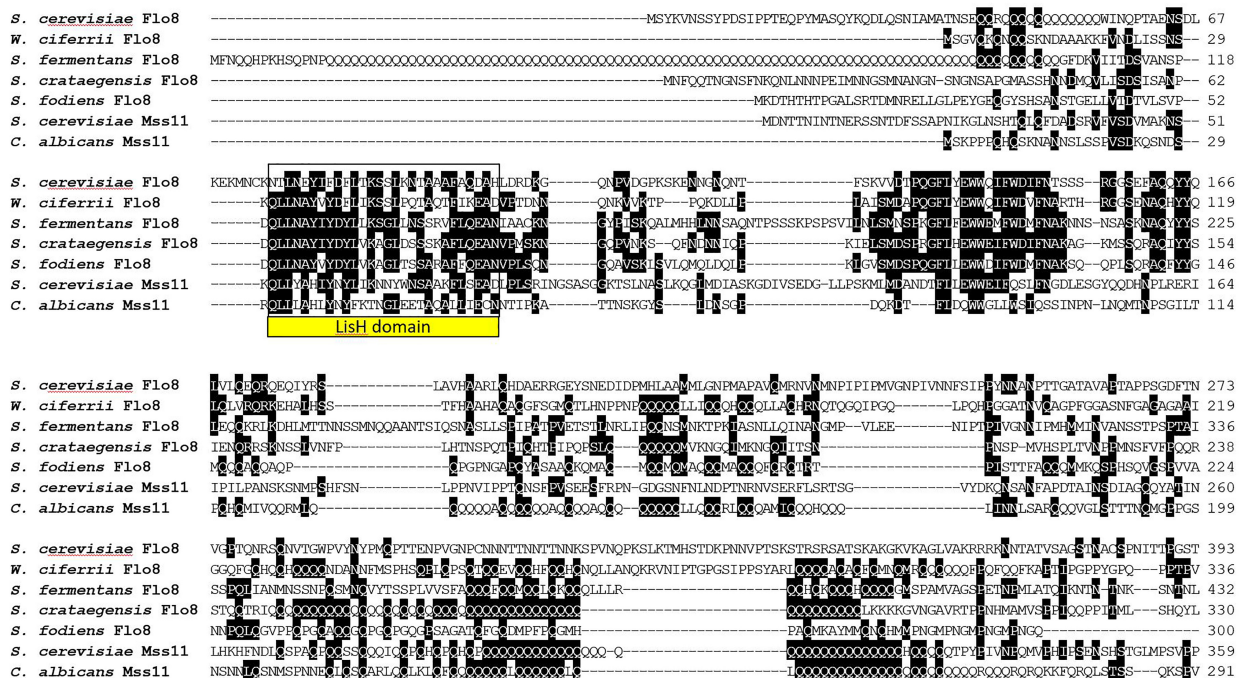
The molecular mechanism of flocculation has been studied for decades and excellent reviews provide detailed insight (Goossens and Willaert, 2010; Soares, 2011; Lipke, 2018). The hall mark of flocculation is based on Flo protein–carbohydrate (mannose) interaction between yeast cells. *S. cerevisiae* harbors different adhesins and particularly the *FLO1/5/9* family is promoting

flocculation, while *FLO11* regulates pseudohyphal and invasive growth and sexual adhesins are expressed during mating of haploid yeast cells (Erdman et al., 1998; Lo and Dranginis, 1998). Flocculation occurs in vegetative cells and is calcium-dependent (Stratford, 1989). In lager yeasts this results in the drop-out of flocs to the bottom of the fermentation vessel from where they efficiently can be collected to initiate a new fermentation. In other fungal systems, adhesins promote fungal virulence, e.g., in *C. albicans* or *C. glabrata* (Sui et al., 2017; Lopez-Fuentes et al., 2018).

In *S. fermentans* only paralogs to the *FLO1/5/9* family of *S. cerevisiae* were found, but not *FLO10* and *FLO11*. A comparison of the *FAS* gene family of *S. fermentans* with the *FLO/ALS* gene families in *S. cerevisiae* and *C. albicans*



**FIGURE 6 |** Dendrogram of Flo8 and Mss11 orthologs from different yeast species. Tree showing the positioning of *Saccharomycopsis* Flo8 proteins with other LisH-domain containing proteins from diverse yeast genera. Protein sequences were obtained via NCBI, aligned with ClustalW and the numbering indicates bootstrap values obtained with 1000 replicas. The Mss11 and Flo8 groups are indicated.



**FIGURE 7 |** Alignment of LisH-domaining proteins from the indicated fungi. Only the N-terminal part of the alignment is shown. The LisH dimerization domain is boxed. Proteins were aligned using ClustalW and two or more matching residues for each position in the alignment are shaded in black.

shows that this gene family is much larger in *S. fermentans* and consists of 34 genes (including three potential pseudogenes). A large degree of copy number variation (CNV) has recently been reported in *S. cerevisiae* wine strains. This particularly involved telomeric gene families and includes *FLO* genes and hexose transporters of the *HXT*-family, but also genes involved in copper

resistance (Steenwyk and Rokas, 2017). This diversity found in wine yeasts may, of course, be the result of human selection and the yeasts' adaptation to different fermentation environments. Besides CNV, there are also substantial size differences between adhesin proteins which are largely due to the number of internal tandem repeats. While individual tandem repeats in Flo-proteins

and Als-proteins are quite highly conserved in sequence, Fas proteins harbor a larger degree of divergence with only a third of the residues within these repeats being highly conserved. Several of the Fas proteins (namely Fas1,6,7,8,12,17,20,25,28,32) harbor a single RK-dibasic motif in the repeat region, while the KK-motif is found four times in Fas16 and once each in Fas25 and Fas32 and the KR sequence is found once each in Fas20 and Fas32. Such dibasic motifs may serve as proteolytic cleavage sites by aspartic proteases, e.g., the *S. cerevisiae* yapsins or the *C. albicans* Sap9/Sap10 proteases (Schild et al., 2011). In *S. fermentans* there is a large family of aspartic proteases available whose genes are also localized at telomeres like the *FAS* genes (see Figure 5).

These conserved residues with the central repeats may be involved in O-linked glycosylation and, in case of the conserved prolines, for structural purposes to establish rod-like structures (Jentoft, 1990). *S. fermentans* is strongly flocculant at the end of fermentation. This flocculation can be abolished by sequestration of  $\text{Ca}^{2+}$  ions by EDTA indicating a closely related flocculation mechanism compared to *S. cerevisiae* (Verstrepen and Klis, 2006).

One of the striking physiological features of *Saccharomycopsis* species is their predacious behavior. *Saccharomycopsis* species are auxotrophic for organic sulfur compounds and, e.g., upon starvation for methionine generate penetration pegs and kill fungal prey cells (Lachance et al., 2000). As one of the initial steps, cell-cell attachment could play a key role toward successful predation. However, how *S. fermentans* differentiates self from non-self to generate either flocs or initiate predation is currently unknown. In three *Saccharomycopsis* species, *S. fermentans*, *S. fodiens*, and *S. crataegensis*, several large gene families have been identified through draft-genome sequencing. This includes, the *FAS* genes, aspartic proteases (paralogs of *S. cerevisiae* YLR120C/YPS1-YLR121C/YPS3), chitinases (similar to YLR286C/CTS1), and transporters (YGR260W/TNA1 in *S. cerevisiae*). This suggests gene family evolution supported predacious behavior in *Saccharomycopsis*. Strikingly, the placement of several of these gene families, and particularly of the *FAS* and yapsins genes, at telomeric regions in *S. fermentans* resembles the evolution of gene families at *S. cerevisiae* telomeres. Similar amplifications of genes at subtelomeric regions were also found for aspartic protease (SAP) genes in *C. albicans* and chitinases in the mycoparasite *Trichoderma reesei* (Naglik et al., 2003; Liti and Louis, 2005; Seidl et al., 2005).

Due to the plasticity of telomeres, efforts to reconstruct ancestral gene orders at these positions are intrinsically difficult (Liti and Louis, 2005). When reconstructing the ancestral genome of yeast prior to the Whole Genome Duplication telomeric regions encompassing the terminal 10 genes could not be assigned to a single chromosome due to the fast turnover particularly within telomeres (Gordon et al., 2009). Here our analysis with *S. fermentans* shows that evolution at telomeres may have led to gene family expansions and relocation of ancestral telomeric genes. Remarkably, six genes that are present at *S. cerevisiae* telomeres were also found at telomeres in *S. fermentans*. For these genes we introduce the term “telologs,”

i.e., paralogs located at telomeric positions. This further suggests that phylogenomics of a sufficient amount of complete yeast genomes will eventually determine the telomere gene set of the yeast ancestor.

Flocculation genes are controlled by several mechanisms, including telomere silencing, epigenetic regulation, the cAMP-dependent protein kinase A pathway, a MAP kinase pathway, negatively by Sfl1 and positively by Flo8 and Mss11 (Teunissen et al., 1995; Kobayashi et al., 1996; Halme et al., 2004; Bester et al., 2006; Verstrepen and Klis, 2006; Fichtner et al., 2007). We have identified two *SFL1* paralogs in *S. fermentans*, one on chromosome 1 and another on chromosome 4, as well as in *S. fodiens* and *S. crataegensis*. On the other hand, extensive searches for *FLO8* and *MSS11* suggested that predator yeasts only contain one ortholog of a *FLO8*-like transcription factor best recognized by its LisH domain. This domain is part of the LUGS domain that is also conserved in *Arabidopsis thaliana* LUG (Leunig) and mouse *LIS1* (Kim et al., 2004; Shrestha et al., 2014). Additionally, also in *S. fibuligera* only one *FLO8* gene is present (Choo et al., 2016). Functional analysis of *S. fermentans FLO8* will determine its role in the expression of *FAS* genes and/or in predation. In *S. cerevisiae* and *C. albicans* Flo8 and Mss11 form heterodimers via their LisH-domains, while both Flo8 and Mss11 of *C. albicans* can also form homodimers (Kim et al., 2014). Overexpression of *CaFLO8* can suppress the *mss11* deletion but *MSS11* overexpression failed to rescue the hyphal growth defect of *flo8* in *C. albicans* (Su et al., 2009). This indicates a more important role of *FLO8* and may provide an explanation for the loss of *MSS11* in *Saccharomycopsis* to rely solely on Flo8 homodimers to direct flocculation. *S. fermentans FLO8* is one of only 24 genes encoding a poly-Q stretch. These poly-glutamines could additionally promote protein-protein interactions and thus in the case of *S. fermentans* Flo8 facilitate homodimerization (Perutz et al., 1994).

## CONCLUSION AND OUTLOOK

This work has generated new insight into the suitability of *S. fermentans* for industrial beverage fermentations. We found a similar telomere association of adhesin genes in *S. fermentans* as it is known for *FLO* genes in *S. cerevisiae*. *S. fermentans* harbor a large set of adhesins encoded by the *FAS* gene family, which could serve distinct purposes in flocculation or predation. Functionally, *S. fermentans* flocculation is phenotypically similar to flocculation in *S. cerevisiae* but apparently is regulated only by the Flo8 transcription factor and not by a heterodimer of Flo8 and Mss11 as in *S. cerevisiae* and *C. albicans*. As a next step the functional analysis of *Saccharomycopsis FLO8* for either flocculation or predation will be interesting to elucidate and also its ability to complement deletion of *ScFLO8* and/or *MSS11*. Introduction of *S. fermentans* as a novel non-conventional yeast in beer and wine fermentations may require selection of strains that are, e.g., adapted to stressful fermentation conditions and higher alcohol concentrations. In contrast to most brewer's yeasts with limited sexual reproduction abilities, *S. fermentans* is a homothallic yeast that is amenable to yeast breeding. Thus,



further characterization of this yeast and other species of the genus may lead to advanced molecular yeast breeding efforts to increase flavor diversity of alcoholic beverages in the future.

## DATA AVAILABILITY

All datasets analyzed for this study are included in the manuscript and the supplementary files or are available online.

## AUTHOR CONTRIBUTIONS

BB, YK, and JW contributed to conception and design of the study. All authors contributed to experimental or bioinformatic analyses. JW analyzed and interpreted the data, introduced the term “telolog,” and wrote the manuscript draft. All authors contributed to manuscript revision.

## REFERENCES

- Aslankooi, E., Herrera-Malaver, B., Rezaei, M. N., Steensels, J., Courtin, C. M., and Verstrepen, K. J. (2016). Non-conventional yeast strains increase the aroma complexity of bread. *PLoS One* 11:e0165126. doi: 10.1371/journal.pone.0165126
- Bester, M. C., Pretorius, I. S., and Bauer, F. F. (2006). The regulation of *Saccharomyces cerevisiae* FLO gene expression and Ca<sup>2+</sup> -dependent flocculation by Flo8p and Mss11p. *Curr. Genet.* 49, 375–383. doi: 10.1007/s00294-006-0068-z
- Blomqvist, J., and Passoth, V. (2015). Dekkera bruxellensis–spoilage yeast with biotechnological potential, and a model for yeast evolution, physiology and competitiveness. *FEMS Yeast Res.* 15:fov021. doi: 10.1093/femsyr/fov021
- Bony, M., Thines-Sempoux, D., Barre, P., and Blondin, B. (1997). Localization and cell surface anchoring of the *Saccharomyces cerevisiae* flocculation protein Flo1p. *J. Bacteriol.* 179, 4929–4936. doi: 10.1128/jb.179.15.4929-4936.1997
- Boulianne, R. P., Liu, Y., Aebi, M., Lu, B. C., and Kues, U. (2000). Fruiting body development in *Coprinus cinereus*: regulated expression of two galectins secreted by a non-classical pathway. *Microbiology* 146(Pt 8), 1841–1853. doi: 10.1099/00221287-146-8-1841
- Boyd, W. C., and Shapleigh, E. (1954). Specific precipitating activity of plant agglutinins (Lectins). *Science* 119:419. doi: 10.1126/science.119.3091.419
- Brückner, S., and Mösch, H. U. (2012). Choosing the right lifestyle: adhesion and development in *Saccharomyces cerevisiae*. *FEMS Microbiol. Rev.* 36, 25–58. doi: 10.1111/j.1574-6976.2011.00275.x
- Bussey, H., Kaback, D. B., Zhong, W., Vo, D. T., Clark, M. W., Fortin, N., et al. (1995). The nucleotide sequence of chromosome I from *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. U.S.A.* 92, 3809–3813. doi: 10.1073/pnas.92.9.3809
- Cao, F., Lane, S., Raniga, P. P., Lu, Y., Zhou, Z., Ramon, K., et al. (2006). The Flo8 transcription factor is essential for hyphal development and virulence in *Candida albicans*. *Mol. Biol. Cell* 17, 295–307. doi: 10.1091/mbc.e05-06-0502
- Capozzi, V., Garofalo, C., Chiriatti, M. A., Grieco, F., and Spano, G. (2015). Microbial terroir and food innovation: the case of yeast biodiversity in wine. *Microbiol. Res.* 181, 75–83. doi: 10.1016/j.micres.2015.10.005
- Carro, D., Garcia-Martinez, J., Perez-Ortin, J. E., and Pina, B. (2003). Structural characterization of chromosome I size variants from a natural yeast strain. *Yeast* 20, 171–183. doi: 10.1002/yea.955
- Carroll, E., Trinh, T. N., Son, H., Lee, Y. W., and Seo, J. A. (2017). Comprehensive analysis of fungal diversity and enzyme activity in nuruk, a Korean fermenting starter, for acquiring useful fungi. *J. Microbiol.* 55, 357–365. doi: 10.1007/s12275-017-7114-z
- Choo, J. H., Hong, C. P., Lim, J. Y., Seo, J. A., Kim, Y. S., Lee, D. W., et al. (2016). Whole-genome de novo sequencing, combined with RNA-Seq analysis, reveals unique genome and physiological features of the amylolytic yeast

## FUNDING

This research was supported by the European Union Marie Skłodowska-Curie Actions Innovative Training Network Aromagenesis (764364).

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2018.00536/full#supplementary-material>

**TABLE S1 |** Table providing sequence details of the *S. fermentans* FAS gene family indicating the length, number of internal repeats, closest relative within *S. fermentans* and the ORF and translated protein sequences of all 34 FAS genes.

**FIGURE S1 |** Table indicating pairwise amino acid sequence identity between adhesins of the *S. fermentans* Fas family, the *S. cerevisiae* Flo family and the *C. albicans* Als family.

- Saccharomycopsis fibuligera* and its interspecies hybrid. *Biotechnol. Biofuels* 9:246. doi: 10.1186/s13068-016-0653-4
- Ciani, M., Morales, P., Comitini, F., Tronchoni, J., Canonico, L., Curiel, J. A., et al. (2016). Non-conventional yeast species for lowering ethanol content of wines. *Front. Microbiol.* 7:642. doi: 10.3389/fmicb.2016.00642
- Cullen, P. J. (2015). The plate-washing assay: a simple test for filamentous growth in budding yeast. *Cold Spring Harb. Protoc.* 2015, 168–171. doi: 10.1101/pdb.prot085068
- De Filippis, F., La Storia, A., and Blaiotta, G. (2017). Monitoring the mycobiota during Greco di Tufo and aglianico wine fermentation by 18S rRNA gene sequencing. *Food Microbiol.* 63, 117–122. doi: 10.1016/j.fm.2016.11.010
- De Groot, P. W., Bader, O., De Boer, A. D., Weig, M., and Chauhan, N. (2013). Adhesins in human fungal pathogens: glue with plenty of stick. *Eukaryot. Cell* 12, 470–481. doi: 10.1128/EC.00364-12
- Dranginis, A. M., Rauceo, J. M., Coronado, J. E., and Lipke, P. N. (2007). A biochemical guide to yeast adhesins: glycoproteins for social and antisocial occasions. *Microbiol. Mol. Biol. Rev.* 71, 282–294. doi: 10.1128/MMBR.00037-06
- Erdman, S., Lin, L., Malczynski, M., and Snyder, M. (1998). Pheromone-regulated genes required for yeast mating differentiation. *J. Cell Biol.* 140, 461–483. doi: 10.1083/jcb.140.3.461
- Farh, M. E., Cho, Y., Lim, J. Y., and Seo, J. A. (2017). A diversity study of *Saccharomycopsis fibuligera* in rice wine starter nuruk, reveals the evolutionary process associated with its interspecies hybrid. *J. Microbiol.* 55, 337–343. doi: 10.1007/s12275-017-7115-y
- Fichtner, L., Schulze, F., and Baus, G. H. (2007). Differential Flo8p-dependent regulation of FLO1 and FLO11 for cell-cell and cell-substrate adherence of *S. cerevisiae* S288c. *Mol. Microbiol.* 66, 1276–1289. doi: 10.1111/j.1365-2958.2007.06014.x
- Finn, R. D., Miller, B. L., Clements, J., and Bateman, A. (2014). iPFam: a database of protein family and domain interactions found in the Protein Data Bank. *Nucleic Acids Res.* 42, D364–D373. doi: 10.1093/nar/gkt1210
- Gabalton, T., Martin, T., Marcet-Houben, M., Durrens, P., Bolotin-Fukuhara, M., Lespinet, O., et al. (2013). Comparative genomics of emerging pathogens in the *Candida glabrata* clade. *BMC Genomics* 14:623. doi: 10.1186/1471-2164-14-623
- Goossens, K., and Willaert, R. (2010). Flocculation protein structure and cell-cell adhesion mechanism in *Saccharomyces cerevisiae*. *Biotechnol. Lett.* 32, 1571–1585. doi: 10.1007/s10529-010-0352-3
- Gordon, J. L., Byrne, K. P., and Wolfe, K. H. (2009). Additions, losses, and rearrangements on the evolutionary route from a reconstructed ancestor to the modern *Saccharomyces cerevisiae* genome. *PLoS Genet.* 5:e1000485. doi: 10.1371/journal.pgen.1000485
- Halme, A., Bumgarner, S., Styles, C., and Fink, G. R. (2004). Genetic and epigenetic regulation of the FLO gene family generates cell-surface variation in yeast. *Cell* 116, 405–415. doi: 10.1016/S0092-8674(04)00118-7



- Hassan, M. A., Rouf, R., Tiralongo, E., May, T. W., and Tiralongo, J. (2015). Mushroom lectins: specificity, structure and bioactivity relevant to human disease. *Int. J. Mol. Sci.* 16, 7802–7838. doi: 10.3390/ijms16047802
- Hesselbart, A., Junker, K., and Wendland, J. (2018). Draft genome sequence of *Saccharomycopsis fermentans* CBS 7830, a predacious yeast belonging to the *Saccharomycetales*. *Genome Announc* 6:e01445-17. doi: 10.1128/genomeA.01445-17
- Hirabayashi, J., Tateno, H., Shikanai, T., Aoki-Kinoshita, K. F., and Narimatsu, H. (2015). The lectin frontier database (LfDB), and data generation based on frontal affinity chromatography. *Molecules* 20, 951–973. doi: 10.3390/molecules20010951
- Hoyer, L. L., and Cota, E. (2016). Candida albicans agglutinin-like sequence (Als) family vignettes: a review of als protein structure and function. *Front. Microbiol.* 7:280. doi: 10.3389/fmicb.2016.00280
- Illegheems, K., De Vuyst, L., Papalexandratou, Z., and Weckx, S. (2012). Phylogenetic analysis of a spontaneous cocoa bean fermentation metagenome reveals new insights into its bacterial and fungal community diversity. *PLoS One* 7:e38040. doi: 10.1371/journal.pone.0038040
- Jentoft, N. (1990). Why are proteins O-glycosylated? *Trends Biochem. Sci.* 15, 291–294.
- Kim, K. H., Hong, S. K., Hwang, K. Y., and Kim, E. E. (2014). Structure of mouse muskellin discoidin domain and biochemical characterization of its self-association. *Acta Crystallogr. D. Biol. Crystallogr.* 70, 2863–2874. doi: 10.1107/S139900471401894X
- Kim, M. H., Cooper, D. R., Oleksy, A., Devedjiev, Y., Derewenda, U., Reiner, O., et al. (2004). The structure of the N-terminal domain of the product of the lissencephaly gene Lis1 and its functional implications. *Structure* 12, 987–998. doi: 10.1016/j.str.2004.03.024
- Klemm, P., and Schembri, M. A. (2000). Fimbriae-assisted bacterial surface display of heterologous peptides. *Int. J. Med. Microbiol.* 290, 215–221. doi: 10.1016/S1438-4221(00)80118-6
- Kobayashi, O., Suda, H., Ohtani, T., and Sone, H. (1996). Molecular cloning and analysis of the dominant flocculation gene FLO8 from *Saccharomyces cerevisiae*. *Mol. Gen. Genet.* 251, 707–715.
- Kurtzman, C. P., and Robnett, C. J. (2013). Relationships among genera of the *Saccharomycotina* (Ascomycota) from multigene phylogenetic analysis of type species. *FEMS Yeast Res.* 13, 23–33. doi: 10.1111/1567-1364.12006
- Lachance, M. A., and Pang, W. M. (1997). Predacious yeasts. *Yeast* 13, 225–232. doi: 10.1002/(SICI)1097-0061(19970315)13:3<225::AID-YEA87>3.0.CO;2-I
- Lachance, M. A., Pupovac-Velikonja, A., Natarajan, S., and Schlag-Edler, B. (2000). Nutrition and phylogeny of predacious yeasts. *Can. J. Microbiol.* 46, 495–505. doi: 10.1139/w00-021
- Lipke, P. N. (2018). What we do not know about fungal cell adhesion molecules. *J. Fungi* 4:E59. doi: 10.3390/jof4020059
- Liti, G., and Louis, E. J. (2005). Yeast evolution and comparative genomics. *Annu. Rev. Microbiol.* 59, 135–153. doi: 10.1146/annurev.micro.59.030804.121400
- Lo, W. S., and Dranginis, A. M. (1998). The cell surface flocculin Flo11 is required for pseudohyphae formation and invasion by *Saccharomyces cerevisiae*. *Mol. Biol. Cell* 9, 161–171. doi: 10.1091/mbc.9.1.161
- Lopez-Fuentes, E., Gutierrez-Escobedo, G., Timmermans, B., Van Dijck, P., De Las Penas, A., and Castano, I. (2018). Candida glabrata's genome plasticity confers a unique pattern of expressed cell wall proteins. *J. Fungi* 4:E67. doi: 10.3390/jof4020067
- Luo, Z., and Van Vuuren, H. J. (2009). Functional analyses of PAU genes in *Saccharomyces cerevisiae*. *Microbiology* 155, 4036–4049. doi: 10.1099/mic.0.030726-0
- Nag, D. K., Suri, M., and Stenson, E. K. (2004). Both CAG repeats and inverted DNA repeats stimulate spontaneous unequal sister-chromatid exchange in *Saccharomyces cerevisiae*. *Nucleic Acids Res.* 32, 5677–5684. doi: 10.1093/nar/gkh901
- Naglik, J. R., Challacombe, S. J., and Hube, B. (2003). Candida albicans secreted aspartyl proteinases in virulence and pathogenesis. *Microbiol. Mol. Biol. Rev.* 67, 400–428. doi: 10.1128/MMBR.67.3.400-428.2003
- Ouoba, L. I., Kando, C., Parkouda, C., Sawadogo-Lingani, H., Diawara, B., and Sutherland, J. P. (2012). The microbiology of Bandji, palm wine of *Borassus akeassii* from Burkina faso: identification and genotypic diversity of yeasts, lactic acid and acetic acid bacteria. *J. Appl. Microbiol.* 113, 1428–1441. doi: 10.1111/jam.12014
- Perutz, M. F., Johnson, T., Suzuki, M., and Finch, J. T. (1994). Glutamine repeats as polar zippers: their possible role in inherited neurodegenerative diseases. *Proc. Natl. Acad. Sci. U.S.A.* 91, 5355–5358. doi: 10.1073/pnas.91.12.5355
- Peter, J., De Chiara, M., Friedrich, A., Yue, J. X., Pflieger, D., Bergstrom, A., et al. (2018). Genome evolution across 1,011 *Saccharomyces cerevisiae* isolates. *Nature* 556, 339–344. doi: 10.1038/s41586-018-0030-5
- Reynolds, T. B., and Fink, G. R. (2001). Bakers' yeast, a model for fungal biofilm formation. *Science* 291, 878–881. doi: 10.1126/science.291.5505.878
- Rigden, D. J., Mello, L. V., and Galperin, M. Y. (2004). The PA14 domain, a conserved all-beta domain in bacterial toxins, enzymes, adhesins and signaling molecules. *Trends Biochem. Sci.* 29, 335–339. doi: 10.1016/j.tibs.2004.05.002
- Rossouw, D., and Bauer, F. F. (2016). Exploring the phenotypic space of non-Saccharomyces wine yeast biodiversity. *Food Microbiol.* 55, 32–46. doi: 10.1016/j.fm.2015.11.017
- Schifferdecker, A. J., Dashko, S., Ishchuk, O. P., and Piskur, J. (2014). The wine and beer yeast *Dekkera bruxellensis*. *Yeast* 31, 323–332. doi: 10.1002/yea.3023
- Schild, L., Heyken, A., De Groot, P. W., Hiller, E., Mock, M., De Koster, C., et al. (2011). Proteolytic cleavage of covalently linked cell wall proteins by Candida albicans Sap9 and Sap10. *Eukaryot. Cell* 10, 98–109. doi: 10.1128/EC.00210-10
- Seidl, V., Huemer, B., Seiboth, B., and Kubicek, C. P. (2005). A complete survey of *Trichoderma chitinases* reveals three distinct subgroups of family 18 chitinases. *FEBS J.* 272, 5923–5939. doi: 10.1111/j.1742-4658.2005.04994.x
- Sharon, N., and Lis, H. (2004). History of lectins: from hemagglutinins to biological recognition molecules. *Glycobiology* 14, 53R–62R. doi: 10.1093/glycob/cwh122
- Shrestha, B., Guragain, B., and Sridhar, V. V. (2014). Involvement of co-repressor LUH and the adapter proteins SLK1 and SLK2 in the regulation of abiotic stress response genes in Arabidopsis. *BMC Plant Biol.* 14:54. doi: 10.1186/1471-2229-14-54
- Silva, C. F., Batista, L. R., Abreu, L. M., Dias, E. S., and Schwan, R. F. (2008). Succession of bacterial and fungal communities during natural coffee (*Coffea arabica*) fermentation. *Food Microbiol.* 25, 951–957. doi: 10.1016/j.fm.2008.07.003
- Snauwaert, I., Roels, S. P., Van Nieuwerburg, F., Van Landschoot, A., De Vuyst, L., and Vandamme, P. (2016). Microbial diversity and metabolite composition of Belgian red-brown acidic ales. *Int. J. Food Microbiol.* 221, 1–11. doi: 10.1016/j.ijfoodmicro.2015.12.009
- Soares, E. V. (2011). Flocculation in *Saccharomyces cerevisiae*: a review. *J. Appl. Microbiol.* 110, 1–18. doi: 10.1111/j.1365-2672.2010.04897.x
- Steensels, J., Daenen, L., Malcorps, P., Derdelinckx, G., Verachtert, H., and Verstrepen, K. J. (2015). Brettanomyces yeasts—from spoilage organisms to valuable contributors to industrial fermentations. *Int. J. Food Microbiol.* 206, 24–38. doi: 10.1016/j.ijfoodmicro.2015.04.005
- Steenwyk, J., and Rokas, A. (2017). Extensive copy number variation in fermentation-related genes among *Saccharomyces cerevisiae* wine strains. *G3* 7, 1475–1485. doi: 10.1534/g3.117.040105
- Sternes, P. R., Lee, D., Kutyna, D. R., and Borneman, A. R. (2017). A combined meta-barcoding and shotgun metagenomic analysis of spontaneous wine fermentation. *Gigascience* 6, 1–10. doi: 10.1093/gigascience/gix040
- Stewart, G. G. (2018). Yeast flocculation—sedimentation and flotation. *Fermentation* 4:28. doi: 10.3390/fermentation4020028
- Stratford, M. (1989). Evidence for two mechanisms of flocculation in *Saccharomyces cerevisiae*. *Yeast* 5, S441–S445.
- Su, C., Li, Y., Lu, Y., and Chen, J. (2009). Mss11, a transcriptional activator, is required for hyphal development in Candida albicans. *Eukaryot. Cell* 8, 1780–1791. doi: 10.1128/EC.00190-09
- Sui, X., Yan, L., and Jiang, Y. Y. (2017). The vaccines and antibodies associated with Als3p for treatment of Candida albicans infections. *Vaccine* 35, 5786–5793. doi: 10.1016/j.vaccine.2017.08.082
- Teunissen, A. W., and Steensma, H. Y. (1995). Review: the dominant flocculation genes of *Saccharomyces cerevisiae* constitute a new subtelomeric gene family. *Yeast* 11, 1001–1013. doi: 10.1002/yea.32011102
- Teunissen, A. W., Van Den Berg, J. A., and Steensma, H. Y. (1995). Transcriptional regulation of flocculation genes in *Saccharomyces cerevisiae*. *Yeast* 11, 435–446. doi: 10.1002/yea.320110506
- Urso, R., Rantsiou, K., Dolci, P., Rolle, L., Comi, G., and Cocolin, L. (2008). Yeast biodiversity and dynamics during sweet wine production as determined by molecular methods. *FEMS Yeast Res.* 8, 1053–1062. doi: 10.1111/j.1567-1364.2008.00364.x

- Verstrepen, K. J., Derdelinckx, G., Verachtert, H., and Delvaux, F. R. (2003). Yeast flocculation: what brewers should know. *Appl. Microbiol. Biotechnol.* 61, 197–205. doi: 10.1007/s00253-002-1200-8
- Verstrepen, K. J., and Fink, G. R. (2009). Genetic and epigenetic mechanisms underlying cell-surface variability in protozoa and fungi. *Annu. Rev. Genet.* 43, 1–24. doi: 10.1146/annurev-genet-102108-134156
- Verstrepen, K. J., Jansen, A., Lewitter, F., and Fink, G. R. (2005). Intragenic tandem repeats generate functional variability. *Nat. Genet.* 37, 986–990. doi: 10.1038/ng1618
- Verstrepen, K. J., and Klis, F. M. (2006). Flocculation, adhesion and biofilm formation in yeasts. *Mol. Microbiol.* 60, 5–15. doi: 10.1111/j.1365-2958.2006.05072.x
- Wälti, M. A., Thore, S., Aebi, M., and Kunzler, M. (2008). Crystal structure of the putative carbohydrate recognition domain of human galectin-related protein. *Proteins* 72, 804–808. doi: 10.1002/prot.22078
- Wu, R., Yu, M., Liu, X., Meng, L., Wang, Q., Xue, Y., et al. (2015). Changes in flavour and microbial diversity during natural fermentation of suan-cai, a traditional food made in Northeast China. *Int. J. Food Microbiol.* 211, 23–31. doi: 10.1016/j.ijfoodmicro.2015.06.028
- Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Bernardi, Kayacan and Wendland. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Multi-Omics Analysis of Fatty Alcohol Production in Engineered Yeasts *Saccharomyces cerevisiae* and *Yarrowia lipolytica*

## OPEN ACCESS

### Edited by:

Isabel Sá-Correia,  
University of Lisbon,  
Portugal

### Reviewed by:

Jennifer Gallagher,  
West Virginia University,  
United States  
Pau Ferrer,  
Autonomous University of Barcelona,  
Spain

### \*Correspondence:

Irina Borodina  
irbo@biosustain.dtu.dk

### †Present address:

Birgitta E. Ebert,  
Vickers Group, Australian  
Institute for Bioengineering and  
Nanotechnology, Brisbane City,  
QLD, Australia

### Specialty section:

This article was submitted to  
Evolutionary and Genomic  
Microbiology,  
a section of the journal  
Frontiers in Genetics

**Received:** 15 April 2019

**Accepted:** 17 July 2019

**Published:** 30 August 2019

### Citation:

Dahlin J, Holkenbrink C,  
Marella ER, Wang G, Liebal U,  
Lieven C, Weber D, McCloskey D,  
Ebert BE, Herrgård MJ, Blank LM  
and Borodina I (2019) Multi-Omics  
Analysis of Fatty Alcohol Production  
in Engineered Yeasts *Saccharomyces  
cerevisiae* and *Yarrowia lipolytica*.  
Front. Genet. 10:747.  
doi: 10.3389/fgene.2019.00747

Jonathan Dahlin<sup>1</sup>, Carina Holkenbrink<sup>1</sup>, Eko Roy Marella<sup>1</sup>, Guokun Wang<sup>1</sup>, Ulf Liebal<sup>2</sup>,  
Christian Lieven<sup>1</sup>, Dieter Weber<sup>2</sup>, Douglas McCloskey<sup>1</sup>, Birgitta E. Ebert<sup>2†</sup>,  
Markus J. Herrgård<sup>1</sup>, Lars Mathias Blank<sup>2</sup> and Irina Borodina<sup>1\*</sup>

<sup>1</sup> The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Kongens Lyngby, Denmark, <sup>2</sup> iAMB – Institute of Applied Microbiology, ABBt – Aachen Biology and Biotechnology, RWTH Aachen University, Aachen, Germany

Fatty alcohols are widely used in various applications within a diverse set of industries, such as the soap and detergent industry, the personal care, and cosmetics industry, as well as the food industry. The total world production of fatty alcohols is over 2 million tons with approximately equal parts derived from fossil oil and from plant oils or animal fats. Due to the environmental impact of these production methods, there is an interest in alternative methods for fatty alcohol production via microbial fermentation using cheap renewable feedstocks. In this study, we aimed to obtain a better understanding of how fatty alcohol biosynthesis impacts the host organism, baker's yeast *Saccharomyces cerevisiae* or oleaginous yeast *Yarrowia lipolytica*. Producing and non-producing strains were compared in growth and nitrogen-depletion cultivation phases. The multi-omics analysis included physiological characterization, transcriptome analysis by RNAseq, <sup>13</sup>C-metabolic flux analysis, and intracellular metabolomics. Both species accumulated fatty alcohols under nitrogen-depletion conditions but not during growth. The fatty alcohol-producing *Y. lipolytica* strain had a higher fatty alcohol production rate than an analogous *S. cerevisiae* strain. Nitrogen-depletion phase was associated with lower glucose uptake rates and a decrease in the intracellular concentration of acetyl-CoA in both yeast species, as well as increased organic acid secretion rates in *Y. lipolytica*. Expression of the fatty alcohol-producing enzyme fatty acyl-CoA reductase alleviated the growth defect caused by deletion of hexadecenal dehydrogenase encoding genes (*HFD1* and *HFD4*) in *Y. lipolytica*. RNAseq analysis showed that fatty alcohol production triggered a cell wall stress response in *S. cerevisiae*. RNAseq analysis also showed that both nitrogen-depletion and fatty alcohol production have substantial effects on the expression of transporter encoding genes in *Y. lipolytica*. In conclusion, through this multi-omics study, we uncovered some effects of fatty alcohol production on the host metabolism. This knowledge can be used as guidance for further strain improvement towards the production of fatty alcohols.

**Keywords:** fatty alcohol, metabolome, <sup>13</sup>C-fluxome, transcriptome, *Yarrowia lipolytica*, *Saccharomyces cerevisiae*

## INTRODUCTION

Fatty alcohols are used as detergents and surfactants in personal care products such as soaps, shampoos, or creams. The global market for fatty alcohols is estimated at 5 billion USD (Grand View Research, Inc. 2016). The major fraction of the fatty alcohols used today are derived from either crude oil or palm kernel oil, both feedstocks being non-sustainable on a long term (Shah et al., 2016). Alternatively, fatty alcohols can be produced from abundant renewable feedstocks *via* microbial fermentation.

Some marine bacteria can produce fatty alcohols naturally; however, the titers are low, and these organisms are not suitable for large-scale fermentation. Hence, several industrially applicable hosts have been engineered to produce fatty alcohols. *Escherichia coli* was engineered to produce 6.3 g/L total fatty alcohols (Liu et al., 2016) (bioreactor, complex media). Baker's yeast *Saccharomyces cerevisiae* was engineered to achieve a titer of 6.0 g/L (d'Espaux et al., 2017) (bioreactor, complex media). Several oleaginous yeast species have also applied for fatty alcohol production. *Yarrowia lipolytica* has been engineered to produce 2.2 g/L (Xu et al., 2016) (bioreactor, minimal media), and *Lipomyces starkeyi* was engineered to produce 1.7 g/L (McNeil and Stuart 2018) (shake flask, minimal media). To this date, the highest reported titer of 8 g/L total alcohols was obtained with oleaginous yeast *Rhodospiridium toruloides* (Fillet et al., 2015) (bioreactor, complex media).

Fatty alcohol biosynthesis is carried out in two enzymatic steps from fatty acyl-CoAs, key intermediates in membrane, and storage lipid biosynthesis. The two enzymatic steps can be carried out by two enzymes, an aldehyde-forming long-chain acyl-CoA reductase (ACR, EC 1.2.1.50), and an aldehyde reductase (AHR, EC 1.1.1.2), where ACR converts the fatty acyl-CoA into a fatty aldehyde, which in turn is converted by AHR into fatty alcohol. The two enzymatic steps can also be carried out by a single enzyme, an alcohol-forming fatty acyl-CoA reductase (FAR, EC 1.2.1.84), where FAR converts fatty acyl-CoA into fatty alcohol, with a fatty aldehyde as a transient intermediate. The conversion of fatty acyl-CoA into corresponding fatty alcohol requires two NADPH molecules (Figure 1).

The process performance parameters reported for fatty alcohols in the literature do not yet meet the requirements for industrial production of bulk chemicals. Further strain improvement and process optimization are required. A better understanding of the effect of fatty alcohol production on a cell factory can guide further strain improvement. In this study, we performed a multi-omics analysis comparing gene expression, fluxes, and intracellular metabolites' concentrations in fatty alcohol-producing *S. cerevisiae* and *Y. lipolytica* strains.

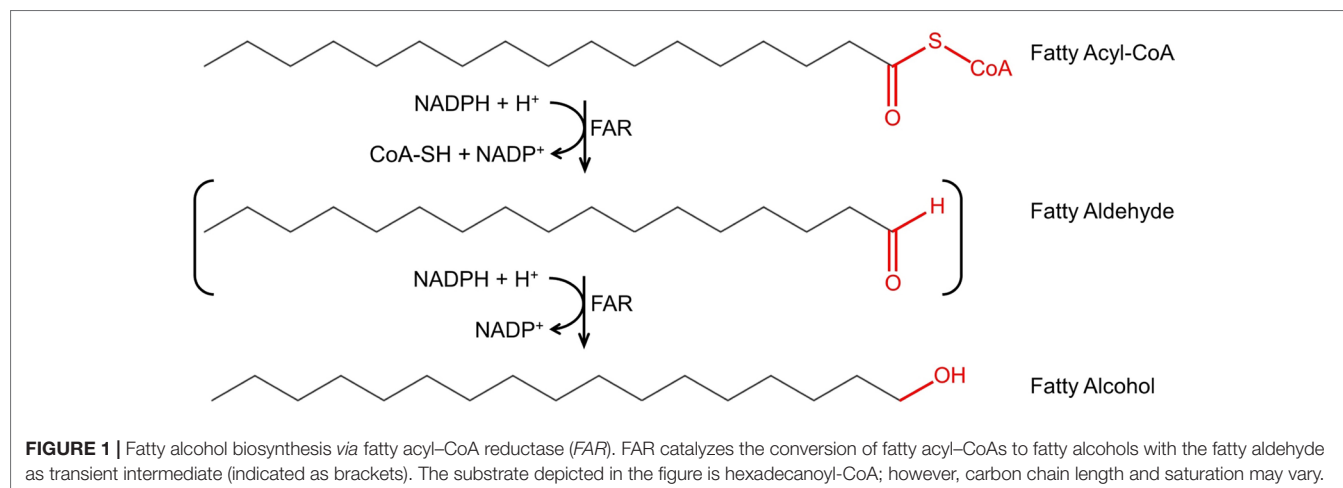
## MATERIALS AND METHODS

### Strains, Reagents, and Chemicals

*Escherichia coli* DH5 $\alpha$  was used for manipulation of DNA during cloning. *S. cerevisiae* CEN.PK113-7D (MATa URA3 HIS3 LEU2 TRP1 MAL2-8<sup>+</sup> SUC2) was a gift from Dr. Peter Kötter (Goethe-Universität, Germany). *Y. lipolytica* GB20 (MATb, ku70 $\Delta$ , nugm-Htg2, ndh2i, lys11<sup>-</sup>, leu2<sup>-</sup>, ura3<sup>-</sup>) was a gift from Prof. Volker Zickermann (Goethe-Universität, Germany). Cloning reagents were sourced according to the EasyClone-Markerfree (Jessop-Fabre et al., 2016) and EasyCloneYALI (Holkenbrink et al., 2018). All chemicals were acquired from Sigma-Aldrich unless otherwise specified. Nourseothricin was acquired from Jena Bioscience GmbH (Germany).

### Strain Construction

The *S. cerevisiae* strains were constructed using the EasyClone-MarkerFree toolbox (Jessop-Fabre et al., 2016). The fatty alcohol degradation-deficient strain (ST6849) was constructed from CEN.PK113-7D by introducing a Cas9 plasmid pCfB2312, followed by a knock out of the *PEX10* and *HFD1* genes using the single guide RNA (sgRNA) plasmid and repair templates (synthesized DNA fragment) as described in **Supplementary Table S1**. The fatty alcohol-producing strain (ST6989) was constructed by expressing a total of four copies of the fatty acyl-CoA reductase from *Marinobacter algicola* (*malFAR*), codon-optimized for *Y. lipolytica* (sequence detailed in **Supplementary**





**File S1**), in the degradation-deficient strain. Integration vectors containing two copies of *malFAR* were constructed, consisting of one copy of *malFAR* under the *TDH3* (*GPD*) promoter and one copy under the *TEF1* promoter, which were inserted into vectors pCfB3034 and pCfB3037 to generate vectors pCfB7082 and pCfB7083, respectively (**Supplementary Table S1**). The NotI-digested plasmids were inserted into integration sites X-3 and XI-5, using the double sgRNA plasmid pCfB5283.

The *Y. lipolytica* degradation-deficient strain (ST6770) was constructed from strain ST6276 (Borodina et al., 2018). ST6276 is derived from *Y. lipolytica* strain GB20 by deleting genes *FAO1*, *HFD1*, *HFD4*, and *PEX10*. The leucine and uracil auxotrophies were closed by integrating an expression cassette containing *URA3* and *LEU2* using vector pCfB7093 and selecting for growth on SC-Ura-Leu agar plates. The lysine auxotrophy was closed by restoring the native sequence of the homocitrate synthase gene (*YALI1\_F38776g*) using DNA fragment BB2251 and selecting for growth on SC-Lys agar plates. The fatty alcohol-producing strain (ST6987) was constructed by expressing a total of four copies of the *malFAR*, codon-optimized for *Y. lipolytica*, in the degradation-deficient strain ST6770. A two-gene construct, consisting of one copy of *malFAR* under the *GPD* promoter and one copy under the *TEF*-intron promoter, were inserted into vector pCfB4796 and pCfB4784 to generate plasmid pCfB7091 and pCfB7092, respectively (**Supplementary Table S1**). The NotI-digested plasmids were inserted at both integration sites D-1 and F-2.

The promoter and coding sequences of all vectors were verified using Sanger sequencing provided by Eurofins Genomics, and integration was verified by colony PCR.

## Cultivation

Unless otherwise specified, all cultures were grown in baffled shake flasks equipped with caps containing air-permeable membranes, incubated at 30°C at 250 rotations per minute (RPM) in a MaxQ 8000 Orbital Shaker (Thermo Fisher). Pre-cultures were grown on minimal media until the late exponential phase, and subsequently washed and resuspended in either fresh minimal media or nitrogen-depleted media for growth phase or nitrogen-depleted stationary phase studies respectively. For the pre-cultures, the media contained 20 g/L glucose, 5 g/L ammonium sulfate, 12 g/L potassium phosphate (pH 6.0), 3.4 g/L yeast nitrogen base (YNB, w/o amino acids, w/o ammonium sulfate), and 1% YPD. Pre-cultures were centrifuged at 5,000 × g for 5 min at room temperature and washed with an equal volume of sterile water; cells were centrifuged again at 5,000 × g for 5 min at room temperature and resuspended to OD<sub>600</sub> 10 in nitrogen-depleted media. For nitrogen-depleted stationary phase cultures, the resuspended cells were used as the starting point. For growth phase cultures, the resuspended cells were inoculated into growth phase media at OD<sub>600</sub> 0.03. For the growth phase cultivations, mineral salt medium was used, containing 20 g/L glucose, 5 g/L ammonium sulfate, 12 g/L potassium phosphate (pH 6.0), and 3.4 g/L YNB (w/o amino acids, w/o ammonium sulfate). For the nitrogen-depleted stationary phase cultivations, the media contained 20 g/L glucose, 12 g/L potassium phosphate

(pH 6.0), and 3.4 g/L YNB (w/o amino acids, w/o ammonium sulfate). For the fatty alcohol degradation test, cells were cultivated in triplicates in 12-ml round bottom glass vials filled with 2-ml media at 30°C at 250 RPM in a MaxQ 8000 Orbital Shaker (Thermo Fisher). The cultivation medium contained 500 mg/L hexadecanol, 500 mg/L octadecanol, 20 g/L ethanol, 2 g/L glucose, 1.7 g/L YNB w/o amino acids and ammonium sulfate, 12 g/L potassium phosphate (pH 6.0), and 5 g/L ammonium sulfate. Biological triplicates were done for each strain and condition.

## HPLC Analysis

Five hundred-microliter samples were taken at each sampling point, centrifuged in Eppendorf tubes at 5,000 × g at 4°C for 5 min, and supernatants were stored at −20°C until analysis. Supernatants were analyzed for the presence of ethanol, glucose, glycerol, and acetate using HPLC UltiMate 3000 (Thermo Fisher) with Aminex HPX87H ion exclusion column. Samples were run for 30 min at 0.600 ml/min at 60°C using 5 mM H<sub>2</sub>SO<sub>4</sub> as eluent. Compounds were detected using a Dionex RI101 and DAD-3000 detectors (Dionex) for RI and UV detection, respectively.

## Liquid Chromatography-Mass Spectrometry (LC-MS) Analysis

For *Y. lipolytica* cultures, supernatant samples from section *HPLC Analysis* were also analyzed for the presence of tricarboxylic acid (TCA) cycle-derived organic acids (malic acid, succinic acid, citric acid, isocitric acid, pyruvic acid, α-ketoglutaric acid, and fumaric acid). LC-MS data were collected on the EVOQ Elite Triple Quadrupole Mass Spectrometer system coupled with an Advance UHPLC pump (Bruker, Fremont, CA). Samples were held in the CTC HTS PAL autosampler at a temperature of 5.0°C during the analysis. 1-μl injections of the sample were made onto a Waters ACQUITY HSS T3 C18 UHPLC column, with a 1.8-μm particle size, 2.1 mm i.d., and 100 mm long. The column was maintained at 30.0°C. The solvent system used was solvent A (MilliQ water with 0.1% formic acid) and solvent B (acetonitrile with 0.1% formic acid). The flow rate was 0.400 ml/min with an initial solvent composition of %A = 100 and %B = 0 held until 0.50 min; the solvent composition was then changed following a linear gradient until it reached %A = 5.0 and %B = 95.0 at 1.00 min. This was held until 1.79 min when the solvent was returned to the initial conditions, and the column was re-equilibrated until 4.00 min. The column eluent flowed directly into the heated ESI probe of the MS, which was held at 250°C and a voltage of 2,500 V. MRM data was collected in negative ion mode; the target masses are shown in **Supplementary Table S2**. The other MS settings were as follows: sheath gas flow rate of 50 units, nebulizer gas flow rate of 50 units, cone gas flow rate of 20 units, cone temp was 350°C, and collision gas pressure 1 mTorr.

## Cell Dry Weight

The cell dry weight (CDW) was determined by pre-weighing dried 0.45-μm cellulose nitrate membrane filters (VWR), filtering 10-ml culture across the membrane, washing with 10 ml water, drying at 60°C, and weighing the dried filter with

biomass. The OD/CDW ratio was determined to be constant for all the strains throughout all the cultivations, and for the cultures throughout this study, the biomass was estimated by measuring the OD<sub>600</sub> using a NanoPhotometer Pearl (Implen) and calculating the biomass dry weight using conversion factors of 0.12 (g/L)/OD or 0.14 (g/L)/OD for *S. cerevisiae* and *Y. lipolytica*, respectively.

### Fatty Alcohol Extraction

Fatty alcohols were analyzed by collecting 1 ml of culture broth into a 4-ml glass vial and adding 10-μl internal standard (2 g/L of methyl cis-10-heptadecanoate dissolved in 100% ethanol). Samples were vortexed for 3 s and frozen at 80°C until further processing. To perform the extraction, samples were freeze-dried for 2–3 days at –30°C under vacuum in a freeze-drying system (Labconco). 1 ml 2:1 chloroform:methanol mixture was added to each freeze-dried sample to disrupt the cells, vortexed for 15 min in a DVX-2500 multi-tube vortexer (VWR), and left at room temperature for 4 h. The solvents were subsequently evaporated under a nitrogen stream. 1 ml hexane was added to the sample vials, vortexed for 15 min in a DVX-2500 multi-tube vortexer (VWR), and incubated at room temperature overnight. Samples were transferred to new vials and stored at –20°C until analyzed.

The analysis was carried out on a GC-MS using an INNOWax column (30 m × 0.25 mm × 0.25 μm) with helium as carrier gas. The injector was set to splitless mode at 220°C; the oven temperature was set to 80°C for 1 min, increased at a rate of 10°C/min to 210°C, followed by a hold at 210°C for 15 min, increased at a rate of 10°C/min to 230°C followed by a hold at 230°C for 20 min. The GC-MS was operated in electron impact mode (70 eV), scanning at the range 30–400 m/z. Compounds were quantified relative to the internal standard (methyl cis-10-heptadecenoate).

### Fatty Alcohol Degradation

Fatty alcohol degradation was estimated by cultivating cells in media containing fatty alcohols. Strains used were CEN.PK113-7D (*S. cerevisiae* reference), ST6849 (*S. cerevisiae* *pex10Δ*, *hfd1Δ*), W29 (*Y. lipolytica* reference), and ST6770 (*Y. lipolytica* *pex10Δ*, *fao1Δ*, *hfd1Δ*, *hfd4Δ*) (Table 1). The negative controls did not contain any cells. Cultures were inoculated to a starting OD<sub>600</sub> of 1.0 and incubated for 96 h. The whole cultivation tube containing 2 ml culture was taken as sample and processed for fatty alcohol quantification according to the previous description

(Cell Dry Weight), but with double volumes to account for the increased sample size. The detection of slightly lower levels of hexadecanol and slightly higher levels of octadecanol than expected is likely due to an experimental error, which is rather high when working with hydrophobic substances in small volumes. Upon addition of fatty alcohols to the media, they form a sticky floating white precipitate and, while we tried to recover the whole remaining fatty alcohol by adding organic solvent directly to the tube, some of it may still remain on the walls of the tube.

### RNA Sequencing and Gene Expression Analysis

Growth phase cultures were sampled at OD 2–3 (~6 generations), when the cultures were in the mid-exponential growth phase and the unlabeled biomass from inoculum was diluted to less than 3%. Nitrogen-depleted cultures were sampled when ca. 50% of glucose was consumed, which was after 12 h for *S. cerevisiae* and after 48 h for *Y. lipolytica*. The sample volume corresponding to 5 × 10<sup>7</sup>–10<sup>8</sup> cells was added into 50-ml Falcon tube, filled with ice. The tubes were centrifuged at 4°C for 1 min, the liquid was discarded, and the pellet was snap frozen in liquid nitrogen and stored at –80°C until further processing. The cell lysis was carried out in 2 ml-screw cap tubes with 600-μl RLT buffer and 500-μl glass beads using a Precellys 24 at 6000 RPM for 4 × 25 s, with 60 s on ice in between. RNA was subsequently extracted using the RNeasy kit (Qiagen), according to the manufacturer's instructions.

Library preparation was carried out using the TruSeq Stranded mRNA Library Prep Kit (Illumina), and the TruSeq RNA CD indexes (Illumina). Sequencing was carried out using a NextSeq 500 system (Illumina), with NextSeq Mid and High Output v2 Kits (150 cycles), as 75 bp paired-end reads. Index (i7 and i5) reads: 8 bp, flow cell loading: 1.08 pM, sequencing chemistry: 2-channel sequencing-by-synthesis (SBS) technology. PhiX was added at 2.5%. Sequencing facility: NGS lab at the Novo Nordisk Foundation Center for Biosustainability.

The RNA-seq data was processed using KBase (Arkin et al., 2018), and unless specified, default settings were used. Reads were trimmed using Trimmomatic v0.36 (post-tail crop length: 73, head crop length: 14) (Bolger et al., 2014). Read quality was assessed by FastQC. Reads were merged using Multiple ReadsLibs to One ReadsLib v1.0.1. Reads were aligned to the

TABLE 1 | Strains used in this study.

Name	ID	Genotype	Reference
<i>S. cerevisiae</i> reference	CEN.PK113-7D	<b>MATa</b>	Entian and Kötter, 2007
<i>S. cerevisiae</i> non-producer	ST6849	<i>pex10Δ</i> , <i>hfd1Δ</i> , MATa	This study
<i>S. cerevisiae</i> fatty alcohol producer	ST6989	<i>4xMalFAR</i> , <i>pex10Δ</i> , <i>hfd1Δ</i> , MATa	This study
<i>Y. lipolytica</i> reference	W29	MATa	Gaillardin et al., 1973
<i>Y. lipolytica</i> non-producer	ST6770	<i>pex10Δ</i> , <i>fao1Δ</i> , <i>hfd1Δ</i> , <i>hfd4Δ</i> , <i>LEU2</i> , <i>LYS11</i> , <i>URA3</i> , MATb, <i>ku70Δ</i> , <i>nugm-Htg2</i> , <i>ndh2i</i> , <i>lys11</i> <sup>–</sup> , <i>leu2</i> <sup>–</sup> , <i>ura3</i> <sup>–</sup>	This study
<i>Y. lipolytica</i> fatty alcohol producer	ST6987	<i>4xMalFAR</i> , <i>pex10Δ</i> , <i>fao1Δ</i> , <i>hfd1Δ</i> , <i>hfd4Δ</i> , <i>LEU2</i> , <i>LYS11</i> , <i>URA3</i> , MATb, <i>ku70Δ</i> , <i>nugm-Htg2</i> , <i>ndh2i</i> , <i>lys11</i> <sup>–</sup> , <i>leu2</i> <sup>–</sup> , <i>ura3</i> <sup>–</sup>	This study

reference genome using HISAT2 v2.1.0 (Kim et al., 2015). The reference genomes used were modified versions of S288C and W29 (Clib89) for *S. cerevisiae* and *Y. lipolytica*, respectively; the modification consisted of the addition of the expressed *malFAR* genes. Alignment quality was assessed using Qualimap2 v2.2.1 (Okonechnikov et al., 2016). Alignments were assembled using StringTie v1.3.3b (not allowing for novel transcripts) (Pertea et al., 2015). The analysis was carried out separately using EdgeR v3.24.3 (Robinson et al., 2010) in R v3.5.1, using TMM normalization as well as false discovery rate (FDR) correction using the Benjamini–Hochberg method. EdgeR was used to calculate the effect of either nitrogen-depletion or fatty alcohol production in both species. Contrasts were set as “(A+B)/2 – (C+D)/2” where A and B belonged to the same condition (e.g., growth phase), and C and D belonged to the same condition (e.g., nitrogen depletion). An exception was made for the effect of fatty alcohol production in *Y. lipolytica*, where only genes differentially expressed between strains in nitrogen-depleted conditions were considered, due to a lack of difference between the strains during growth phase. Differentially expressed genes predicted by EdgeR were filtered for a *p*-value or 0.01, log2 CPM of 1, and log2-fold change of 2.

GO term enrichment analysis was carried out by the PANTHER v14.0–based online tool (Ashburner et al., 2000; Thomas et al., 2003; Mi et al., 2017; The Gene Ontology Consortium, 2017) available at [www.geneontology.org](http://www.geneontology.org). Enrichment analysis was carried out for biological process GO terms using Fisher’s exact test and Bonferroni correction. Results were filtered for a *p*-value less than 0.05 and fold enrichment greater than 2.5.

RNA-seq data has been uploaded to European Nucleotide Archive (PRJEB32352). A list of differentially expressed genes is attached as **Supplementary File S3**.

### <sup>13</sup>C-Metabolic Flux Analysis

Cells were cultivated as previously described (*Cultivation*), with the exception that labeled glucose was used. One replicate was made with 20% U-<sup>13</sup>C glucose (99% purity, Euriso-Top GmbH, Saarbrücken, Germany) and 80% non-labeled glucose. Two replicates were made with 20% U-<sup>13</sup>C glucose and 80% 1-<sup>13</sup>C glucose (99% purity, Euriso-Top GmbH, Saarbrücken, Germany). The two isotopomers (1-<sup>13</sup>C glucose and U-<sup>13</sup>C glucose) are commonly used in <sup>13</sup>C metabolic flux analysis as they provide a good flux resolution at a reasonable cost. Furthermore, the single replicate of 20% U<sup>13</sup>C glucose can be used to estimate the quality of the MS data. The labeling strategy was informed by previous publication (Zamboni et al., 2009). A cell amount of 0.3 mg CDW were harvested at OD 2 (~6 generations) and washed with 1 ml cold NaCl (0.9%). The pellet was stored at –80°C until processing. The samples were resuspended in 150 µl HCl (6 M), transferred to a glass vial, and incubated at 105°C for 6 h to hydrolyze the cell pellet as previously described (Schmitz et al., 2017). Samples were dried until only a dark brown residue remained by heating the open vial at 80°C under a fume hood. The residue was resuspended in 30 µl acetonitrile. Samples were derivatized by the addition of MBDSTFA at a 1:1 ratio and incubated at 85°C for 1 h. Samples were analyzed by GC-MS according to a previously

described protocol (Kildegaard et al., 2016). Raw GC-MS data was corrected using iMS2Flux v7.2.1 (Poskar et al., 2012). Fluxes were calculated using parameter continuation in the INCA v1.7 software (Young, 2014). The model used was adapted from previous publication (Wasylenko and Stephanolous, 2015), see **Supplementary File S4** for the final model used, as well as for the full set of calculated fluxes.

### Intracellular Metabolome Analysis

Samples corresponding to approximately 0.25 mg CDW were taken for each replicate. For growth phase samples, samples were taken at mid-growth phase (OD 1–2, ~6 generations). For the nitrogen-depletion samples, samples were taken at 12 or 48 h, respectively, for *S. cerevisiae* and *Y. lipolytica* strains. The sampling method used was adapted from a previous publication (McCloskey et al., 2015b), with the addition of an extra extraction step, in which the filter with the quenched biomass was transferred into a 50-ml Falcon tube containing 5 ml boiling ethanol, together with the 1 ml extraction solvent used for quenching (containing internal standard). The filter was incubated at 80°C in the boiling ethanol for 90 s, the tube was quickly vortexed, the filter was flipped, and the tube was incubated for another 90 s. The solution was aliquoted into 2-ml Eppendorf tubes and processed according to the previously described protocol. See **Supplementary File S5** for full details. The analysis was conducted according to previous publication (McCloskey et al., 2015a).

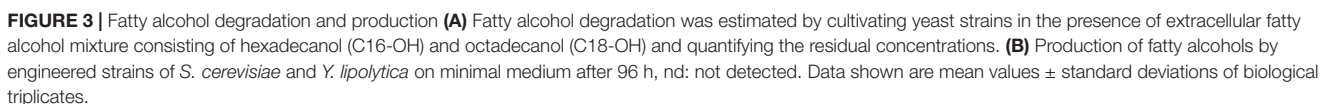
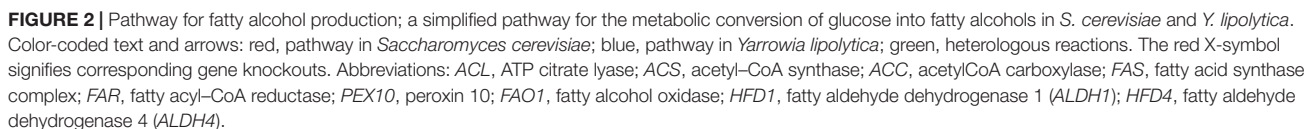
## RESULTS

### Establishing Fatty Alcohol Production in *S. cerevisiae* and *Y. lipolytica*

The first step towards creating fatty alcohol–producing yeast strains was to reduce the degradation of fatty alcohols as described in Borodina et al. (2018). In *S. cerevisiae*, we chose to delete the genes encoding peroxisomal biogenesis factor Pex10p and aldehyde dehydrogenase Hfd1p (**Figure 2**). The deletion of *PEX10* prevents the formation of peroxisomes, where β-oxidation of fatty acids occurs. The *HFD1* gene was shown in a previous study (Buijs et al., 2015) to be responsible for the degradation of fatty alcohols in *S. cerevisiae*. In *Y. lipolytica*, we deleted *PEX10* and two out of four aldehyde dehydrogenase-coding genes *HFD1* and *HFD4*, the ones that were previously reported to have the highest activity (Iwama et al., 2014). Additionally, we deleted the *FAO1* gene encoding a fatty alcohol oxidase in *Y. lipolytica* (**Figure 2**).

In order to investigate if the chosen gene knockouts reduced the degradation of fatty alcohols, we cultivated the non-engineered and engineered strains in the medium supplemented with approximately 0.5 g/L each of hexadecanol and octadecanol for 96 h and analyzed the remaining fatty alcohol concentration (**Figure 3A**). There was no significant difference between the final concentrations of fatty alcohols between the cultures of *S. cerevisiae* strains and the control experiment without cell addition. As it has been shown previously that *S. cerevisiae* can





In the next step, we integrated four copies of the *malFAR* fatty acyl-CoA reductase (*FAR*) gene into the yeast strains with reduced fatty alcohol degradation. The *FAR* genes were expressed from strong constitutive promoters, *TEF1* and *TDH3* (*GPD*) for *S. cerevisiae* and from *TEFintron* (Tai and Stephanopoulos, 2013) and *GPD* promoters for *Y. lipolytica*. The strong constitutive promoters were selected to ensure that fatty alcohol biosynthetic

Throughout this study, non-producing strains of *S. cerevisiae* (*pex10Δ*, *hfd1Δ*) and *Y. lipolytica* (*pex10Δ*, *fao1Δ*, *hfd1Δ*, *hfd4Δ*) were compared with fatty alcohol-producing strains of *S. cerevisiae* (4x *malFAR*, *pex10Δ*, *hfd1Δ*) and *Y. lipolytica* (4x *malFAR*, *pex10Δ*, *fao1Δ*, *hfd1Δ*, *hfd4Δ*), during both the



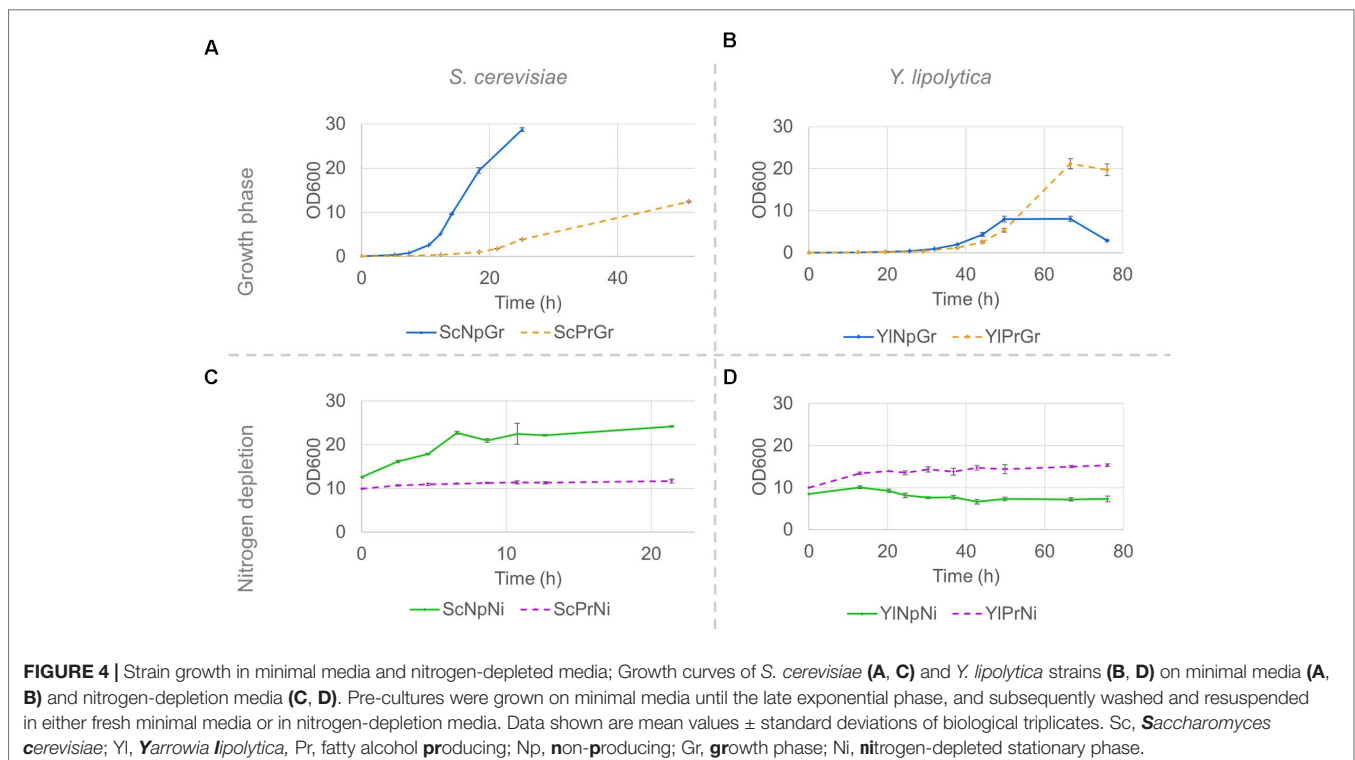
exponential growth phase as well as during a nitrogen-depleted stationary phase (Table 1). Pre-cultures were grown on minimal media until the late exponential phase, and subsequently washed and resuspended in fresh minimal media for growth or in nitrogen-depletion media for stationary phase studies. This experimental set-up allowed for simplified parallel investigation of each phase individually, which could be made to occur in sequence as a result of nitrogen consumption in an industrial setting. The nitrogen-depleted stationary phase is of interest since it has previously been shown to increase the flux *via* fatty acyl-CoA to triacylglycerides and lipid accumulation (Pomraning et al., 2016).

The fatty alcohol-producing strain of *S. cerevisiae* had a 50% lower maximum specific growth rate,  $\mu_{\max}$ , and reached a lower final OD, ~40% of the final OD reached by the parental strain not expressing *FAR* genes (Figure 4A and Table 2). Although no time series data is available beyond the 50 h, we made some separate experiments under the same conditions, and no further increase in OD was observed between 50 and 96 h for the producing strain. In contrast, the  $\mu_{\max}$  of *Y. lipolytica* was not affected by expression of fatty alcohol reductase genes, and the strain with fatty alcohol production grew to a 2.6-fold higher final OD (Figure 4B and Table 2). When the strains were transferred to a medium without nitrogen and cultivated for 21 h, the OD of the non-producing *S. cerevisiae* strain increased 1.9-fold (Figure 4C). The OD of the non-producing *Y. lipolytica* strain decreased slightly, indicating some negative effect of the introduced gene deletions (Figure 4D).

Glucose uptake rates varied greatly between the hosts and conditions. *S. cerevisiae* exhibited a ~13-fold and ~5-fold higher glucose uptake rate in growth phase and nitrogen-depleted stationary phase, respectively, compared to *Y. lipolytica*. Furthermore, the

uptake rate was ~17-fold higher and ~6-fold higher in the growth phase than in the nitrogen-depletion phase in *S. cerevisiae* and *Y. lipolytica*, respectively (Table 2).

As for by-products, *S. cerevisiae* primarily secreted ethanol, as well as some acetate and glycerol. There was no major change in the by-product secretion of the producing and non-producing strains of *S. cerevisiae*. *Y. lipolytica* secreted several TCA cycle-associated organic acids, primarily pyruvate,  $\alpha$ -ketoglutarate, citrate, and malate. The fatty alcohol-producing strain of *Y. lipolytica* exhibited a 2–3-fold lower by-product secretion rates for all the measured metabolites, during both exponential growth and nitrogen-depleted stationary phase, plausibly due to a redirection of the flux toward fatty alcohols. The relative secretion rate (relative to glucose uptake rate) of isocitrate in the producing *Y. lipolytica* strain was reduced by ~20-fold in the nitrogen-depleted condition. The reduced isocitrate secretion rate could potentially be explained by an increased transport of mitochondrial citrate to the cytoplasm, followed by conversion of citrate to acetyl-CoA (catalyzed by ATP-citrate lyase), which in turn would be used for fatty alcohol production. In response to nitrogen-depletion, *Y. lipolytica* exhibited a 5–25-fold relative increase in secretion rate of all by-products analyzed, except for citrate, which was secreted with a ~200-fold increased relative rate. During nitrogen-depletion, *Y. lipolytica* secreted ~50% and ~25% of the total carbon consumed, in the form of organic acid by-products for the non-producing and producing strain, respectively. Additionally, the fatty alcohol-producing strain produced fatty alcohols corresponding to ~8% of the carbon consumed. Both *S. cerevisiae* and *Y. lipolytica* showed detectable fatty alcohol production only in nitrogen-depleted conditions (Table 2).



**TABLE 2 |** Uptake-, secretion-, and growth rate. Sc, *Saccharomyces cerevisiae*; Yl, *Yarrowia lipolytica*, Pr, fatty alcohol producing; Np, non-producing; Gr, growth phase; Ni, nitrogen-depleted stationary phase; nd, not determined. Data shown are mean values  $\pm$  standard deviations of biological triplicates.

	Glucose	Ethanol	Acetate	Glycerol	Succinate	Malate	Isocitrate
	mmol g <sup>-1</sup> h <sup>-1</sup>	mmol g <sup>-1</sup> h <sup>-1</sup>	mmol g <sup>-1</sup> h <sup>-1</sup>	mmol g <sup>-1</sup> h <sup>-1</sup>	$\mu$ mol g <sup>-1</sup> h <sup>-1</sup>	$\mu$ mol g <sup>-1</sup> h <sup>-1</sup>	$\mu$ mol g <sup>-1</sup> h <sup>-1</sup>
ScNpGr	15.7 $\pm$ 0.29	22. $\pm$ 0.3	1.41 $\pm$ 0.1	2.81 $\pm$ 0.0	nd	nd	nd
ScPrGr	15.4 $\pm$ 0.26	20.6 $\pm$ 0.4	2.3 $\pm$ 0.3	3.89 $\pm$ 0.3	nd	nd	nd
ScNpNi	1.17 $\pm$ 0.12	0.9 $\pm$ 0.5	0.2 $\pm$ 0.0	0.0 $\pm$ 0.0	nd	nd	nd
ScPrNi	1.61 $\pm$ 0.10	2.4 $\pm$ 0.9	0.1 $\pm$ 0.0	0.3 $\pm$ 0.0	nd	nd	nd
YlNpGr	1.52 $\pm$ 0.19	nd	nd	nd	1.2 $\pm$ 0.4	8.3 $\pm$ 1.9	1.1 $\pm$ 0.1
YlPrGr	2.04 $\pm$ 0.20	nd	nd	nd	0.8 $\pm$ 0.0	5.5 $\pm$ 0.2	0.4 $\pm$ 0.0
YlNpNi	0.35 $\pm$ 0.01	nd	nd	nd	5.0 $\pm$ 0.7	31.3 $\pm$ 3.8	6.6 $\pm$ 1.2
YlPrNi	0.22 $\pm$ 0.03	nd	nd	nd	1.8 $\pm$ 0.3	9.1 $\pm$ 0.6	0.2 $\pm$ 0.5
	Citrate	$\alpha$ -Ketoglutarate	Fumarate	Pyruvate	Fatty alcohols	Growth rate	
	$\mu$ mol g <sup>-1</sup> h <sup>-1</sup>	$\mu$ mol g <sup>-1</sup> h <sup>-1</sup>	$\mu$ mol g <sup>-1</sup> h <sup>-1</sup>	$\mu$ mol g <sup>-1</sup> h <sup>-1</sup>	$\mu$ mol g <sup>-1</sup> h <sup>-1</sup>	h <sup>-1</sup>	
ScNpGr	nd	nd	nd	nd	nd	0.33 $\pm$ 0.00	
ScPrGr	nd	nd	nd	nd	nd	0.17 $\pm$ 0.00	
ScNpNi	nd	nd	nd	nd	nd	nd	
ScPrNi	nd	nd	nd	nd	1.9 $\pm$ 0.7	nd	
YlNpGr	1.1 $\pm$ 0.8	30.2 $\pm$ 3.9	1.7 $\pm$ 0.4	79.7 $\pm$ 8.9	nd	0.13 $\pm$ 0.00	
YlPrGr	0.6 $\pm$ 0.1	11.9 $\pm$ 0.9	1.0 $\pm$ 0.1	33.2 $\pm$ 5.0	nd	0.12 $\pm$ 0.00	
YlNpNi	54.9 $\pm$ 3.2	39.1 $\pm$ 13.4	7.2 $\pm$ 1.4	102.7 $\pm$ 5.4	nd	nd	
YlPrNi	12.6 $\pm$ 1.2	19.9 $\pm$ 2.2	1.3 $\pm$ 0.1	29.9 $\pm$ 2.7	6.0 $\pm$ 0.9	nd	

### <sup>13</sup>C-Metabolic Flux Analysis

In order to get insight into the strains' response to fatty alcohol production on a metabolic level, <sup>13</sup>C-fluxomics and targeted metabolomics analyses were conducted. The flux analysis determines the fluxes (conversion rates) between metabolites. Fluxes reveal the flow of carbon through the cell and the distribution of carbon flux between alternative pathways. For <sup>13</sup>C-flux analysis, the cells were cultivated on a mix of labeled and unlabeled glucose. The incorporation of labeled carbons (<sup>13</sup>C) into proteinogenic amino acids was measured by GC-MS. The <sup>13</sup>C-flux analysis method requires a metabolic steady state and, if based on measurements of proteinogenic amino acids, growing cells. Therefore, the fluxes were only estimated for the exponential growth phase, which can be considered to represent a quasi-steady-state condition.

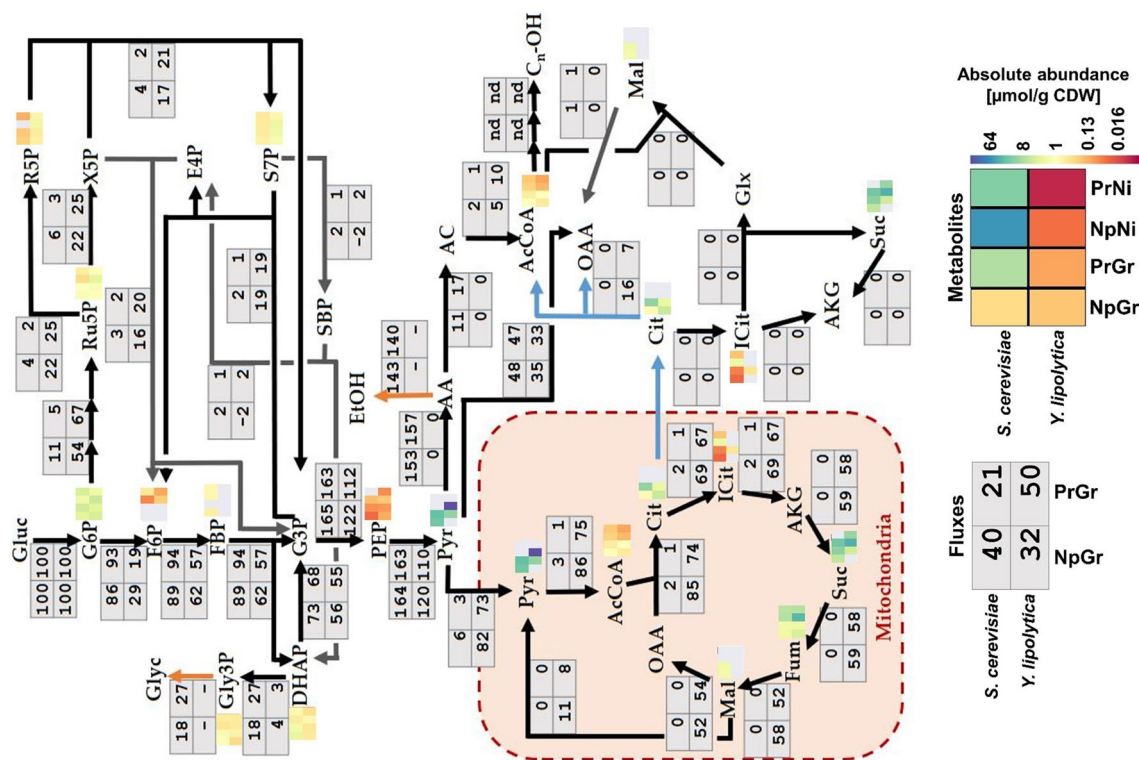
The central carbon flux distributions in *S. cerevisiae* and *Y. lipolytica* were very different (Figure 5). The *S. cerevisiae* strains processed approximately 90% of the carbon from glucose through glycolysis, whereas only around 10% of the carbon went into the pentose phosphate pathway. The cells channeled ~50% of the carbon from glucose to ethanol, primarily relying on the energy generated in the fermentation process. *Y. lipolytica* is a non-fermenting yeast, so the energy is generated by oxidative phosphorylation in the mitochondria. Besides the higher TCA cycle flux associated with this respiratory lifestyle, *Y. lipolytica* differed from *S. cerevisiae* by diverting nearly half of the internalized glucose into the pentose phosphate pathway generating substantial amounts of NADPH. NADPH is a key redox co-factor for fatty acid and fatty alcohol biosynthesis. Additionally, *Y. lipolytica* had a ~5-fold higher relative flux (relative to glucose uptake) towards cytosolic acetyl-CoA than *S. cerevisiae*. Acetyl-CoA is the precursor for fatty acyl-CoA, which in turn is the precursor for fatty alcohols, and other fatty acid-derived compounds. It is worth noting that *S. cerevisiae* and

*Y. lipolytica* utilize different pathways for the synthesis of acetyl-CoA from glucose (Figure 2). In *S. cerevisiae*, most of cytosolic acetyl-CoA is produced by the action of pyruvate decarboxylase, aldehyde dehydrogenase, and acetyl-CoA synthase. The NADP-dependent *ALD6* has previously been shown to contribute ~40% of the NADPH generated in *S. cerevisiae*, with the remaining 60% being generated via the pentose phosphate pathway (Blank et al., 2005). In contrast, *Y. lipolytica* uses ATP-citrate lyase (*ACL*) to make acetyl-CoA from citrate, which is exported from mitochondria with a simultaneous import of malate. The by-product of the *ACL* reaction is oxaloacetate, which is converted into malate. In *Y. lipolytica*, the NADPH is primarily generated through the pentose phosphate pathway (Wasylenko et al. 2015).

### Metabolomics

Metabolomics analysis determines the intracellular metabolite concentrations. Metabolite pools can help to reveal limitations in precursor availability, as well as potentially limiting steps in the production pathway. Samples for metabolome analysis were taken during growth or nitrogen-depletion in producing and non-producing strains of both *S. cerevisiae* and *Y. lipolytica* and were rapidly filtered and quenched.

The metabolomics data revealed a decreased abundance of acetyl-CoA during nitrogen-depletion in both species, ~3- and ~6.5-fold lower in *S. cerevisiae* and *Y. lipolytica*, respectively (Figure 5). This might indicate that fatty alcohol-producing strains encounter a limited precursor supply under these conditions. *Y. lipolytica* had increased levels of intermediates of the pentose phosphate pathway, which is consistent with the high flux through this pathway. Ribose 5-phosphate (r5p), ribulose 5-phosphate (ru5p), and sedoheptulose 7-phosphate (s7p) had a ~3-fold higher abundance in *Y. lipolytica* compared to *S. cerevisiae* strain. Given that this difference was observed with



**FIGURE 5 |** Fluxes and metabolite levels in central carbon metabolism. Fluxes (gray boxes) are relative values, defined as the percentage of glucose uptake rate. Negative values indicate that the reaction is taking place in the opposite direction to what is indicated by the arrow. Metabolites (heatmaps) correspond to absolute abundances in  $\mu\text{mol/g CDW}$ ; gray means data missing (below the detection limit, saturated signal, or too high background in control sample). Sc: *Saccharomyces cerevisiae*, Yl: *Yarrowia lipolytica*, Pr: fatty alcohol producing, Np: non-producing, Gr: growth phase, Ni: nitrogen-depleted stationary phase.

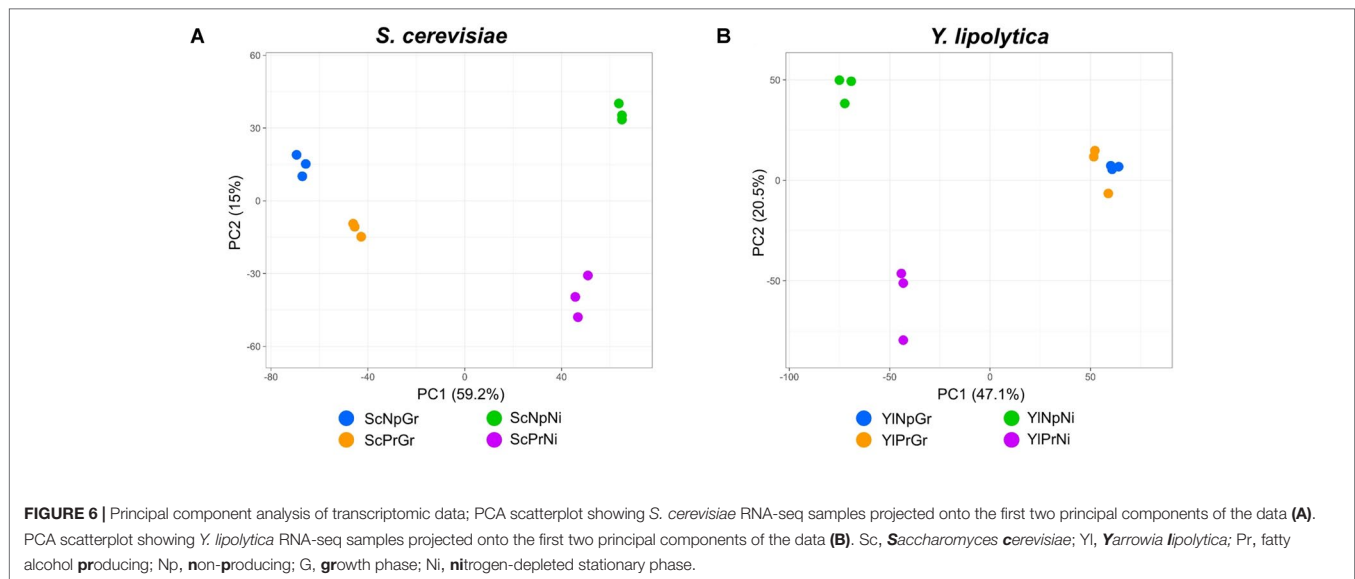
both producing and non-producing strains, it appears to be an inherent feature of *Y. lipolytica* metabolism.

## Transcriptomics

The gene expression profiles of fatty alcohol producing and non-producing strains were analyzed during growth and nitrogen-depletion *via* RNA sequencing. Principal component analysis (PCA) plots of the data revealed four separate sample groups in *S. cerevisiae* and three separate groups in *Y. lipolytica* (Figure 6). In *S. cerevisiae*, these groups represented the four different conditions (fatty alcohol-producing strain in growth and nitrogen-depletion, and non-producing strain in growth and nitrogen-depletion), indicating clear differences between all analyzed strains/conditions. In contrast, for *Y. lipolytica*, the three groups revealed that the producing and non-producing strains in the nitrogen-depletion phase were clearly different from each other and from the strains in the growth phase. But producing and non-producing strains in the growth phase were similar. Furthermore, direct comparison of the differentially expressed genes between the two conditions revealed only nine functionally annotated genes, out of which only four were also differentially expressed between the same two strains in nitrogen-depleted conditions. Considering this similarity, only *Y. lipolytica* strains subjected to the nitrogen-depleted conditions were used

to identify the differentially expressed genes in response to fatty alcohol production in *Y. lipolytica*. In both species, PC1 separates growth phase from nitrogen-depleted stationary phase and explains 59% and 47% of the difference in *S. cerevisiae* and *Y. lipolytica*, respectively.

In response to nitrogen-depletion (nitrogen-depleted stationary phase vs. growth phase), *S. cerevisiae* differentially expressed 716 genes, 401 of which were upregulated in nitrogen-depleted stationary phase, and 315 were downregulated (Table 3). Out of the 331 characterized upregulated genes, the most enriched GO terms were those related to the TCA cycle, carbohydrate metabolism, as well as various metabolic processes. These points toward a shift in the metabolic profile, switching from the rapid but wasteful Crabtree overflow metabolism to a more energy conservative strategy utilizing the mitochondria, and storing excess carbon as glycogen and trehalose. Furthermore, the GO categories “response to oxidative stress” and “response to toxic substance” were also enriched. The 286 characterized downregulated genes were enriched for ribosome-/translation-related processes as well as RNA metabolic processes. The enrichment analysis of the downregulated genes indicated a slowdown of the central cellular processes associated with adaptation to the stationary phase, i.e., quiescence (Coller, 2011), induced by the nitrogen depletion. The response is similar to what has been seen in previous studies (Boer et al., 2003), and even



though it is not revealed in the enrichment analysis, individual inspection of differentially expressed genes showed that the strains underwent transcriptional changes associated with the release of nitrogen catabolite repression (Daugherty et al., 1993; ter Schure et al., 2000) (Table S3). Nitrogen depletion triggered the upregulation of nitrogen transporters such as *GAP1* (general amino acid permease, YKR039W) and *PUT4* (proline permease, YOR348C), as well as the upregulation of enzymes involved in nitrogen metabolism such as *DUR1,2* (urea amidolyase, YBR208C) and *DAL1* (allantoinase, YIR027C).

The *Y. lipolytica* strains differentially expressed 631 genes in response to nitrogen depletion, out of which 500 were upregulated in nitrogen-depleted stationary phase, and 131 were downregulated. However, of the 500 upregulated genes 355 (71%) were uncharacterized. Of the 145 characterized upregulated genes, GO terms relating to transporters were the only terms significantly enriched. A large part of these belonged to transporters for nitrogen-containing compounds such as ammonium, amino acids, oligopeptides, and urea. The upregulation of these transporters is likely the result of the alleviation of nitrogen catabolite repression imposed in the presence of ammonia, which is a logical biological response as the cell needs to find alternative nitrogen sources. Furthermore, considering the large increase in carboxylic acid secretion rate in response to nitrogen depletion (Table 2), carboxylic acid transporters were of particular interest. However, due to genes being associated with multiple GO terms, all the 11 carboxylic acid transporters predicted to be enriched were also annotated as amino acid transporters. Hence, the enrichment analysis didn't predict any known transporters of the TCA-associated organic acids. Of the 77 characterized down-regulated genes, only a single GO term, "sulfur compound metabolic process," was enriched. This does not appear to be related to sulfur-containing amino acids, but rather to the metabolism of other various sulfur-containing compounds possibly due to the unintended consequence of lowering (but not depleting) the extracellular sulfate concentration when depleting the media of nitrogen, which is added to the media in the form of ammonium sulfate.

In response to fatty alcohol production (producing strain vs. non-producing strain), the *S. cerevisiae* strain differentially expressed 24 genes, 21 of which were upregulated in the producing strain (Table 3). GO term enrichment analysis revealed that 9 out of the 17 characterized upregulated genes were associated with "cell wall organization or biogenesis." A previous meta-study (Arroyo et al., 2009) compared the transcriptional response to three different compounds (zymolase, congo red, and pneumocandins) triggering a cell wall stress response. Lowering the cutoff threshold of the differential gene expression from a 4-fold to a 2-fold increase [same as used by the cell wall stress studies (Lagorce et al., 2003; Boorsma et al., 2004; García et al., 2004; Rodríguez-Peña et al., 2005) in the meta-analysis] revealed a significant overlap. Out of the 18 genes upregulated in all three cell wall stress conditions, 16 were found to be upregulated in response to *FAR* expression, indicating that the *S. cerevisiae* fatty alcohol-producing strain is experiencing cell wall stress. No other GO terms were enriched. Three genes were downregulated, all of which were uncharacterized.

The *Y. lipolytica* strains differentially expressed 215 genes in response to fatty alcohol production, out of which 60 genes were upregulated, and 155 were downregulated in the producing strain. GO term enrichment analysis singled out antibiotic catabolic process (especially formate catabolism) as a key factor among the 39 characterized genes upregulated during fatty alcohol production. Three of the 10 most upregulated characterized genes were formate dehydrogenases (YALI0\_B22506g, YALI0\_B19976g, and YALI0\_E14256g) with relatively high RNA abundance (CPM) and very low p-values (Supplementary Table S3), which further supports that the formate dehydrogenase upregulation is biologically significant. Formate dehydrogenases catalyze the reversible reaction between formate and carbon dioxide (formate +  $\text{NAD}^+ \rightleftharpoons \text{CO}_2 + \text{NADH} + \text{H}^+$ ). Upregulation of formate dehydrogenases was also found to correlate with lipid accumulation in a recent study (Zhang et al., 2019), but the biological significance of the upregulation remains unclear. Genes annotated with the broad GO term "transporters" were



**TABLE 3 |** GO term enrichment in differentially expressed genes. Go terms enriched more than 2.5-fold. Indented GO terms in brackets are a sub-group of the preceding broader higher-level GO term. Differentially expressed genes described as condition 1 vs. condition 2, where upregulated signifies that condition 1 has a higher transcript abundance, and downregulated signifies that condition 2 has a higher transcript abundance. Uncharacterized genes consist of unclassified or unknown genes.

### ***S. cerevisiae*: nitrogen depletion vs. growth**

#### **Upregulated genes**

Total: 401 genes, uncharacterized: 70 genes

<b>GO biological process complete</b>	<b>Gene count (331)</b>	<b>Expected</b>	<b>Fold enrichment</b>	<b>P-value</b>
Glyoxylate metabolic process (GO:0046487)	6	0.6	<b>11</b>	9.60E-05
Oligopeptide transmembrane transport (GO:0035672)	5	0.5	<b>10.2</b>	5.00E-04
Glutamate metabolic process (GO:0006536)	6	0.9	<b>6.5</b>	8.90E-04
Antibiotic metabolic process (GO:0016999)	18	3.4	<b>5.3</b>	8.70E-08
[Tricarboxylic acid cycle (GO:0006099)]	11	1.6	<b>6.7</b>	4.80E-06
'De novo' protein folding (GO:0006458)	7	1.3	<b>5.3</b>	8.50E-04
Ammonium transport (GO:0015696)	6	1.2	<b>5.2</b>	2.20E-03
Carbohydrate biosynthetic process (GO:0016051)	17	3.5	<b>4.9</b>	4.80E-07
Monosaccharide metabolic process (GO:0005996)	16	3.7	<b>4.4</b>	4.10E-06
Carbohydrate catabolic process (GO:0016052)	20	4.7	<b>4.3</b>	3.10E-07
Response to toxic substance (GO:0009636)	15	3.8	<b>3.9</b>	2.50E-05
Monocarboxylic acid metabolic process (GO:0032787)	35	3.8	<b>3.8</b>	1.50E-10
Carbohydrate metabolic process (GO:0005975)	50	13.2	<b>3.8</b>	1.70E-14
Small molecule catabolic process (GO:0044282)	30	8	<b>3.8</b>	5.00E-09
Response to oxidative stress (GO:0006979)	25	6.6	<b>3.8</b>	8.00E-08
Cellular glucan metabolic process (GO:0006073)	11	3	<b>3.7</b>	4.50E-04
Oxidation-reduction process (GO:0055114)	75	24.1	<b>3.1</b>	2.30E-17
[Carboxylic acid catabolic process (GO:0046395)]	15	4.8	<b>3.2</b>	2.20E-04
[Fatty acid beta-oxidation (GO:0006635)]	7	0.7	<b>10.7</b>	2.80E-05
Response to heat (GO:0009408)	15	4.9	<b>3.1</b>	2.70E-04
Drug metabolic process (GO:0017144)	43	15.3	<b>2.8</b>	6.10E-09
Cellular response to external stimulus (GO:0071496)	18	6.7	<b>2.7</b>	3.30E-04
Nucleobase-containing small molecule metabolic process (GO:0055086)	41	15.7	<b>2.6</b>	9.80E-08
[Nucleotide metabolic process (GO:0009117)]	34	13.6	<b>2.5</b>	3.00E-06
Cofactor metabolic process (GO:0051186)	37	14	<b>2.6</b>	4.20E-07
Coenzyme metabolic process (GO:0006732)	26	10.3	<b>2.5</b>	5.00E-05
[Glycolytic process (GO:0006096)]	7	1.4	<b>4.9</b>	1.30E-03

#### **Downregulated genes**

Total: 315 genes, uncharacterized: 29 genes

<b>GO biological process complete</b>	<b>Gene count (293)</b>	<b>Expected</b>	<b>Fold enrichment</b>	<b>P-value</b>
S-adenosylmethionine metabolic process (GO:0046500)	5	0.3	<b>14.8</b>	1.20E-04
Regulation of establishment or maintenance of cell polarity (GO:0032878)	4	0.3	<b>11.8</b>	1.10E-03
Nucleobase biosynthetic process (GO:0046112)	7	1.1	<b>6.6</b>	2.70E-04
DNA integration (GO:0015074)	14	2.4	<b>5.8</b>	8.80E-07
rRNA-containing ribonucleoprotein complex export from nucleus (GO:0071428)	14	3.1	<b>4.6</b>	8.90E-06
Ribosomal small subunit biogenesis (GO:0042274)	28	7	<b>4</b>	3.60E-09
DNA biosynthetic process (GO:0071897)	16	4.7	<b>3.4</b>	4.90E-05
Amide biosynthetic process (GO:0043604)	65	20.3	<b>3.2</b>	6.40E-16
[Translation (GO:0006412)]	61	17.7	<b>3.4</b>	3.10E-16
Transposition, RNA-mediated (GO:0032197)	14	4.4	<b>3.2</b>	3.00E-04
Ribosomal large subunit biogenesis (GO:0042273)	18	5.9	<b>3.1</b>	6.90E-05
Cellular amino acid metabolic process (GO:0006520)	37	12.3	<b>3</b>	1.10E-08
RNA phosphodiester bond hydrolysis (GO:0090501)	27	9.1	<b>3</b>	1.70E-06
Cellular amide metabolic process (GO:0043603)	69	23.6	<b>2.9</b>	4.30E-15
Nucleoside phosphate biosynthetic process (GO:1901293)	19	7.4	<b>2.6</b>	3.10E-04

### ***S. cerevisiae*: producing vs. non-producing**

#### **Upregulated genes**

Total: 21 genes, uncharacterized: four genes

<b>GO biological process complete</b>	<b>Gene count (18)</b>	<b>Expected</b>	<b>Fold enrichment</b>	<b>P-value</b>
Cell wall organization or biogenesis (GO:0071554)	9	0.8	<b>11.0</b>	7.2E-05

#### **Downregulated genes**

Total: three genes, uncharacterized: three genes

<b>GO biological process complete</b>	<b>Gene count (0)</b>	<b>Expected</b>	<b>Fold enrichment</b>	<b>P-value</b>
None	–	–	–	–

(Continued)

TABLE 3 | Continued

<i>Y. lipolytica</i> : nitrogen depletion vs. growth				
<b>Upregulated genes</b>				
Total: 500 genes, uncharacterized: 355 genes				
<b>GO biological process complete</b>	<b>Gene count (315)</b>	<b>Expected</b>	<b>Fold enrichment</b>	<b>P-value</b>
Transmembrane transport (GO:0055085)	70	20.6	<b>3.4</b>	2.8E-18
[Ammonium transmembrane transport (GO:0072488)]	9	0.6	<b>15.4</b>	1.7E-07
[Carboxylic acid transmembrane transport (GO:1905039)]	11	2.3	<b>4.9</b>	5.1E-05
[Amino acid transmembrane transport (GO:0003333)]	11	1.6	<b>6.8</b>	3.6E-06
<b>Downregulated genes</b>				
Total: 131 genes, uncharacterized: 54 genes				
<b>GO biological process complete</b>	<b>Gene count (121)</b>	<b>Expected</b>	<b>Fold enrichment</b>	<b>P-value</b>
Sulfur compound metabolic process (GO:0006790)	13	1.8	<b>7.3</b>	6.9E-08
<i>Y. lipolytica</i> producing vs. non-producing				
<b>Upregulated genes</b>				
Total: 60 genes, uncharacterized: 21 genes				
<b>GO biological process complete</b>	<b>Gene count (43)</b>	<b>Expected</b>	<b>Fold enrichment</b>	<b>P-value</b>
Antibiotic catabolic process (GO:0017001)	6	0.1	<b>42.8</b>	1.57E-08
[Formate catabolic process (GO:0042183)]	4	0.1	<b>60.0</b>	1.59E-06
<b>Downregulated genes</b>				
Total: 155 genes, uncharacterized: 79 genes				
<b>GO biological process complete</b>	<b>Gene count (120)</b>	<b>Expected</b>	<b>Fold enrichment</b>	<b>P-value</b>
Transport (GO:0006810)	51	16.5	<b>3.1</b>	2.2E-11
[Nitrogen compound transport (GO:0071705)]	26	7.7	<b>3.4</b>	6.8E-05
[Carboxylic acid transmembrane transport (GO:1905039)]	11	1.3	<b>8.7</b>	1.7E-04
[Amino acid transmembrane transport (GO:0003333)]	10	0.8	<b>12.1</b>	4.3E-05

significantly enriched among the 76 downregulated characterized genes. Among the enriched transporters, the carboxylic acid transmembrane transporters were again of particular interest as downregulation of those genes could explain the decrease of byproduct secretion rate observed in response to fatty alcohol production (Table 2). However, due to genes being associated with multiple GO terms, out of the 11 transporters associated with carboxylic acids, 10 of them were amino acid transporters. The remaining transporter, YALI0\_B19470g, was homologous to the *S. cerevisiae* transporter *JEN1*, annotated as a monocarboxylic acid transporter.

## DISCUSSION

<sup>13</sup>C-flux analysis revealed fundamentally different metabolic profiles of the two yeast species *S. cerevisiae* and *Y. lipolytica* (Figure 5), which is in line with previous findings (Christen and Sauer, 2011). This has implications for engineering strategies. Given that fatty alcohols have a much higher energy content than glucose, energy efficiency and abundant reducing power (NADPH) will likely be key factors for high-performing cell factories. For *S. cerevisiae*, this means that removal of ethanol production and increased NADPH generation are necessary. Ethanol production is a major drain of carbon atoms and is energetically inefficient. As for NADPH generation, this could, for example, be achieved by re-routing more carbon through the pentose phosphate pathway; another solution might be to replace NAD-dependent glycolytic enzymes with NADP-dependent ones (Kildegaard et al., 2016). It is likely that fatty alcohol production in *Y. lipolytica* would also benefit from increased NADPH supply. An increased proportion of carbon going through the pentose

phosphate pathway has previously been shown to improve lipid accumulation (Wasylenko et al., 2015). Efficient NADPH generation has also been achieved by expressing the NADP-dependent glycerol-3-phosphate dehydrogenase (*caGAPC*), with either an NADH kinase (*ylYEF1*) or a cytosolic NADP-dependent malic enzyme (*mcMCE2*) (Qiao et al., 2017).

Transcriptomic analysis comparing producing and non-producing strains of *S. cerevisiae* revealed a cell wall stress response. The cell wall stress response was indicated by both the GO term enrichment analysis (Table 3) as well as comparisons with previous studies (Arroyo et al., 2009). The cell wall-related upregulated genes have various functions. Crh1p and Crh2p are involved in cross-linking between the (1,6)- $\beta$ -glucan and chitin (Cabib et al., 2007). Slr2p is the MAPK (mitogen-activated protein kinase) responsible for triggering the cell wall stress response. Cwp1p, Pir2p, Pir3p, Ncw2p, Cis3p, Sed1p, Pst1p, and Ccw14p among others are covalently attached structural components of the cell wall. The reason for the triggering is unclear, but could conceivably be due to some kind of disturbance of the cell envelope caused by the produced fatty alcohols. It is interesting that no such stress response appears to be triggered in *Y. lipolytica*. The reasons for this difference are undetermined, but one hypothesis is that *Y. lipolytica*'s lipid accumulating abilities allow it to incorporate the produced fatty alcohols into lipid bodies, reducing disturbance to the cell envelope such as the one observed in *S. cerevisiae*. In terms of design considerations, avoiding the toxic effects of fatty alcohols will likely be a part of any beneficial design. The cell wall stress response seen in *S. cerevisiae* might prove difficult to overcome by rational design; it is, however, possible that adaptive laboratory evolution selecting for increased growth in the presence of fatty alcohols could alleviate the problem.

In *Y. lipolytica*, toxic effects were instead revealed in the non-producing strain in the form of a growth defect (lower growth rate and final OD), which to some degree seems to be alleviated by the expression by the *FAR* (Figures 4B, D). The reasons for this phenomenon are unclear. However, we hypothesize that the growth defect might be due to the accumulation of medium chain fatty aldehydes due to a reaction between unsaturated long chain fatty acids and free radicals and/or molecular oxygen in *hfd*-negative cells as described in a previous study (Xu et al., 2017). The generation of free radicals might be worsened by the *PEX10* knockout, which disrupts the peroxisome biogenesis and leads to oxidative stress (Van der Leij, 1992). Given that *malFAR* likely acts upon fatty aldehydes as a transient intermediate (Figure 1), it is possible that *malFAR* is also able to act upon medium chain fatty aldehydes, converting them into medium chain fatty alcohols, which possibly have a reduced toxic effect. At this point, this is mere speculation, and further experiments are needed to validate the hypothesis. Given the relocation of peroxisomal matrix proteins to the cytoplasm as a result of the *PEX10* knockout, it might prove a better approach to leave the peroxisome intact and all its enzymes enclosed. An alternative approach to *PEX10* deletion is to prevent  $\beta$ -oxidation by knocking out the *POX* enzymes responsible for the first step in the  $\beta$ -oxidation cycle.

Nitrogen depletion appears to have some major drawbacks, such as reduced cellular and metabolic activities (Tables 2 and 3). This can be due to quiescence, which is a beneficial adaptation in a natural environment (Coller, 2011). Several studies on the cellular response to nitrogen limitation were made (Morin et al., 2011; Kerkhoven et al., 2016; Pomraning et al., 2016); though more research is needed to apply this knowledge to improve the cellular performance under the nitrogen limitation.

Although the metabolomics data doesn't directly reveal the limiting steps of fatty alcohol synthesis, it may help narrow down the possible options by the assumption that the limiting steps occur in metabolic pathways where data is lacking. The combination of the depleted acetyl-CoA pool (Figure 5), the low glucose uptake rate, and high organic acid secretion rate (Table 2) indicate that there are one or more limiting steps occurring upstream of acetyl-CoA with the glucose uptake and glucose phosphorylation and conversion to F6P/FBP as possible candidates. In the case of *S. cerevisiae*, the pyruvate/acetaldehyde/acetate conversion is a potential target. The secretion profile (Table 2) of *Y. lipolytica* indicates that there are limiting steps following both pyruvate (pyruvate translocase and/or pyruvate dehydrogenase complex) and citrate (ATP-citrate lyase), or a possible competition with the plasma membrane-bound pyruvate and citrate transporters.

Furthermore, in *Y. lipolytica*, nitrogen depletion results in a significant increase in organic acid secretion, with the non-producing strain secreting ~50% of the total carbon consumed in the form of organic acids. Knocking out organic acid transporters might help to limit this carbon loss. Based on the correlation of changing secretion rates with differential gene expression, and excluding amino acid transporters, a single carboxylic acid transporter (YALI0\_B19470g) could be predicted and might be a contributor to the organic acid secretion. However, there are likely multiple other carboxylic transporters being upregulated

among the currently unknown genes. A recent study (Zhang et al., 2019) indicates that lowering the pH of the media may shift *Y. lipolytica* from citrate secretion to lipid accumulation. Furthermore, the secretion of organic acids could also be reduced by expressing downstream enzymes using cytosolic organic acids as substrates, such as ATP-citrate lyase (citrate + ATP  $\rightarrow$  acetyl-CoA + oxaloacetate) or pyruvate formate lyase (pyruvate  $\rightarrow$  acetyl-CoA + formate); both of which has previously been shown to boost lipid accumulation (Xu et al., 2016). Overexpression of native *Y. lipolytica* ATP-citrate lyase resulted in modest improvements; however, heterologous ATP-citrate lyases might be proven more beneficial. Organic acid secretion might also be decreased by increasing the indirect pull from enzymes further downstream in the pathway (Figure 2). Overexpression of acetyl-CoA carboxylase (*ACCI*) and stearoyl-CoA desaturase (*SGD1*) in combination with diacylglyceride acyl-transferase (*DGA1*) has been shown to greatly boost lipid accumulation in *Y. lipolytica* (Qiao et al., 2015). In terms of fatty alcohol production, it's possible that the same strategy could be implemented, but replacing *DGA1* with *FAR*.

This study describes a multi-omics analysis of the cellular and metabolic response to fatty alcohol production in two yeasts. It revealed cell wall stress response in fatty alcohol-producing *S. cerevisiae*. Furthermore, we have suggested designs that might aid in the engineering of fatty alcohol-producing cell factories.

## DATA AVAILABILITY

The datasets generated for this study can be found in European Nucleotide Archive, PRJEB32352.

## AUTHOR CONTRIBUTIONS

IB, JD, and CH conceived the study. JD performed the experiments. EM, CH, GW, and IB aided in troubleshooting and data interpretation. JD and DM performed method development, sampling and data processing of metabolomics data. JD and DW performed sample preparation and GCMS analysis for  $^{13}\text{C}$ -flux analysis. JD and UL performed data analysis of  $^{13}\text{C}$ -flux analysis data with the aid of BE. JD and CL performed data analysis of transcriptomic data. JD and IB wrote the manuscript with support from CH, EM, GW, UL, BE, MH, and LB. IB, LB, MH, and BE supervised the project. CH and GW helped supervise the project.

## FUNDING

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 722287. IB acknowledges the financial support from the Novo Nordisk Foundation (Grant agreement NNF15OC0016592 and NNF10CC1016517) and from the European Research Council under the European Union's Horizon 2020 research and innovation programme (YEAST-TRANS project, Grant Agreement No 757384). MH, CL, and IB

acknowledge the received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 760798 (OLEFINE project). LMB acknowledges funding by the Cluster of Excellence "The Fuel Science Center – Adaptive Conversion Systems for Renewable Energy and Carbon Sources," which is funded by the Excellence Initiative of the German federal and state governments to promote science and research at German universities.

## ACKNOWLEDGMENTS

We thank Alexandra Hoffmeyer and Pannipa Pornpitakpong, for performing the RNA-seq analysis. We would also like to

thank Dr. Hanne Bjerre Christensen and Dr. Lars Schrübbers for LC-MS and GC-MS analysis. We thank Prof. Volker Zickermann (Goethe-Universität) for the gift of *Y. lipolytica* GB20 strain and Dr. Peter Kötter, Johann Wolfgang Goethe University, Frankfurt, Germany for the gift of the *S. cerevisiae* CEN.PK113-7D strain. We would also like to thank Suresh Sudarsan and Marie Inger Dam for fruitful discussions.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2019.00747/full#supplementary-material>

## REFERENCES

- Arkin, A. P., Cottingham, R. W., Henry, C. S., Harris, N. L., Stevens, R. L., Maslov, S., et al. (2018). KBase: the United States department of energy systems biology knowledgebase. *Nat. Biotechnol.* 36, 566–569. doi: 10.1038/nbt.4163
- Arroyo, J., Bermejo, C., García, R., and Rodríguez-Peña, J. M. (2009). Genomics in the detection of damage in microbial systems: cell wall stress in yeast. *Clin. Microbiol. Infect.* 15, 44–46. doi: 10.1111/j.1469-0691.2008.02676.x
- Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., et al. (2000). Gene Ontology: tool for the unification of biology. *Nat. Genet.* 25, 25–29. doi: 10.1038/75556
- Blank, L. M., Lehmbeck, F., and Sauer, U. (2005). Metabolic-flux and network analysis in fourteen Hemiascomycetous yeasts. *FEMS Yeast Res.* 5 (6–7), 545–558. doi: 10.1016/j.femsyr.2004.09.008
- Boer, V. M., de Winde, J. H., Pronk, J. T., and Piper, M. D. (2003). The genome-wide transcriptional responses of *Saccharomyces cerevisiae* grown on glucose in aerobic chemostat cultures limited for carbon, nitrogen, phosphorus, or sulfur. *J. Biol. Chem.* 278 (5), 3265–3274. doi: 10.1074/jbc.M209759200
- Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30 (15), 2114–2120. doi: 10.1093/bioinformatics/btu170
- Boorsma, A., de Nobel, H., ter Riet, B., Bargmann, B., Brul, S., Hellingwerf, K. J., et al. (2004). Characterization of the transcriptional response to cell wall stress in *Saccharomyces cerevisiae*. *Yeast* 21 (5), 413–427. doi: 10.1002/yea.1109
- Borodina, I., Holkenbrink, C., Dam, M. I., and Löfstedt, C., (2018). *Methods for producing fatty alcohols and derivatives thereof in yeast*. WO2018109163A1. Geneva, Switzerland: World Intellectual Property Organization.
- Buijs, N. A., Zhou, Y. J., Siewers, V., and Nielsen, J. (2015). Long-chain alkane production by the yeast *Saccharomyces cerevisiae*. *Biotechnol. Bioeng.* 112 (6), 1275–1279. doi: 10.1002/bit.25522
- Cabib, E., Blanco, N., Grau, C., Rodríguez-Peña, J. M., and Arroyo, J. (2007). Crh1p and Crh2p are required for the cross-linking of chitin to  $\beta(1-6)$ Glucan in the *Saccharomyces cerevisiae* cell wall. *Mol. Microbiol.* 63 (3), 921–935. doi: 10.1111/j.1365-2958.2006.05565.x
- Christen, S., and Sauer, U. (2011). Intracellular characterization of aerobic glucose metabolism in seven yeast species by 13C flux analysis and metabolomics. *FEMS Yeast Res.* 11 (3), 263–272. doi: 10.1111/j.1567-1364.2010.00713.x
- Coller, H. A. (2011). The essence of quiescence. *Science* 334 (6059), 1074–1075. doi: 10.1126/science.1216242
- Daugherty, J. R., Rai, R., el Berry, H. M., and Cooper, T. G. (1993). Regulatory circuit for responses of nitrogen catabolic gene expression to the GLN3 and DAL80 proteins and nitrogen catabolite repression in *Saccharomyces cerevisiae*. *J. Bacteriol.* 175 (1), 64–73. doi: 10.1128/jb.175.1.64-73.1993
- d'Espaux, L., Ghosh, A., Runguphan, W., Wehrs, M., Xu, F., Konzock, O., et al. (2017). Engineering high-level production of fatty alcohols by *Saccharomyces cerevisiae* from lignocellulosic feedstocks. *Metab. Eng.* 42, 115–125. doi: 10.1016/j.ymben.2017.06.004
- Entian, K. D., and Kötter, P. (2007). Yeast genetic strain and plasmid collections. *Method Microbiol.* 36, 629–666. doi: 10.1016/S0580-9517(06)36025-4
- Fillet, S., Gibert, J., Suárez, B., Lara, A., Ronchel, C., and Adrio, J. L. (2015). Fatty alcohols production by oleaginous yeast. *J. Ind. Microbiol. Biotechnol.* 42 (11), 1463–1472. doi: 10.1007/s10295-015-1674-x
- Gaillardin, C. M., Charoy, V., and Heslot, H. (1973). A study of copulation, sporulation and meiotic segregation in *Candida lipolytica*. *Arch. Mikrobiol.* 92 (1), 69–83. doi: 10.1007/BF00409513
- García, R., Bermejo, C., Grau, C., Pérez, R., Rodríguez-Peña, J. M., Francois, J., et al. (2004). The global transcriptional response to transient cell wall damage in *Saccharomyces cerevisiae* and its regulation by the cell integrity signaling pathway. *J. Biol. Chem.* 279 (15), 15183–15195. doi: 10.1074/jbc.M312954200
- Grand View Research, Inc. (2016). Fatty alcohols market by product (C6-C10, C11-C14, C15-C22), by application (soaps & detergents, personal care, lubricants, amines) expected to reach USD 7.51 billion by 2022. Available at: <https://www.grandviewresearch.com/press-release/global-fatty-alcohols-market>.
- Holkenbrink, C., Dam, M. I., Kildegaard, K. R., Beder, J., Dahlin, J., Belda, D. D., et al. (2018). EasyCloneYALI: CRISPR/Cas9-based synthetic toolbox for engineering of the yeast *Yarrowia lipolytica*. *Biotechnol. J.* 13, 1700543. doi: 10.1002/biot.201700543
- Iwama, R., Kobayashi, S., Ohta, A., Horiuchi, H., and Fukuda, R. (2014). Fatty aldehyde dehydrogenase multigene family involved in the assimilation of N-alkanes in *Yarrowia lipolytica*. *J. Biol. Chem.* 289 (48), 33275–33286. doi: 10.1074/jbc.M114.596890
- Jessop-Fabre, M. M., Jakočiūnas, T., Stovicek, V., Dai, Z., Jensen, M. K., Keasling, J. D., et al. (2016). EasyClone-MarkerFree: a vector toolkit for marker-less integration of genes into *Saccharomyces cerevisiae* via CRISPR-Cas9. *Biotechnol. J.* 11 (8), 1110–1117. doi: 10.1002/biot.201600147
- Kerkhoven, E. J., Pomraning, K. R., Baker, S. E., and Nielsen, J. (2016). Regulation of amino-acid metabolism controls flux to lipid accumulation in *Yarrowia lipolytica*. *NPJ Syst. Biol. Appl.* 2, 16005. doi: 10.1038/npsba.2016.5
- Kildegaard, K. R., Jensen, N. B., Schneider, K., Czarnotta, E., Özdemir, E., Klein, T., et al. (2016). Engineering and systems-level analysis of *Saccharomyces cerevisiae* for production of 3-hydroxypropionic acid via malonyl-CoA reductase-dependent pathway. *Microb. Cell Fact.* 15, 53. doi: 10.1186/s12934-016-0451-5
- Kim, D., Langmead, B., and Salzberg, S. L. (2015). HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* 12 (4), 357–360. doi: 10.1038/nmeth.3317
- Lagorce, A., Hauser, N. C., Labourdette, D., Rodriguez, C., Martin-Yken, H., Arroyo, J., et al. (2003). Genome-wide analysis of the response to cell wall mutations in the yeast *Saccharomyces cerevisiae*. *J. Biol. Chem.* 278 (22), 20345–20357. doi: 10.1074/jbc.M211604200
- Liu, Y., Chen, S., Chen, J., Zhou, J., Wang, Y., Yang, M., et al. (2016). High production of fatty alcohols in *Escherichia coli* with fatty acid starvation. *Microb. Cell Fact.* 15 (1), 129. doi: 10.1186/s12934-016-0524-5
- McCloskey, D., Gangoti, J. A., Palsson, B. O., and Feist, A. M. (2015a). A PH and solvent optimized reverse-phase ion-pairing-LC-MS/MS method that leverages multiple scan-types for targeted absolute quantification of intracellular metabolites. *Metabolomics* 11 (5), 1338–1350. doi: 10.1007/s11306-015-0790-y



- McCloskey, D., Utrilla, J., Naviaux, R. K., Palsson, B. O., and Feist, A. M. (2015b). Fast Swinnex Filtration (FSF): a fast and robust sampling and extraction method suitable for metabolomics analysis of cultures grown in complex media. *Metabolomics* 11 (1), 198–209. doi: 10.1007/s11306-014-0686-2
- McNeil, B. A., and Stuart, D. T. (2018). Optimization of C16 and C18 fatty alcohol production by an engineered strain of *Lipomyces starkeyi*. *J. Ind. Microbiol. Biotechnol.* 45 (1), 1–14. doi: 10.1007/s10295-017-1985-1
- Mi, H., Huang, X., Muruganujan, A., Tang, H., Mills, C., Kang, D., et al. (2017). PANTHER version 11: expanded annotation data from gene ontology and reactome pathways, and data analysis tool enhancements. *Nucleic Acids Res.* 45 (D1), D183–D189. doi: 10.1093/nar/gkw1138
- Morin, N., Cescut, J., Beopoulos, A., Lelandaïs, G., Le Berre, V., Uribealarea, J.-L., et al. (2011). Transcriptomic analyses during the transition from biomass production to lipid accumulation in the oleaginous yeast *Yarrowia lipolytica*. *PLOS ONE* 6 (11), e27966. doi: 10.1371/journal.pone.0027966
- Okonechnikov, K., Conesa, A., and García-Alcalde, F. (2016). Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics* 32 (2), 292–294. doi: 10.1093/bioinformatics/btv566
- Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T.-C., Mendell, J. T., and Salzberg, S. L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* 33 (3), 290–295. doi: 10.1038/nbt.3122
- Pomraning, K. R., Kim, Y.-M., Nicora, C. D., Chu, R. K., Bredeweg, E. L., Purvine, S. O., et al. (2016). Multi-omics analysis reveals regulators of the response to nitrogen limitation in *Yarrowia lipolytica*. *BMC Genomics* 17, 138. doi: 10.1186/s12864-016-2471-2
- Poskar, C. H., Huege, J., Krach, C., Franke, M., Shachar-Hill, Y., and Junker, B. H. (2012). IMS2Flux—a high-throughput processing tool for stable isotope labeled mass spectrometric data used for metabolic flux analysis. *BMC Bioinformatics* 13 (1), 295. doi: 10.1186/1471-2105-13-295
- Qiao, K., Imam Abidi, S. H., Liu, H., Zhang, H., Chakraborty, S., Watson, N., et al. (2015). Engineering lipid overproduction in the oleaginous yeast *Yarrowia lipolytica*. *Metab. Eng.* 29, 56–65. doi: 10.1016/j.ymben.2015.02.005
- Qiao, K., Wasylenko, T. M., Zhou, K., Xu, P., and Stephanopoulos, G. (2017). Lipid production in *Yarrowia lipolytica* is maximized by engineering cytosolic redox metabolism. *Nat. Biotechnol.* 35, 173–177. doi: 10.1038/nbt.3763
- Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2010). EdgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26 (1), 139–140. doi: 10.1093/bioinformatics/btp616
- Rodríguez-Peña, J. M., Pérez-Díaz, R. M., Álvarez, S., Bermejo, C., García, R., Santiago, C., et al. (2005). The ‘Yeast Cell Wall Chip’—a tool to analyse the regulation of cell wall biogenesis in *Saccharomyces cerevisiae*. *Microbiology* 151 (7), 2241–2249. doi: 10.1099/mic.0.27989-0
- Schmitz, A., Ebert, B. E., and Blank, L. M., (2017). “GC-MS-based determination of mass isotopomer distributions for 13C-based metabolic flux analysis,” in *Hydrocarbon and Lipid Microbiology Protocols: Genetic, Genomic and System Analyses of Pure Cultures*. Eds. T. J. McGenity, K. N. Timmis, and B. Nogales (Berlin, Heidelberg: Springer Berlin Heidelberg: Springer Protocols Handbooks), 223–243. doi: 10.1007/8623\_2015\_78
- Shah, J., Arslan, E., Cirucci, J., O’Brien, J., and Moss, D. (2016). Comparison of oleo- vs petro-sourcing of fatty alcohols via cradle-to-gate life cycle assessment. *J. Surfactants Deterg.* 19 (6), 1333–1351. doi: 10.1007/s11743-016-1867-y
- Tai, M., and Stephanopoulos, G. (2013). Engineering the push and pull of lipid biosynthesis in oleaginous yeast *Yarrowia lipolytica* for biofuel production. *Metab. Eng.* 15, 1–9. doi: 10.1016/j.ymben.2012.08.007
- ter Schure, E. G., van Riel, N. A., and Verrips, C. T. (2000). The role of ammonia metabolism in nitrogen catabolite repression in *Saccharomyces cerevisiae*. *FEMS Microbiol. Rev.* 24 (1), 67–83. doi: 10.1111/j.1574-6976.2000.tb00533.x
- The Gene Ontology Consortium (2017). Expansion of the gene ontology knowledgebase and resources. *Nucleic Acids Res.* 45 (D1), D331–D338. doi: 10.1093/nar/gkw1108
- Thomas, P. D., Campbell, M. J., Kejariwal, A., Mi, H., Karlak, B., Daverman, R., et al. (2003). PANTHER: a library of protein families and subfamilies indexed by function. *Genome Res.* 13 (9), 2129–2141. doi: 10.1101/gr.772403
- Van der Leij, I. (1992). Isolation of peroxisome assembly mutants from *Saccharomyces cerevisiae* with different morphologies using a novel positive selection procedure. *J. Cell Biol.* 119 (1), 153–162. doi: 10.1083/jcb.119.1.153
- Wasylenko, T. M., and Stephanopoulos, G. (2015). Metabolomic and 13C-Metabolic flux analysis of a xylose-consuming *Saccharomyces cerevisiae* strain expressing xylose isomerase. *Biotechnol. Bioeng.* 112 (3), 470–483. doi: 10.1002/bit.25447
- Wasylenko, T. M., Ahn, W. S., and Stephanopoulos, G. (2015). The oxidative pentose phosphate pathway is the primary source of NADPH for lipid overproduction from glucose in *Yarrowia lipolytica*. *Metab. Eng.* 30, 27–39. doi: 10.1016/j.ymben.2015.02.007
- Xu, P., Qiao, K., Ahn, W. S., and Stephanopoulos, G. (2016). Engineering *Yarrowia lipolytica* as a platform for synthesis of drop-in transportation fuels and oleochemicals. *Proc. Natl. Acad. Sci.* 113 (39), 10848–10853. doi: 10.1073/pnas.1607295113
- Xu, P., Qiao, K., and Stephanopoulos, G. (2017). Engineering oxidative stress defense pathways to build a robust lipid production platform in *Yarrowia lipolytica*. *Biotechnol. Bioeng.* 114 (7), 1521–1530. doi: 10.1002/bit.26285
- Young, J. D. (2014). INCA: a computational platform for isotopically non-stationary metabolic flux analysis. *Bioinformatics* 30 (9), 1333–1335. doi: 10.1093/bioinformatics/btu015
- Zamboni, N., Fendt, S.-M., Rühl, M., and Sauer, U. (2009). 13C-based metabolic flux analysis. *Nat. Protoc.* 4 (6), 878–892. doi: 10.1038/nprot.2009.58
- Zhang, S., Jagtap, S. S., Deewan, A., and Rao, C. V. (2019). PH selectively regulates citric acid and lipid production in *Yarrowia lipolytica* w29 during nitrogen-limited growth on glucose. *J. Biotechnol.* 290, 10–15. doi: 10.1016/j.jbiotec.2018.10.012

**Conflict of Interest Statement:** IB and CH have a financial interest in BioPhero ApS. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Dahlin, Holkenbrink, Marella, Wang, Liebal, Lieven, Weber, McCloskey, Ebert, Herrgård, Blank and Borodina. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Ploidy Variation in *Kluyveromyces marxianus* Separates Dairy and Non-dairy Isolates

Raúl A. Ortiz-Merino<sup>1†</sup>, Javier A. Varela<sup>2†</sup>, Aisling Y. Coughlan<sup>1</sup>, Hisashi Hoshida<sup>3</sup>, Wendel B. da Silveira<sup>4</sup>, Caroline Wilde<sup>5</sup>, Niels G. A. Kuijpers<sup>6</sup>, Jan-Maarten Geertman<sup>6</sup>, Kenneth H. Wolfe<sup>1</sup> and John P. Morrissey<sup>2\*</sup>

<sup>1</sup> School of Medicine, UCD Conway Institute, University College Dublin, Dublin, Ireland, <sup>2</sup> School of Microbiology, Centre for Synthetic Biology and Biotechnology, Environmental Research Institute, APC Microbiome Institute, University College Cork, Cork, Ireland, <sup>3</sup> Department of Applied Chemistry, Graduate School of Sciences and Technology for Innovation, Yamaguchi University, Yamaguchi, Japan, <sup>4</sup> Department of Microbiology, Universidade Federal de Viçosa, Viçosa, Brazil, <sup>5</sup> Lallemand Inc., Montreal, QC, Canada, <sup>6</sup> Heineken Supply Chain, Zoeterwoude, Netherlands

## OPEN ACCESS

### Edited by:

Isabel Sá-Correia,  
Instituto Superior Técnico,  
Universidade de Lisboa, Portugal

### Reviewed by:

Amparo Querol,  
Consejo Superior de Investigaciones  
Científicas (CSIC), Spain  
José Manuel Guzmán,  
Consejo Superior de Investigaciones  
Científicas (CSIC), Spain

### \*Correspondence:

John P. Morrissey  
j.morrissey@ucc.ie

<sup>†</sup>These authors have contributed  
equally to this work.

### Specialty section:

This article was submitted to  
Evolutionary and Genomic  
Microbiology,  
a section of the journal  
Frontiers in Genetics

Received: 20 January 2018

Accepted: 05 March 2018

Published: 21 March 2018

### Citation:

Ortiz-Merino RA, Varela JA, Coughlan AY, Hoshida H, Silveira WB, Wilde C, Kuijpers NGA, Geertman J-M, Wolfe KH and Morrissey JP (2018) Ploidy Variation in *Kluyveromyces marxianus* Separates Dairy and Non-dairy Isolates. *Front. Genet.* 9:94. doi: 10.3389/fgene.2018.00094

*Kluyveromyces marxianus* is traditionally associated with fermented dairy products, but can also be isolated from diverse non-dairy environments. Because of thermotolerance, rapid growth and other traits, many different strains are being developed for food and industrial applications but there is, as yet, little understanding of the genetic diversity or population genetics of this species. *K. marxianus* shows a high level of phenotypic variation but the only phenotype that has been clearly linked to a genetic polymorphism is lactose utilisation, which is controlled by variation in the *LAC12* gene. The genomes of several strains have been sequenced in recent years and, in this study, we sequenced a further nine strains from different origins. Analysis of the Single Nucleotide Polymorphisms (SNPs) in 14 strains was carried out to examine genome structure and genetic diversity. SNP diversity in *K. marxianus* is relatively high, with up to 3% DNA sequence divergence between alleles. It was found that the isolates include haploid, diploid, and triploid strains, as shown by both SNP analysis and flow cytometry. Diploids and triploids contain long genomic tracts showing loss of heterozygosity (LOH). All six isolates from dairy environments were diploid or triploid, whereas 6 out of 7 isolates from non-dairy environment were haploid. This also correlated with the presence of functional *LAC12* alleles only in dairy haplotypes. The diploids were hybrids between a non-dairy and a dairy haplotype, whereas triploids included three copies of a dairy haplotype.

**Keywords:** lactose transport, non-conventional yeast, yeast evolution, industrial yeast, dairy, *Kluyveromyces*, *LAC12*

## INTRODUCTION

The yeast *Kluyveromyces marxianus* is best-known because of its frequent association with traditional dairy products such as kefir and cheese (Lachance, 2011; Gethins et al., 2016; Coloretto et al., 2017). This association with fermented dairy beverages, a consequence of its capacity to use the milk sugar lactose as a carbon source, has led to inclusion of *K. marxianus* on GRAS (FDA) and QPS (EU) lists of safe micro-organism for use in foods (Lane and Morrissey, 2010; Ricci et al., 2017).

The yeast is also regularly isolated from non-dairy environments (e.g., decaying fruit) and is part of the natural flora involved in production of *Agave*-based alcoholic beverages such as tequila and mezcal (Lappe-Oliveras et al., 2008; Verdugo Valdez et al., 2011). In the latter case, the production of enzymes that degrade plant fructans to simpler sugars (inulinases) undoubtedly contributes to its growth in this environment (Arrizon et al., 2012). The capacity of *K. marxianus* to utilise a broad array of sugars also creates potential for biotechnological applications (Fonseca et al., 2008; Lane and Morrissey, 2010), which is illustrated by the many studies exploring potential for bioethanol production from diverse substrates such as whey permeate, crop plants, and lignocellulosic biomass (Nonklang et al., 2009; Guimarães et al., 2010; Wu et al., 2016; Kobayashi et al., 2017). This yeast is used commercially for production of the flavour molecule 2-phenylethanol, and there is considerable interest in development of *K. marxianus* as a cell factory for production of other bioflavours (Morrissey et al., 2015). *K. marxianus* is also distinguished by thermotolerance (Lane et al., 2011), and the fastest reported growth rate of any eukaryote (Groeneveld et al., 2009). Recent years have seen increasing interest in new applications such as production of biomolecules (Hughes et al., 2017; Lin et al., 2017), biocatalysis (Oliveira et al., 2017; Wang et al., 2017) and heterologous protein production (Gombert et al., 2016; Lee et al., 2017).

One of the interesting aspects of the nascent development of *K. marxianus* as an important yeast for biotechnology is the wide variety of strains that are being used, both for research and for application. This contrasts with the traditional yeast, *Saccharomyces cerevisiae*, where, until recently, there was a very strong focus on a relatively narrow set of model strains. While giving access to the broad diversity that exists within any species, the non-reliance on model strains also creates challenges since findings with one isolate are not automatically transferrable to other isolates. This is illustrated well by studies that demonstrate wide variance in tolerance to different external stresses (Lane et al., 2011; Rocha et al., 2011). Indeed, it has emerged that even a trait such as lactose utilisation, long considered one of the defining characteristics of *K. marxianus*, is not universal, and many strains exhibit very poor growth on lactose, a phenotype that was shown to be due to polymorphisms in the *LAC12* gene, which encodes a permease responsible for transport of lactose into the cell (Varela et al., 2017). Although recent studies on sugar transport and physiology are starting to address the deficit (Fonseca et al., 2013; Signori et al., 2014; Beniwal et al., 2017; Dias et al., 2017; Diniz et al., 2017), it is true to say that a lot of the underlying knowledge about the biology of *K. marxianus* is based on inference of similarity with its sister species, *Kluyveromyces lactis*, which was developed as a model for studying lactose-positive yeasts since the 1960's (Fukuhara, 2006). The genome of *K. lactis* was sequenced more than a decade ago (Souciet et al., 2000), with a more recent functional reannotation and genome scale model that provides a deeper understanding of the core metabolism of this species (Dias et al., 2012, 2014). Notwithstanding the utility of a related species for comparison, the many metabolic and physiological differences between *K. lactis* and *K. marxianus*

necessitate independent studies of *K. marxianus* to provide the comprehensive understanding of its genetics and physiology that will underpin future developments in fundamental biology and biotechnology.

Genomic and transcriptomic studies have started to shed light on *K. marxianus* and a growing number of genome sequences of *K. marxianus* strains are now available (Jeong et al., 2012; Silveira et al., 2014; Inokuma et al., 2015; Lertwattanasakul et al., 2015; Quarella et al., 2016). As yet, however, there has not been a systematic comparison of the sequenced *K. marxianus* genomes, nor a comparison to the single *K. lactis* genome that is in the public domain. In contrast to *K. lactis*, whose genome comprises 6 chromosomes, several studies have reported that *K. marxianus* has a full complement of 8 chromosomes, with many areas of local synteny between the species. There is strong conservation of the mating type locus (Lane et al., 2011) and thus *K. marxianus* could be expected to be capable of mating type switching and mating in a manner similar to *K. lactis* (Barsoum et al., 2010; Rajaei et al., 2014). Based on information to date, however, there does appear to be a fundamental difference in life-cycles. Studies of natural isolates of *K. lactis* suggest that this yeast is primarily a haploid (haplontic) species. Mating is induced by depletion of nitrogen or phosphate in the environment, and zygotes formed by mating usually sporulate immediately (although diploids can be maintained in the lab, e.g., by selection for auxotrophic markers) (Schaffrath and Breunig, 2000; Zonneveld and Steensma, 2003; Booth et al., 2010; Rodicio and Heinisch, 2013). In contrast, analysis of the mating-type locus of natural and culture collection *K. marxianus* isolates identified both haploid and diploid strains (Lane et al., 2011; Fasoli et al., 2016).

To put the phenotypic diversity of *K. marxianus* into context, it is important to characterise its genomic diversity and to assess the population structure of the species. There have been some pre-whole genome sequence studies that addressed this question using different methods. Pulsed-field gel electrophoresis studies suggested that there were variable numbers of chromosomes in *K. marxianus* strains (Belloch et al., 1998; Fasoli et al., 2015), a finding not in accordance with genome sequence data, which has consistently indicated 8 chromosomes. Mitochondrial DNA haplotypes and variation at some genomic loci was used to try to determine population structure in a collection from Italian cheeses (Fasoli et al., 2016). That particular study identified variations in population structure and proposed the occurrence of homozygous and heterozygous strains. A Multi Locus Sequence Typing (MLST) method was developed to further explore the diversity in that collection and in this case, the analysis was extended to other strains that are sequenced or available in culture collections (Tittarelli et al., 2018). MLST analysis did not identify distinct sub-populations but while the method was very diagnostic for strain identification, the surprisingly high level of heterozygosity in diploid strains reduced resolution to a level too low for population-type analysis. Analysis of population structure in diploid yeasts is challenging and, in many cases, has relied on SNPs identified in genome sequences derived from haploids or from completely homozygous diploids (made by self-mating of

single spore derivatives) (Liti et al., 2009; Schacherer et al., 2009; Strobe et al., 2015). In highly heterozygous species, this method may not generate an accurate view of the relationships among haplotypes or among strains.

In this study, we set out to explore the genomic diversity of *K. marxianus* by analysing whole-genome data from 14 strains isolated from different sources. Some of these strains had been previously sequenced and published, whereas others were sequenced for this study. To take heterozygosity into account, raw sequence reads were used to allow analysis of single nucleotide polymorphisms (SNPs) between strains. The results indicate a high degree of variation among isolates, in both ploidy and heterozygosity, and show a correlation between ploidy and environmental niche. Our work raises important questions about the life cycle of *K. marxianus*, and emphasizes the need to take ploidy and heterozygosity into account when considering using *K. marxianus* for biotechnological purposes.

## MATERIALS AND METHODS

### Yeast Strains, Growth, and Phenotypic Analysis

The 14 *K. marxianus* strains analysed in this study are listed in **Table 1**. Two strains (DMKU3-1042 and UFS-Y2791) were not available for phenotypic assessment but the remaining 12 strains were obtained from the sources indicated in **Table 1** and were routinely cultured at 30°C in YPD medium (10 g/L yeast extract, 20 g/L bactopectone, 20 g/L glucose). For lactose utilisation tests, yeast strains were first grown overnight in 5 mL minimal media (MM) supplemented with 2% glucose (Fonseca et al., 2007). Cells from the overnight cultures were harvested by centrifugation, washed twice with 5 mL of water and used to inoculate MM supplemented with 2% lactose to an OD<sub>600</sub> of 0.1. These cultures were incubated for 15 h, when the final OD<sub>600</sub> was determined. Lactose concentration was determined by HPLC at 0 and 15 h and

**TABLE 1** | Sources of *K. marxianus* strains and genomes analysed in this study.

Strain	Synonym	Country	Sample source	Strain source	Reference for genome sequence	Source of Illumina FASTQ data	Accession numbers for Illumina data
L01		Unknown	Dairy	Lallemand Inc.	This study	University College Dublin (K.H. Wolfe)	SRX3541360
L02		Unknown	Dairy	Lallemand Inc.	This study	University College Dublin (K.H. Wolfe)	SRX3541359
L03		Unknown	Dairy	Lallemand Inc.	This study	University College Dublin (K.H. Wolfe)	SRX3541362
L04		Unknown	Baking	Lallemand Inc.	This study	University College Dublin (K.H. Wolfe)	SRX3541361
L05		Unknown	Distillery	Lallemand Inc.	This study	University College Dublin (K.H. Wolfe)	SRX3541364
CBS397		Netherlands	Yoghurt	Westerdijk Institute, Netherlands	This study	University College Cork (J.P. Morrissey)	SRX3541363
NBRC0272		Unknown	Miso	Biological Resource Center, NITE (NBRC), Japan	This study	Yamaguchi University (H. Hoshida)	SRX3541366
NBRC0288	DSM4906	Unknown	Unknown	Biological Resource Center, NITE (NBRC), Japan	This study	Yamaguchi University (H. Hoshida)	SRX3541365
NBRC0617	ATCC8622	Denmark	Yoghurt	Biological Resource Center, NITE (NBRC), Japan	This study	Yamaguchi University (H. Hoshida)	SRX3541358
NBRC1777		Japan	Soil	Biological Resource Center, NITE (NBRC), Japan	Inokuma et al., 2015*	Yamaguchi University (H. Hoshida)	SRX3541357
CBS6556	KCTC17555, ATCC26548	Mexico	Pozol	Westerdijk Institute, Netherlands	Jeong et al., 2012	Yonsei University (J. F. Kim)	SRX3637961
UFV-3	CCT7735	Brazil	Dairy	Universidade Federal de Viçosa, Brazil	Silveira et al., 2014	BIOAGRO, Brazil (F. M. L. Passos)	SRX3637959
DMKU3-1042†		Thailand	Soil	Strain not obtained	Lertwattanasakul et al., 2015	Yamaguchi University (H. Hoshida)	SRX3541367
UFS-Y2791		South Africa	Agave americana juice	Strain not obtained	Schabot et al., 2016	Univ. of the Free State (D. T. W. P. Schabot)	SRX3637960

\*The reference genome sequence of NBRC1777 (Inokuma et al., 2015) was based on an assembly of Pacific Biosciences and Ion Torrent data but the SNP analysis in this study used newly-generated Illumina FASTQ data. †Genome sequence obtained from a *ura3* derivative generated by UV mutagenesis.



used to calculate lactose consumption as previously described (Varela et al., 2017). Experiments were performed in triplicate with error bars showing standard deviation.

## Flow Cytometry

DNA content was determined by flow cytometry using SYTOX green (Thermo-Fisher) as previously described (Haase and Reed, 2002). Yeast strains were grown in YPD at 30°C with 200 rpm agitation in a New Brunswick Innova 40/40 R orbital shaker (Eppendorf, Hamburg, Germany). Cultures were harvested by centrifugation and resuspended in 1 mL sterile water. Cells were then washed, resuspended in 400  $\mu$ L sterile water and fixed by adding 950  $\mu$ L 100% ethanol. The suspensions were incubated overnight at 4°C, then centrifuged and washed in 50 mM sodium citrate (pH 7.2). The cells were resuspended in 500  $\mu$ L RNase A solution (0.25 mg/mL RNase A, 50 mM sodium citrate pH 7.2) and incubated for 1 h at 37°C. Then, 100  $\mu$ L of 20 mg/mL Proteinase K was added to each sample and the tubes were incubated at 50°C for 2 h. Finally, 500  $\mu$ L of SYTOX Green solution (4  $\mu$ M SYTOX Green, 50 mM sodium citrate pH 7.2) was added to each tube. Samples were analysed using a BD FACSCelesta system (BD Biosciences, CA, USA) and the data was processed using FlowJo software v10 (BD Biosciences, CA, USA).

## Genome Data, Sequencing, and Read Mapping

The genomes of five strains (NBRC1777, CBS6556, UFV-3, DMKU3-1042, and UFS-Y2791) had previously been published and some of the authors kindly made the source Illumina FASTQ data available for this analysis. The 11 strains sequenced in this study are indicated in **Table 1**, accession numbers are provided for all strains and references are given when applicable. All strains were sequenced on Illumina HiSeq 2000 or 2500 instruments after Truseq genomic library preparation. Details of the sequencing data type and coverage for all 14 strains, including those sequenced elsewhere, are summarised in **Table 2**. We performed quality control checks for all libraries with FastQC v. 0.10.1 (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Reads for strain CBS6556 were trimmed using skewer v. 0.2.2 (Jiang et al., 2014) with parameters -q 20 -m pe -l 70. For BLAST analyses, *de novo* assemblies of each newly sequenced genome were made using SPAdes 3.5.0 (Bankevich et al., 2012).

The NBRC1777 genome sequence was selected as a reference because of the high quality of its assembly into 8 chromosomes, and because our analysis confirmed the haploid nature of this strain (Inokuma et al., 2015). Sequencing libraries from all strains were aligned to the NBRC1777 reference using the Burrows-Wheeler Aligner (BWA) v. 0.7.9a-r786 (Li and Durbin, 2009) with default parameters. The BWA “mem” alignment algorithm was used for libraries with read length  $\geq 100$  bp, the “aln” alignment algorithm for libraries with read length  $< 100$  bp, and the alignment modes were set to “samse” and “sampe” for single-end and paired-end data respectively. Samtools v. 0.1.19-44428cd (Li et al., 2009) was used to remove unmapped reads from the BWA output

**TABLE 2 |** Summary of Illumina sequencing strategies and coverage for 14 strains used in SNP analysis.

Strain	Mean coverage (x)	Pairedness*	Read length	Million reads	BWA algorithm
L01	47	SE	50	16.6	aln samse
L02	47	SE	50	16.7	aln samse
L03	45	SE	50	16.7	aln samse
L04	45	SE	50	15.6	aln samse
L05	47	SE	50	16.2	aln samse
CBS397	141	PE	126	15.1	mem
NBRC0272	160	PE	100	20.4	mem
NBRC0288	212	PE	100	25.6	mem
NBRC0617	35	SE	50	11.6	aln samse
NBRC1777	110	PE	100	12.8	mem
CBS6556	623	PE	70-150	56.9	mem
UFV-3	359	PE	90	50.3	mem
DMKU3-1042	341	PE	100	44.6	mem
UFS-Y2791	50	PE	75-100	17.1	mem

\*SE, single-end; PE, paired-end.

files and to generate indexes for downstream steps. Picard tools v. 2.0.1 (<http://broadinstitute.github.io/picard>) function AddOrReplaceReadGroups was used to add identifiers to the BAM files, followed by MarkDuplicates to mark and discard PCR duplicates. Indel realignment and coverage calculation were performed using the RealignerTargetCreator, IndelRealigner, and DepthOfCoverage tools from the Genome Analysis Tool Kit (GATK) v. 3.5-0-g36282e4 (Van der Auwera et al., 2013). Mean coverage was calculated omitting a 19 kb region on chromosome 5 that contains the array encoding the rRNA genes. Coverage plots were obtained by calculating the average in 10-kb windows. Segment means were calculated using the R Bioconductor package DNACopy v 1.50.1 (DOI: 10.18129/B9.bioc.DNACopy).

## Variant Calling

“Variable sites” were defined as the set of sites in the genome that contain a non-reference base, hereafter called “variants,” in at least one of the 14 strains. Variant calling in the 14 strains was done using the GATK tool HaplotypeCaller in DISCOVERY and GVCF modes, requiring a minimum quality score of 20. The output files were then used for multi-sample analysis using the GenotypeGVCF tool and a custom Perl script was used to remove all variants that had low genotype quality (GQ  $< 20$ ), or had low approximate read depth (below 10% of the mean coverage for the sample excluding the rDNA locus and telomeric regions). For every remaining variable site, the output from GATK enabled us to calculate the empirical allele frequencies of the reference base (designated *f<sub>A</sub>*) and the variant (designated *f<sub>B</sub>* and referred as alternative allele frequencies), which sum to 1. Empirical allele frequencies were calculated at each variable site on each strain by dividing the allelic depth of the variant by the approximate read depth observed at that site (AD and DP fields in GATK’s HaplotypeCaller). Each variant remaining after the filtering steps and having an *f<sub>B</sub>*  $> 0.15$  is considered to be SNPs. Accordingly, variable sites were only

used for further analysis when having a SNP. If  $f_A \geq 0.85$  the strain was called homozygous for the reference base (AA) at this variable site, and if  $f_B \geq 0.85$  it was called homozygous for the alternative base (BB). Sites with intermediate allele frequencies ( $0.15 < f_A < 0.85$ ) were called heterozygous (AB). Using these thresholds, a small number of SNPs were called in some strains later shown to be haploid. These apparent SNPs clustered in sub-telomeric regions known to contain repetitive DNA and are likely to be technical artefacts due to misaligning of reads to different repeats or to copy number variants. Note that these calls were only made for variable sites; the genomes also contain a much larger number of invariant sites that are considered identical among all strains and are therefore AA.

Nucleotide diversity ( $\pi$ ) and average SNP density were calculated with VariScan v 2.0.3 (Hutter et al., 2006) using a non-overlapping window size of 1 kb along all chromosomes. In the case of heterozygous variants, only the variant with highest  $f_B$  was used. Sites considered for the analysis of the 14 strains were required to show variation in a minimum number of 4 strains. SnpEff v 4.3s (Cingolani et al., 2012) was used to produce summary statistics and annotate the SNPs using the public NBRC1777 genome annotation as a reference (Inokuma et al., 2015).

## Phylogenetic Analysis

Because the data consisted of a mixture of strains with different ploidies, we developed a custom method for phylogenetic analysis of haplotypes. This method is based on a window approach similar to our previous development for an interspecies hybrid (Schröder et al., 2016). Homozygosity and heterozygosity were first assessed in 1 kb windows of the genome of each diploid strain. For each 1 kb window in each strain, the total numbers of variable sites that were called with each genotype (#AA, #BB, and #AB) in the window were calculated. The whole window was then classified as either heterozygous for the two haplotypes if  $\#AB \geq 3$ , or homozygous otherwise. Homozygous windows were then classified as either homozygous for the alternative haplotype if  $\#BB \geq 9$ , or homozygous for the reference haplotype otherwise. The cut-off values of 3 and 9 were chosen based on analysis of the distributions of window frequencies, using strain L01 as a test case (Figure S1). Each 1 kb window of the genome was only used if it was heterozygous in all five diploid strains, and the regions with aberrant allele frequencies in NBRC0272 were excluded (chromosome 6). Concatenated nucleotide sequences of the shared heterozygous windows were extracted from the NBRC1777 genome and used to infer the sequences of alleles in each strain, depending on its ploidy.

For haploid strains, the variant was used to replace the reference base at the corresponding position of the variable site in the concatenated sequence. For diploid strains, two different putative A and B alleles were first generated and used as a template for base replacement depending on the type of variable site. For homozygous alternative (BB) sites, the variant was used for replacement in both A and B alleles. For heterozygous (AB) variable sites, the variant was only used for replacement in the

B allele. For triploid strains, putative alleles 1, 2, and 3 were first generated and used as a template for base replacement depending on the types of variable sites and their allele frequencies. For homozygous alternative (BB) sites, the variant was used for replacement in all three putative alleles. For heterozygous (AB) sites, the variant was used for replacement in both alleles 2 and 3 if  $f_B > 0.6$ , or for replacement only in allele 3 if  $f_B < 0.4$ . The phylogenetic tree of the inferred and concatenated haplotype sequences was generated using PhyML v. 3.1 (Guindon et al., 2010) selecting for the best of NNI and SPR methods, using five random starts, and with empirical estimation of base frequencies and proportions of invariable sites (parameters: –search BEST –rand\_start –n\_rand\_starts 5 –f e –v e). The tree was visualized using FigTree v. 1.4.3 (<http://tree.bio.ed.ac.uk/software/figtree/>).

## Data Availability

The Illumina sequences generated and analyzed for this study can be found in the NCBI Sequence Read Archive under the accession number SRP128575, strain-specific accessions are provided in Table 1.

## RESULTS

### *K. marxianus* Displays a High Level of Genomic Variation

The genome sequences of 14 strains of *K. marxianus* were analysed for SNPs to determine the extent of variability in this species. The strains selected for analysis included the five strains with published whole genome sequences (at the time of the study) and 9 other strains from different collections (Table 1). The five sequenced strains (NBRC1777, CBS6556, UFV3, DMKU3-1042, and UFS-Y2791) have been phenotypically analysed to different extents by a number of research teams and are of interest for biotechnological applications (Nonklang et al., 2008; Fonseca et al., 2013; Costa et al., 2014; Schabort et al., 2016; Nambu-Nishida et al., 2017). This is also the case for the previously unsequenced strains CBS397, NBRC0272, NBRC0288, and NBRC0617, which were obtained from national culture collections (Lane et al., 2011; Foukis et al., 2012; Yurimizu et al., 2013). The additional five strains (L01–L05) are from the in-house culture collection of the Lallemand company and there are no published data available for these strains. The original source of isolation is known for 13/14 strains and is almost evenly divided between dairy and non-dairy environments. The genome sequences of the 9 previously un-sequenced strains were obtained as described in methods and thus all 14 genomes could be analysed and compared. It should be noted that although all 14 strains were sequenced using Illumina technology, this was performed by different laboratories using a diversity of sequencing strategies and thus the depth of coverage is quite variable (Table 2).

We used the genome sequence of strain NBRC1777, which was assembled into eight complete chromosomes using Pacific Biosciences technology, as the reference for SNP analysis (Inokuma et al., 2015). GATK software was used to identify sequence variants present in the Illumina reads from all 14 strains

relative to this reference. Variable sites were defined as the set of sites in the genome that contain a non-reference base in at least one of the 14 strains. For each variable site in each strain, the empirical allele frequency in the Illumina reads from that strain was calculated (see Methods). Depending on frequency, the variant was classified as homozygous SNP ( $f_B \geq 0.85$ ), or heterozygous SNP ( $0.15 < f_B < 0.85$ ). Variants appearing at frequencies below 0.15 were assumed to be due to sequencing errors and were ignored. Only 249 SNPs were identified in the Illumina data from the reference NBRC1777, confirming that the reference sequence is accurate and this strain is haploid. All other strains contained more than 30,000 SNPs relative to the reference (Table 3). The highest number of SNPs is in strain UFS-Y2791 (Schabert et al., 2016) and corresponds to 3.0% nucleotide sequence divergence in the 10.9 Mb genome. From the numbers of heterozygous and homozygous (non-reference) SNPs, the other strains fall into three groups: one group with low numbers of heterozygous SNPs ( $<2000$ ), one group with more heterozygous than homozygous SNPs, and one group with fewer heterozygous than homozygous SNPs. Below, we show that these three groups are haploid, diploid, and triploid strains, respectively.

In total, 667,472 variable sites, comprising 597,466 SNPs, and 70,006 indels were found. Indels were not analysed further, and, after filtering (see methods) a subset of 571,339 SNPs was retained for analysis. SNP diversity in *K. marxianus* is relatively high, with average pairwise difference between strains ( $\pi$ ) of  $12 \times 10^{-3}$ . The average density of SNPs in *K. marxianus* is 7.6 SNPs/kb in coding regions and 11.1 SNPs/kb in intergenic regions (Table S1). Of the 359,354 variants located within the coding regions of genes, 71.6% were predicted to be silent, 28.1% missense, and 0.2% (745)

nonsense mutations, when compared against the NBRC1777 annotation (Inokuma et al., 2015).

## K. marxianus Isolates Show Different Ploidy States

The distribution of allele frequencies for each strain was assessed using a graphical method similar to that used in a recent study in *S. cerevisiae* (Zhu et al., 2016). Histograms (Figure 1A), of the distribution of allele frequencies at all the variable sites in the genome created three sets of strains that corresponded to those identified on the basis of the numbers of heterozygous and homozygous SNPs (Table 3). The five strains with the highest numbers of heterozygous SNPs in (L02, L01, CBS397, NBRC0288, and NBRC0272) all show a symmetrical peak of allele frequencies centred on 0.5, suggesting that they are diploid. The three strains with high numbers of both homozygous and heterozygous SNPs (NBRC0617, L03, and UFV-3) all show bimodal distributions, with peaks at 0.33 and 0.66 for the frequency of the variant, suggesting that they are triploid. In contrast, in the six strains UFS-Y2791, DMKU3-1042, L05, L04, CBS6556, and NBRC1777, only a low number of sites were designated as heterozygous, and these sites show little pattern in their frequency distributions. These data are most consistent with a haploid genome.

To provide an independent measurement of ploidy, DNA content in each strain was measured by flow cytometry (Figure 1C). Each analysed strain shows a bimodal distribution of DNA content, corresponding to the G1 and G2 phases of the cell cycle. For the haploid strains, the DNA content is  $1n$  (in G1 phase) and  $2n$  (in G2 phase), where  $n$  is the DNA content of the haploid genome. For the diploids, it is  $2n$  and  $4n$ , and for the triploids it is  $3n$  and  $6n$ . The patterns observed were consistent with the designation based on allele frequencies and confirmed that this set of strains was comprised of 6 haploid, 5 diploid and 3 triploid strains.

## Common Patterns of Loss of Heterozygosity in Diploid Strains

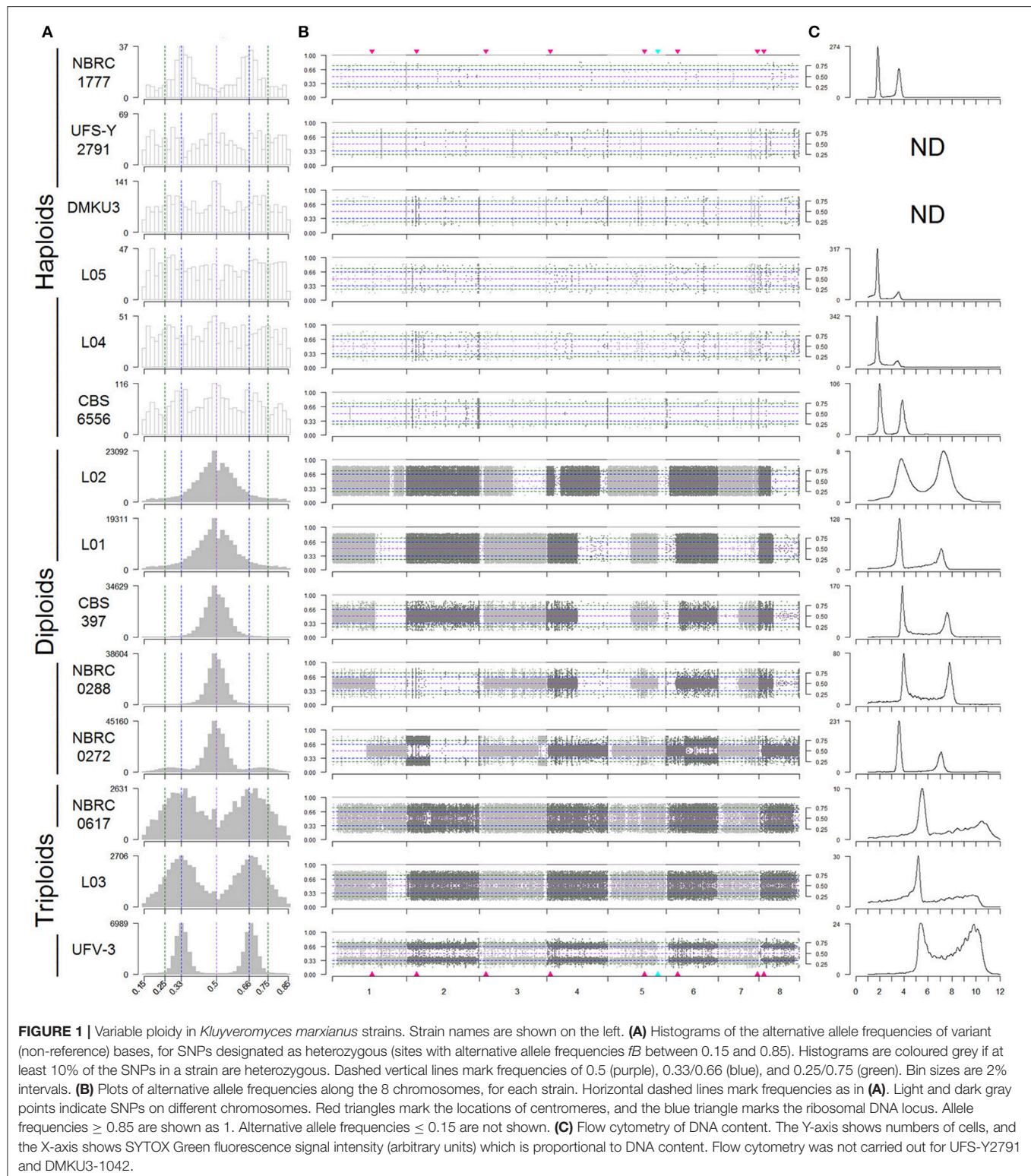
For each SNP, its allele frequency vs. its chromosomal location in the reference assembly was plotted to determine whether heterozygosity was uniformly distributed (Figure 1B). For haploid strains, the low level of variation does not show any particular pattern, whereas in all five diploid strains, large regions of the genome with loss of heterozygosity (LOH) are apparent. In heterozygous regions of the genome, allele frequency should be distributed about 0.5 but there are regions where no variability is seen. These predominantly white areas in the plots indicate stretches of chromosome that are homozygous in that strain (Figure 1B). For example, strain L02 is heterozygous through most of its genome but shows homozygosity on the right half of chromosome 3 and over most of chromosome 8. In the diploids, most chromosomes are heterozygous over at least some of their length, but chromosome 2 in NBRC0288 and chromosome 7 in L01 are essentially completely homozygous. Three diploid strains L01, CBS397 and NBRC0288 exhibit almost identical extents of partial LOH on chromosomes 1, 3 (left

**TABLE 3 |** SNPs identified in 14 *K. marxianus* strains.

Strain	Heterozygous SNPs*	Homozygous SNPs†	Total SNPs
<b>HAPLOID STRAINS</b>			
NBRC1777	248	1	249
L05	611	33,326	33,937
L04	726	35,138	35,864
CBS6556	1,439	40,922	42,361
DMKU3-1042	1,647	39,648	41,295
UFS-Y2791	714	325,190	325,904
<b>DIPLOID STRAINS</b>			
L02	115,648	42,561	158,209
L01	96,202	49,325	145,527
CBS397	110,132	52,415	162,547
NBRC0288	96,304	60,554	156,858
NBRC0272	147,241	40,085	187,326
<b>TRIPLOID STRAINS</b>			
NBRC0617	29,010	145,016	174,026
L03	26,388	145,856	172,244
UFV-3	27,347	177,365	204,712

\*Number of sites at which a variant (non-reference base) was present, at a frequency in the reads between 0.15 and 0.85. †Number of sites at which a variant (non-reference base) was present, at a frequency  $\geq 0.85$  in the reads.





end), 4, 5, 6, and 8 (**Figure 1B**), but different patterns on chromosomes 2 and 7. On chromosome 6, the region of LOH is slightly larger in CBS397 than in L01 and NBRC0288 (it crosses the centromere only in CBS397). Strain L02 shares

two LOH boundaries with this group of three strains, on chromosome 5 (the boundary occurs at the rDNA locus) and chromosome 3 (left end). Only one region of LOH is visible in our triploids, on chromosome 1 in strain L03. This LOH



region occupies ~4% of the L03 genome, whereas in the diploid strains the LOH regions total 25–51% of the genome by length.

## Copy Number Variation and Partial Aneuploidy in Multiple Strains

Read coverage in 10 kb windows was determined across the genome of each strain to investigate whether any of the strains displayed aneuploidy (**Figure 2**). In this analysis, a value of zero indicates no variation between expected and actual numbers of reads, whereas higher or lower numbers could indicate DNA duplications or deletions. Although the SNP and flow cytometry results (**Figure 1**) indicated that there are three groups of strains with genomes that are primarily haploid, diploid, and triploid, there is also evidence in multiple strains of partial aneuploidy, segmental duplications, or deletions that alter the copy number of some parts of the genome. These possible aneuploidies do not correspond to the regions of LOH described in **Figure 1**, indicating that these are distinct phenomena.

The clearest case of aneuploidy is in strain NBRC0617, which has an extra copy of chromosome 7 (**Figure 2**). The analysis of ratios shows that reads from chromosome 7 are present at 1.18x the expected frequency ( $\log_2$  value 0.249) in this strain. Since NBRC0617 is primarily triploid, a fourth copy of a chromosome should increase the read coverage on that chromosome by  $\sim 4/3 = 1.33$ -fold (cyan line in **Figure 2**) relative to the genome average, and many of the 10-kb windows in chromosome 7 are not significantly different from this value (**Figure S2A**). Furthermore, the distribution of allele frequency values for SNPs on chromosome 7 of NBRC0617 shows a peak at 0.25 (**Figure S2B**), which is consistent with the presence of four copies of the chromosome, and contrasts with the peaks at 0.33 and 0.66 that are seen when the whole genome of NBRC0617 is considered (**Figure 1A**). NBRC0617 also shows increased copy number of a circa 150 kb segment of chromosome 2 (coordinates 420–570 kb) (**Figure 2**). Within this segment, allele frequencies of 0.25 and 0.75 are visible (**Figure S2B**), indicating that it is present in 4 copies. With the current data, we are unable to determine the precise structure of the chromosomal rearrangement that increased the segment's copy number. The region immediately to the left of it (0–420 kb) may be present at a reduced copy number.

Some intriguing possible examples of copy number variation are seen in strain NBRC0272, which was designated as a diploid based on its overall allele frequency and flow cytometry patterns. Examination of allele frequencies indicates that there are three genomic regions (in chromosomes 2, 3, and 6) present at higher copy numbers (**Figure 1B**). This can be seen in more detail in the allele frequency plots of those chromosomes for this strain (**Figure S3**). Chromosomes 2 (left end) and 3 (right end) contain SNPs with 0.25/0.75 allele frequencies, indicating the presence of four copies. Chromosome 6 (right end) contains SNPs with allele frequency peaks close to 0.33/0.66, indicating three copies, and contrasting with the peaks at 0.5 on the left part of this chromosome. Since the flow cytometry (**Figure 1C**) shows that the total DNA content of NBRC0272 is close to diploid, it is concluded that the extra copies of these three chromosomal

regions must be the result of segmental duplications and not extra copies of the whole chromosome. Puzzlingly, however, these putative segmental duplications were not apparent in the coverage plot of NBRC0272 (**Figure 2**).

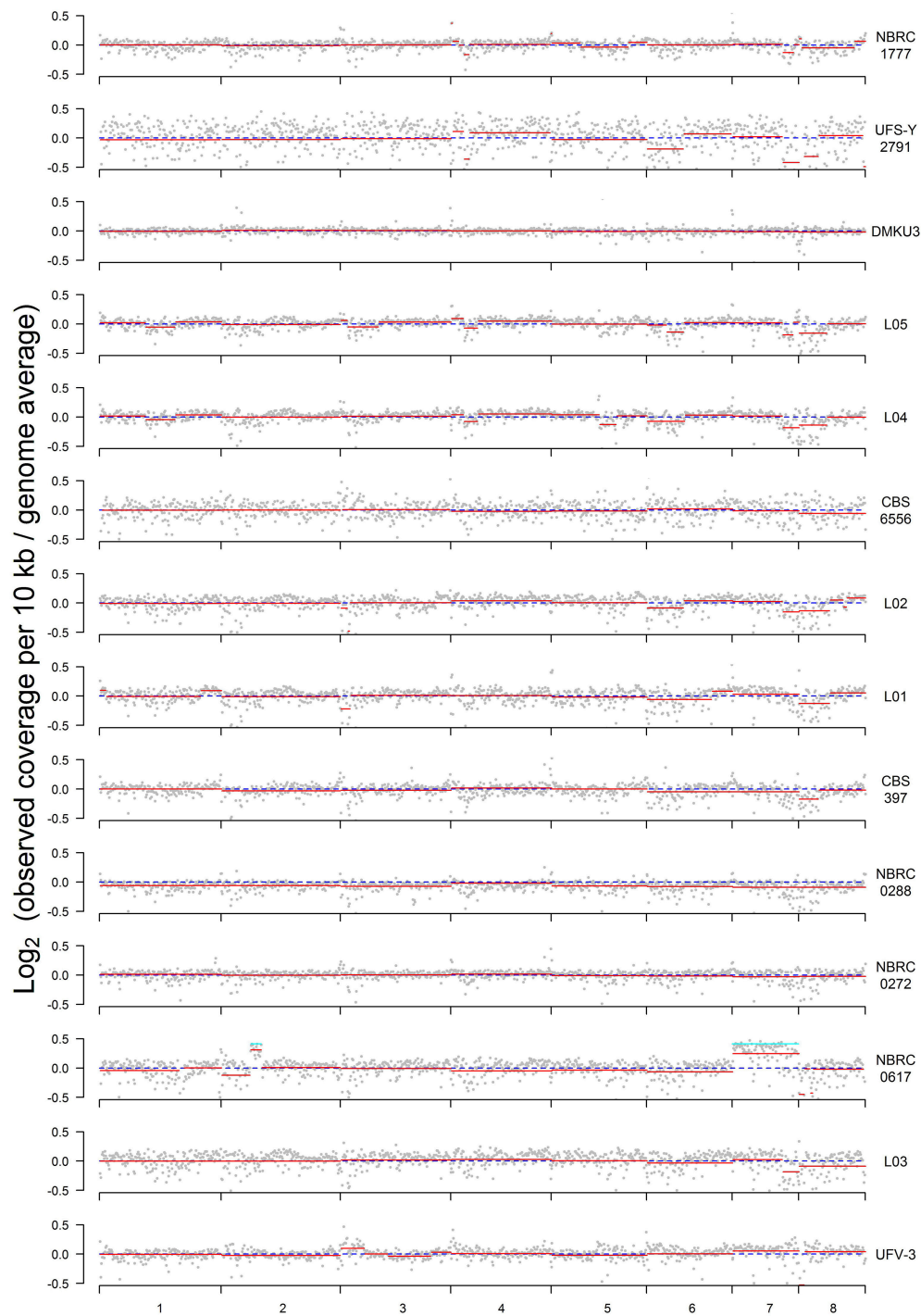
## Phylogenetic Analysis Separates Dairy From Non-dairy Haplotypes

The presence of strains with different ploidy in the dataset presents a problem for phylogenetic analysis. In studies on diploid eukaryotes such as mammals, the standard approaches for constructing phylogenies of individuals from SNP data either exclude all heterozygous sites, or randomly choose one of the alleles at these sites (Lischer et al., 2014). In our preliminary analyses of the *K. marxianus* data, it was noticed that in the diploid strains, one allele was often very similar to the NBRC1777 reference sequence, but the other allele was considerably different. We were therefore motivated to construct a phylogenetic tree of the *K. marxianus* strains that kept the alleles separate—namely, a tree of haplotypes rather than a tree of strains.

To make a tree of haplotypes, we used a method previously developed to investigate the pathogenic yeast *Candida orthopsilosis*, which is an interspecies hybrid (Schröder et al., 2016). In each of the 5 diploid *K. marxianus* strains, each region of the genome was classified as either heterozygous, homozygous for the “A” haplotype (the haplotype more similar to the NBRC1777 reference), or homozygous for the “B” haplotype (the haplotype less similar to the reference) (**Figure 3**; see Methods for details). Approximately 18% of the genome was heterozygous in all 5 diploid strains, and we then extracted the sequences of the “A” and “B” haplotypes from these regions in each diploid. The aberrant (trisomic or tetrasomic) regions of the NBRC0272 genome were excluded from this dataset. A similar process was used to estimate the sequences of the three haplotypes present in each of the three triploid strains in these regions (see Materials and Methods). In the phylogenetic tree of haplotypes then generated, each haploid strain appears once, each diploid strain appears twice, and each triploid strain appears three times (**Figure 4**). Three clades are evident, but Clade 3 contains only the haploid strain UFS-Y2791. Despite the high divergence of UFS-Y2791 from the other strains, a phylogenetic tree using *K. lactis* and *Lachancea thermotolerans* as outgroups confirms that it is indeed a strain of *K. marxianus* (**Figure 4**, inset). All other *K. marxianus* haplotypes lie in Clades 1 and 2. Clade 1 contains all the haploid strains except the outlier UFS-Y2791, and the “A” haplotypes of each of the five diploid strains. Clade 2 contains the “B” haplotypes of the diploid strains, and all three haplotypes of the triploid strains.

## Lactose Consumption Phenotypes and LAC12 Genotypes

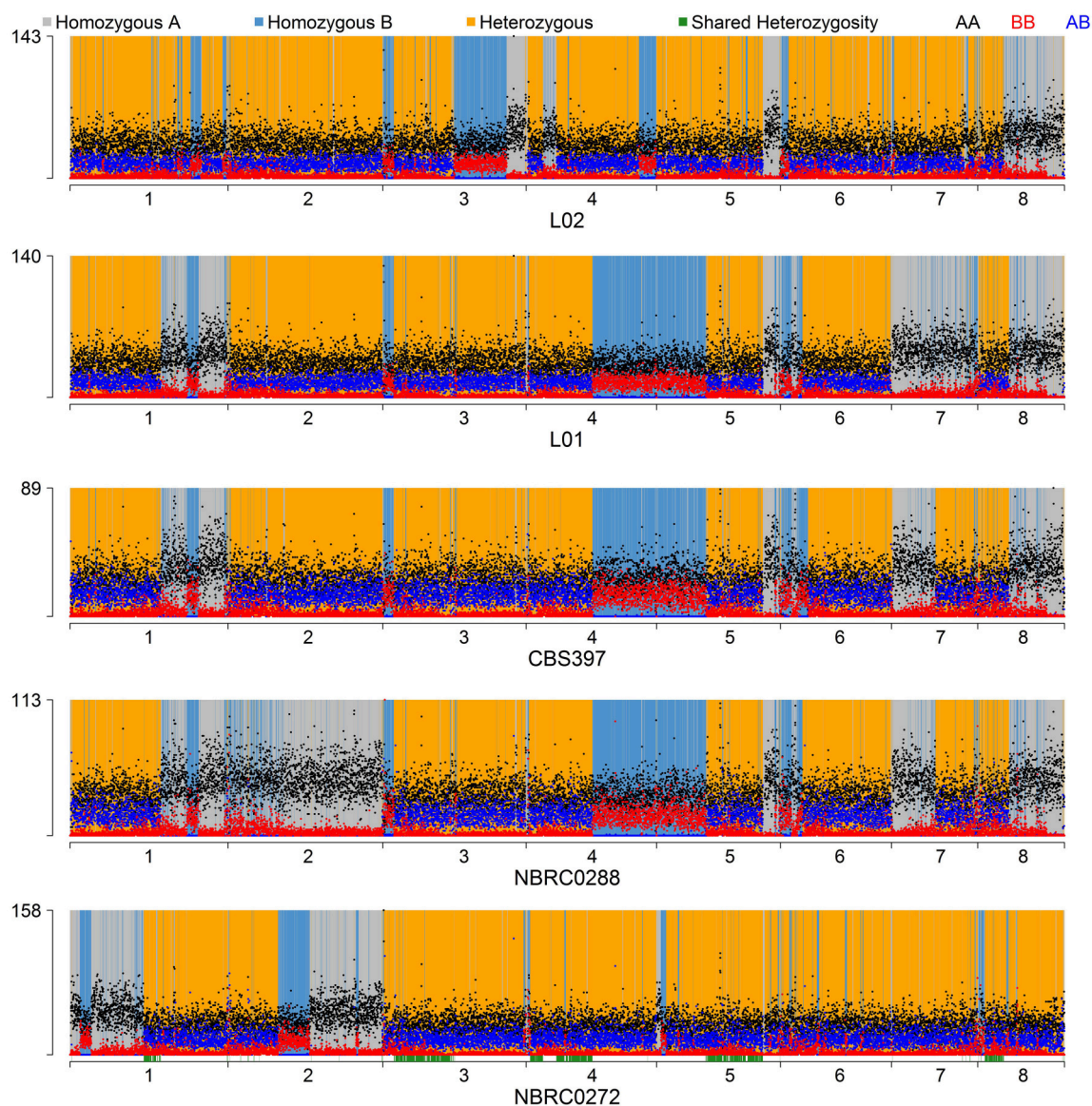
When the environments from which the strains were isolated are considered, an unexpected relationship is apparent between environment, ploidy and clade (**Figure 4**). The six strains from “dairy” environments (and strain NBRC0288, from an unknown source) are all either diploid or triploid whereas, with the



**FIGURE 2 |** Plots of sequence coverage in each strain. The Y-axis is  $\log_2$  of the ratio between the observed and expected coverage, for 10-kb windows through the genome; a value of zero (dashed blue line) indicates no difference. Expected coverage is based on the average in the whole genome. Red lines show the segmental means for consecutive 10-kb windows calculated using the Bioconductor package DNACopy. Cyan lines for NBRC0617 indicate the value expected for the 1.33-fold increase in coverage that would result from a fourth copy of a region in a triploid.

exception of strain NBRC0272 (isolated from miso), all the strains from “non-dairy” environments are haploid (Table 1). Since the use of lactose as a sugar source is considered important

for growth of dairy yeasts, the capacity of the strains to grow on lactose was assessed to determine whether a similar pattern would emerge (Figure 5). Indeed, none of the tested haploid strains

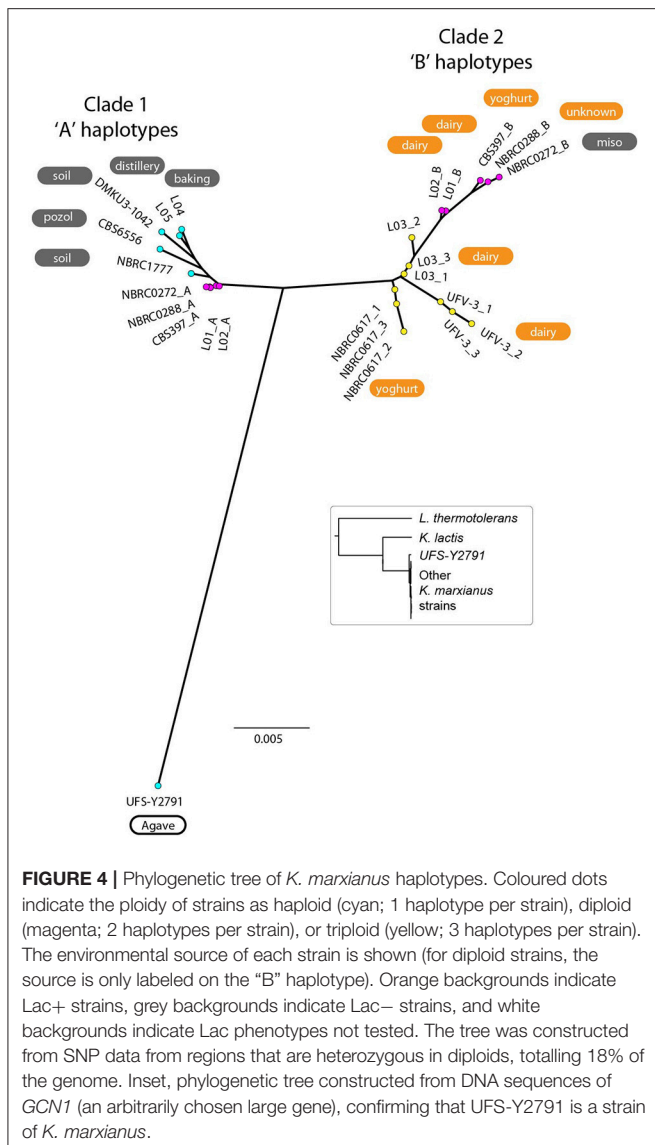


**FIGURE 3 |** Haplotype assignment in diploid strains. For each 1-kb window through the genome, dots show the number of variable sites in the window with genotypes AA (black dots), BB (red dots), or AB (blue dots). Each window was then classified as either heterozygous (yellow background), homozygous for the *B* haplotype (blue background), or homozygous for the reference *A* haplotype (grey background) following rules as described in Methods and **Figure S1**. The green bars along the bottom axis show the regions (18%) that were heterozygous in all diploids and used for the haplotype phylogeny analysis.

used lactose, whereas all the diploid and triploid strains were Lac<sup>+</sup>, again with the exception of NBRC0272, which is diploid but Lac<sup>−</sup>. Although DMKU3-1042 was not available for this study, our previous work has demonstrated that it is Lac<sup>−</sup> (Varela et al., 2017). We also previously established that the variable ability of *K. marxianus* strains to consume lactose is explained by polymorphism of a single gene, the lactose transporter *LAC12*, and that functional and non-functional (in terms of lactose transport) alleles of this gene differ by 13 key amino acid substitutions (Varela et al., 2017). BLAST searches against *de novo* assemblies of the genomes showed that the key polymorphisms

in all the Lac<sup>+</sup> strains in **Figure 5** match were an exact match to the functional *LAC12*<sup>+</sup> allele, except for NBRC0617 which matched in 11 positions. Similarly, all the Lac<sup>−</sup> strains exactly matched the non-functional *LAC12* allele, except for NBRC0272, which diverged at a single amino acid (**Figure S4**). None of the diploid strains is a *LAC12*<sup>+/−</sup> heterozygote, due to LOH at the left end of chromosome 3 where *LAC12* is located (the gene is only 15 kb from the telomere). Thus, although all five diploids are AB heterozygous for most of chromosome 3, the four Lac<sup>+</sup> diploids are “BB” homozygous and the Lac<sup>−</sup> diploid strain NBRC0272 is “AA” homozygous in this region (**Figure 3**). In





addition to the polymorphisms associated with a non-functional allele, the *LAC12* gene in NBRC0272 contains an internal stop codon (Figure S4).

## DISCUSSION

### Ploidy in *K. marxianus* Distinguishes Dairy and Non-dairy Strains

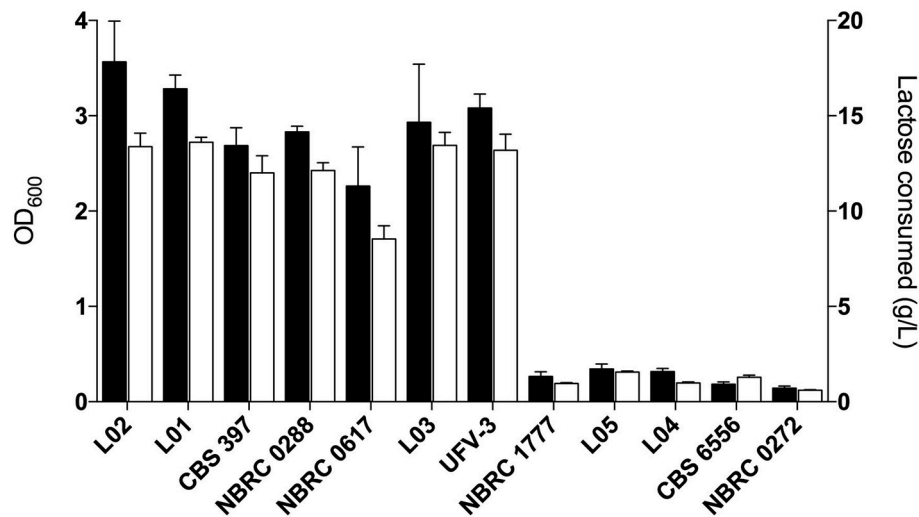
Although this study relied on a relatively small set of strains (14), it delivered some remarkable insights into the life-cycle of *K. marxianus*. Our previous report that natural isolates of this yeast can be either haploid or diploid (Lane et al., 2011) was confirmed and then extended by the discovery of three triploid strains. The sample size is too small to draw statistical conclusions but it can be said that haploids, diploids, and triploids were present at roughly equal frequencies (43, 36, and 21% respectively) in this set. This appears to contrast

with *K. lactis*, which is considered to be haploid, though it must be borne in mind that there are as yet no published production-level studies with that yeast. Variable ploidy is not uncommon in yeasts; for example, among 144 mainly clinical isolates of *S. cerevisiae*, the basal ploidy levels (i.e., ignoring aneuploid chromosomes) were haploid (11%), diploid (57%), triploid (16%), and tetraploid (16%) (Zhu et al., 2016).

The most striking finding was that all the isolates from a dairy environment were either diploid or triploid, whereas non-dairy isolates were haploid. Furthermore, it was possible to distinguish two genomic haplotypes, described here as “A” and “B” that mapped 13/14 strains into distinct clades (Clades 1 and 2). The 14th strain, UFS-Y2791 may represent a third clade. All the dairy isolates contained at least one of the B haplotype genomes, suggesting that this is a dairy-niche associated genome. In the case of the three triploid strains, the B genome was represented three times, whereas the diploid strains contained one A haplotype genome and one B haplotype genome. The sequence divergence (~2%) between the “A” and “B” haplotypes indicates that the diploid dairy strains were probably formed by mating between haploid representatives of Clades 1 and 2, as opposed to any other mechanism of ploidy change. Nevertheless, we were unable to examine *MAT* locus genotypes because the sequence assemblies are too fragmented in the *MAT/HML/HMR* regions. The phylogeny of the haplotypes indicates that the diploid dairy strains were formed by at least two independent matings between parents from the A and B clades, and that the A parents in these matings were very closely related to each other (much more so than the B parents). Triploids may have arisen by self-mating of B-haplotype (clade 2) strains, with one scenario being mating of a BB diploid with a B haploid to form a triploid; other routes to a triploid are also possible. It is implicit in these scenarios that B-haplotype haploid strains should also exist, though none were found in the current study. It is also notable that in our recent study developing an MLST method for *K. marxianus*, all 57 strains that were listed as coming from (6 different) dairy environments were heterozygous and therefore presumably diploid (Tittarelli et al., 2018). In fact, in that study, only 13/83 strains were homozygous in the regions included in the MLST. It is noted that one well-studied strain in the literature is *K. marxianus* CBS397 and this study and those of Fasoli et al. (2016) and Tittarelli et al. (2018) show that this strain is diploid (Fasoli et al., 2016; Tittarelli et al., 2018), which contrasts with what appears to be a previous erroneous suggestion based on long range PCR of the *MAT* locus that it was haploid (Lane et al., 2011).

The data suggest that the B-haplotype is a dairy-associated genome. It was gratifying, therefore, to identify one locus in this haplotype that confers a growth advantage in milk, the *LAC12* gene. Our previous work identified positions in the *Lac12p* where the functional protein had one particular amino acid and the non-functional protein a different one (Varela et al., 2017). In six of the strains with a B-haplotype genome, there was an exact match to this functional sequence and in the seventh (NBRC0617), there was a match in 11 positions (Figure S4). Since all these strains grew on lactose as a sole sugar source, this now allows us to propose that the number of amino





**FIGURE 5 |** Growth and lactose consumption by *K. marxianus* strains. Growth in MM + 2 % lactose and lactose consumption are shown in black and white bars, respectively. Lactose consumption was calculated by subtracting the final and initial lactose concentrations in the medium.

acid positions that distinguish a functional and non-functional lactose-transporting Lac12p protein can be refined to these 11 amino acids though confirmation would require functional tests. The Lac<sup>−</sup> strain with the B-haplotype genome was NBRC0272, which is homozygous for the non-functional *LAC12* allele. This strain was isolated from a non-dairy environment (miso) and the most likely explanation is that it arose like the other diploids as a hybrid between an A and a B strain but since lactose transport was not required in its niche, it was possible for it to lose the B-haplotype *LAC12* allele through a LOH event at the left end of chromosome 3 whereas the other diploids lost the non-functional A-haplotype *LAC12* allele via a similar LOH event (Figure 4).

The situation in *K. marxianus* seems to resemble a pattern seen in *Saccharomyces* and *Zygosaccharomyces* species, where strains used in industrial processes or isolated from industrial environments are often polyploids or interspecies hybrids, whereas “natural” isolates (e.g., from non-anthropogenic environments) tend to be haploid or homozygous diploid (Hittinger, 2013; Suh et al., 2013; Wendland, 2014; Ortiz-Merino et al., 2017). This pattern is thought to reflect selection toward stress tolerance in the industrial environment, but toward maintenance of the ability to mate and sporulate in natural environments. If this is also the case with *K. marxianus*, it could be expected that the AB diploids display enhanced stress tolerance, at least over B-haplotype haploid strains in a dairy environment. Experiments to date have not succeeded in identifying any correlations between ploidy and stress tolerance (data not shown), but more studies that also include B-haplotype haploids are required to further address this question. As mentioned, B-haplotype strains have not yet been positively identified but there are Lac<sup>+</sup> candidates worth investigating, for example, *K. marxianus* NCYC1424, shown to be homozygous by Tittarelli et al. (2018), and *K. marxianus* NCYC1429, which appears to be haploid based on genetic crossing (Varela et al., 2017).

## Sequence Diversity, Aneuploidy, and LOH in *K. marxianus*

One of the aims of this study was to assess if the wide phenotypic diversity that has been observed in *K. marxianus* was reflected in its genome diversity. The large number of SNPs (>500 k) observed in the set of *K. marxianus* strains used in our study shows relatively high SNP diversity. We found an average pairwise difference ( $\pi$ ) of  $12 \times 10^{-3}$ , which is comparable with reported values from other yeasts; for example,  $4 \times 10^{-3}$  in *S. cerevisiae*,  $12 \times 10^{-3}$  in *S. uvarum*, and  $17 \times 10^{-3}$  in *L. kluyveri* (Peter and Schacherer, 2016).

Previous studies with *S. cerevisiae* isolates showed variable ploidy (from 1 to 4 copies of the genome), aneuploidy (unequal copy numbers of different chromosomes), or variation in the copy number of segments of chromosomes (Hose et al., 2015; Strope et al., 2015; Zhu et al., 2016). Similar to *S. cerevisiae* but unlike *L. kluyveri* (Friedrich et al., 2015), all these phenomena were observed in this study of *K. marxianus*. The most unambiguous example of aneuploidy in *K. marxianus* is the presence of an extra copy of chromosome 7 in NBRC0617, but, as described in the results, there are multiple other likely cases of aneuploidies or copy number variation that would need to be investigated in more detail.

There are also quite extensive regions of LOH in the diploid strains (25–51% by genome length) but not in the triploids. LOH arises when the genome homogenises in a region and it is expected to be rarer in triploid strains, because it will only be apparent if all three copies of a genomic region were homogenised. The shared patterns of LOH in different diploid *K. marxianus* isolates was unexpected. This observation could indicate that these strains are closely related and are mitotic descendants of a recent diploid common ancestor that had already lost heterozygosity in the shared regions. Alternatively, it could indicate that this species mostly reproduces by mitosis and rarely goes through meiosis and sporulation, at least in

the dairy environment. Nonetheless, the divergence between the *K. marxianus* clades is low enough that it would not be expected to cause problems in meiosis. In *S. paradoxus*, crosses between strains with sequence divergence of up to 4.6% can still produce viable gametes, as long as the genomes are collinear (Liti et al., 2006). It should also be considered that dairy is probably not the original niche for *K. marxianus* and the strains that we studied were most likely selected during a fermentation process. This could have specifically selected hybrids between A and B clade strains, and may also promote LOH. Other than the preservation of the functional lactose transporting allele of *LAC12* (B-haplotype), there was not an obvious preference for either the A-haplotype or the B-haplotype during LOH events (Figure 3). Because lactose utilisation confers a benefit during growth in milk, one can speculate that the LOH of the region containing the functional *LAC12* gene is an adaptive response. It is not possible to say, however, whether or not other selective pressures played a role in determining the overall patterns of LOH.

## Implications for Biotechnology

This study focused on a small set of strains of biotechnological interest and therefore may not be fully representative of the species diversity. Indeed, one strain (UFS-Y2791) was far more diverse than the others, suggesting that there is further diversity to be accessed. Given that UFS-Y2791 was isolated from agave juice (in South Africa), it will be interesting to see whether strains associated with tequila/mezcal fermentation (also from agave) in Mexico show any relationship to this strain. The divergence between strains used, either deliberately or traditionally, in the food biotechnology sector is very significant in comparison to those isolated from “natural” environments. Perhaps the natural state of *K. marxianus* is haploid like its sister *K. lactis*, and diploids only arise after biotechnological selection. It is possible that diploids will have advantages, though, other than for lactose utilisation, these are not yet apparent. Haploid strains are much easier to engineer and manipulate so, for most biotechnological applications, it may be preferable to choose Clade 1 (A haplotype) strains. Nonetheless, divergent alleles in Clade 2 (B haplotype) may also be functionally important (for example *LAC12*) so this will still need to be considered in future studies.

## AUTHOR CONTRIBUTIONS

RO-M and JV: contributed equally to the paper, they carried out the bulk of the experimental and bioinformatic analysis and wrote the manuscript; AC: contributed with strain handling and sample preparation for DNA sequencing; CW, NK, J-MG, WdS, and HH: sequenced *K. marxianus* strains for the study; KW and JM: conceived the study, supervised the research, analysed and interpreted data and contributed to writing the manuscript.

## REFERENCES

Arrizon, J., Morel, S., Gschaedler, A., and Monsan, P. (2012). Fructanase and fructosyltransferase activity of non-Saccharomyces yeasts isolated

## ACKNOWLEDGMENTS

RO-M and JV were supported by the YEASTCELL Marie Curie ITN project which received funding from the People Programme (Marie Curie Actions) of the European Union's Seventh Framework Programme FP7/2007-2013/ under REA grant agreement n° 606795. RO-M was also partially supported by CONACyT, Mexico (fellowship number 440667). AC was supported by Science Foundation Ireland (13/IA/1910). This study was supported in part by the Adaptable and the Advanced Low Carbon Technology R&D Program (JST, Japan). We thank Kevin Byrne for computational support and colleagues listed in Table 1 for providing raw Illumina source files of genome sequences.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2018.00094/full#supplementary-material>

**Figure S1** | Method of classifying 1-kb genomic windows into haplotypes. The cutoff values chosen are marked by triangles in all panels (windows with  $\geq 3$  AB SNPs were classified as AB heterozygous windows; windows with  $\geq 9$  BB SNPs were classified as BB heterozygous windows; other windows were classified as AA homozygous windows). The data plotted are for the diploid strain L01. **(A)** Histogram of numbers of SNPs of each type, in all 1-kb windows in the genome. **(B)** Heatmap showing the distribution of numbers of AB and BB SNPs per window. Cells in the matrix show numbers of windows. The cutoff values were chosen to coincide with minima on the two axes. **(C)** Distribution of numbers of BB SNPs in windows that have zero AB SNPs.

**Figure S2** | Allele frequencies and sequence coverage in three strains on **(A)** chromosome 7 and **(B)** chromosome 2. The left and middle panels show distributions of allele frequency on each chromosome, as in Figure 1. The centromere is marked by a vertical red line. The right panels show  $\log_2$  ratios between observed and expected sequence coverage, as in Figure 2. Cyan lines for NBRC0617 indicate the value expected for the 1.33-fold increase in coverage that would result from a fourth copy of a region in a triploid. The vertical orange lines in **(B)** mark the 150 kb region from with increased coverage in NBRC0617 (coordinates 420–570 kb).

**Figure S3** | Allele frequencies on each chromosome of NBRC0272. Details are as in Figures 1A,B. In the plots on the right, red vertical lines mark centromeres and blue lines mark the rDNA array.

**Figure S4** | Multiple alignment of Lac12p sequences from the strains used in this study. Amino acid sequences shown were derived from the haplotypes of each strain. Haplotypes from diploid and triploid strains are denoted by letters and numbers, respectively. The residues marked in color are those associated with either functional or non-functional alleles. The positions marked in blue were previously part of the set that distinguished alleles but these are not conserved in all haplotype B Lac12p. Those in red are those that still distinguish based on functionality and an additional differentiating AA at position 475 (pink shading) that was overlooked in the previous study is also marked. Thus, based on sequences currently available, there are 11 differentiating amino acids. The stop codon in the NBRC0272\_B sequence at position 139 is indicated with a green asterisk.

**Table S1** | SNP density.

from fermenting musts of Mezcal. *Bioresour. Technol.* 110, 560–565. doi: 10.1016/j.biortech.2012.01.112

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., et al. (2012). SPAdes: a new genome assembly algorithm and

- its applications to single-cell sequencing. *J. Comput. Biol.* 19, 455–477. doi: 10.1089/cmb.2012.0021
- Barsoum, E., Martinez, P., and Aström, S. U. (2010). Alpha3, a transposable element that promotes host sexual reproduction. *Genes Dev.* 24, 33–44. doi: 10.1101/gad.557310
- Belloch, C., Barrio, E., García, M. D., and Querol, A. (1998). Inter- and intraspecific chromosome pattern variation in the yeast genus *Kluyveromyces*. *Yeast* 14, 1341–1354.
- Beniwal, A., Saini, P., Kokkilgadda, A., and Vij, S. (2017). Physiological growth and galactose utilization by dairy yeast *Kluyveromyces marxianus* in mixed sugars and whey during fermentation. *3 Biotech* 7:349. doi: 10.1007/s13205-017-0985-1
- Booth, L. N., Tuch, B. B., and Johnson, A. D. (2010). Intercalation of a new tier of transcription regulation into an ancient circuit. *Nature* 468, 959–963. doi: 10.1038/nature09560
- Cingolani, P., Platts, A., Wang le, L., Coon, M., Nguyen, T., Wang, L., et al. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly* 6, 80–92. doi: 10.4161/fly.19695
- Coloretti, F., Chiavari, C., Luise, D., Tofalo, R., Fasoli, G., Suzzi, G., et al. (2017). Detection and identification of yeasts in natural whey starter for Parmigiano Reggiano cheese-making. *Int. Dairy J.* 66(Suppl. C), 13–17. doi: 10.1016/j.idairyj.2016.10.013
- Costa, D. A., de Souza, C. J., Costa, P. S., Rodrigues, M. Q., dos Santos, A. F., Lopes, M. R., et al. (2014). Physiological characterization of thermotolerant yeast for cellulosic ethanol production. *Appl. Microbiol. Biotechnol.* 98, 3829–3840. doi: 10.1007/s00253-014-5580-3
- Dias, O., Basso, T. O., Rocha, I., Ferreira, E. C., and Gombert, A. K. (2017). Quantitative physiology and elemental composition of *Kluyveromyces lactis* CBS 2359 during growth on glucose at different specific growth rates. *Antonie Van Leeuwenhoek* 111, 183–195. doi: 10.1007/s10482-017-0940-5
- Dias, O., Gombert, A. K., Ferreira, E. C., and Rocha, I. (2012). Genome-wide metabolic (re-) annotation of *Kluyveromyces lactis*. *BMC Genomics* 13:517. doi: 10.1186/1471-2164-13-517
- Dias, O., Pereira, R., Gombert, A. K., Ferreira, E. C., and Rocha, I. (2014). iOD907, the first genome-scale metabolic model for the milk yeast *Kluyveromyces lactis*. *Biotechnol. J.* 9, 776–790. doi: 10.1002/biot.201300242
- Diniz, R. H. S., Villada, J. C., Alvim, M. C. T., Vidigal, P. M. P., Vieira, N. M., Lamas-Maceiras, M., et al. (2017). Transcriptome analysis of the thermotolerant yeast *Kluyveromyces marxianus* CCT 7735 under ethanol stress. *Appl. Microbiol. Biotechnol.* 101, 6969–6980. doi: 10.1007/s00253-017-8432-0
- Fasoli, G., Barrio, E., Tofalo, R., Suzzi, G., and Belloch, C. (2016). Multilocus analysis reveals large genetic diversity in *Kluyveromyces marxianus* strains isolated from Parmigiano Reggiano and Pecorino di Farindola cheeses. *Int. J. Food Microbiol.* 233, 1–10. doi: 10.1016/j.ijfoodmicro.2016.05.028
- Fasoli, G., Tofalo, R., Lanciotti, R., Schirone, M., Patrignani, F., Perpetuini, G., et al. (2015). Chromosome arrangement, differentiation of growth kinetics and volatile molecule profiles in *Kluyveromyces marxianus* strains from Italian cheeses. *Int. J. Food Microbiol.* 214, 151–158. doi: 10.1016/j.ijfoodmicro.2015.08.001
- Fonseca, G. G., de Carvalho, N. M., and Gombert, A. K. (2013). Growth of the yeast *Kluyveromyces marxianus* CBS 6556 on different sugar combinations as sole carbon and energy source. *Appl. Microbiol. Biotechnol.* 97, 5055–5067. doi: 10.1007/s00253-013-4748-6
- Fonseca, G. G., Gombert, A. K., Heinzle, E., and Wittmann, C. (2007). Physiology of the yeast *Kluyveromyces marxianus* during batch and chemostat cultures with glucose as the sole carbon source. *FEMS Yeast Res.* 7, 422–435. doi: 10.1111/j.1567-1364.2006.00192.x
- Fonseca, G. G., Heinzle, E., Wittmann, C., and Gombert, A. K. (2008). The yeast *Kluyveromyces marxianus* and its biotechnological potential. *Appl. Microbiol. Biotechnol.* 79, 339–354. doi: 10.1007/s00253-008-1458-6
- Foukis, A., Stergiou, P. Y., Theodorou, L. G., Papagianni, M., and Papamichael, E. M. (2012). Purification, kinetic characterization and properties of a novel thermo-tolerant extracellular protease from *Kluyveromyces marxianus* IFO 0288 with potential biotechnological interest. *Bioresour. Technol.* 123, 214–220. doi: 10.1016/j.biortech.2012.06.090
- Friedrich, A., Jung, P., Reisser, C., Fischer, G., and Schacherer, J. (2015). Population genomics reveals chromosome-scale heterogeneous evolution in a protoploid yeast. *Mol. Biol. Evol.* 32, 184–192. doi: 10.1093/molbev/msu295
- Fukuhara, H. (2006). *Kluyveromyces lactis*- a retrospective. *FEMS Yeast Res.* 6, 323–324. doi: 10.1111/j.1567-1364.2005.00012.x
- Gethins, L., Rea, M. C., Stanton, C., Ross, R. P., Kilcawley, K., O'Sullivan, M., et al. (2016). Acquisition of the yeast *Kluyveromyces marxianus* from unpasteurised milk by a kefir grain enhances kefir quality. *FEMS Microbiol. Lett.* 363:fnw165. doi: 10.1093/femsle/fnw165
- Gombert, A. K., Madeira, J. V. Jr., Cerdán, M. E., and González-Siso, M. I. (2016). *Kluyveromyces marxianus* as a host for heterologous protein synthesis. *Appl. Microbiol. Biotechnol.* 100, 6193–6208. doi: 10.1007/s00253-016-7645-y
- Groeneveld, P., Stouthamer, A. H., and Westerhoff, H. V. (2009). Super life—how and why ‘cell selection’ leads to the fastest-growing eukaryote. *FEBS J.* 276, 254–270. doi: 10.1111/j.1742-4658.2008.06778.x
- Guimarães, P. M., Teixeira, J. A., and Domingues, L. (2010). Fermentation of lactose to bio-ethanol by yeasts as part of integrated solutions for the valorisation of cheese whey. *Biotechnol. Adv.* 28, 375–384. doi: 10.1016/j.biotechadv.2010.02.002
- Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* 59, 307–321. doi: 10.1093/sysbio/syq010
- Haase, S. B., and Reed, S. I. (2002). Improved flow cytometric analysis of the budding yeast cell cycle. *Cell Cycle* 1, 132–136. doi: 10.4161/cc.1.2.114
- Hittinger, C. T. (2013). *Saccharomyces* diversity and evolution: a budding model genus. *Trends Genet.* 29, 309–317. doi: 10.1016/j.tig.2013.01.002
- Hose, J., Yong, C. M., Sardi, M., Wang, Z., Newton, M. A., and Gasch, A. P. (2015). Dosage compensation can buffer copy-number variation in wild yeast. *Elife* 4:e05462. doi: 10.7554/eLife.05462
- Hughes, S. R., Qureshi, N., López-Núñez, J. C., Jones, M. A., Jarodsky, J. M., Galindo-Leva, L. Á., et al. (2017). Utilization of inulin-containing waste in industrial fermentations to produce biofuels and bio-based chemicals. *World J. Microbiol. Biotechnol.* 33:78. doi: 10.1007/s11274-017-2241-6
- Hutter, S., Vilella, A. J., and Rozas, J. (2006). Genome-wide DNA polymorphism analyses using VariScan. *BMC Bioinformatics* 7:409. doi: 10.1186/1471-2105-7-409
- Inokuma, K., Ishii, J., Hara, K. Y., Mochizuki, M., Hasunuma, T., and Kondo, A. (2015). Complete genome sequence of *Kluyveromyces marxianus* NBRC1777, a nonconventional thermotolerant yeast. *Genome Announc* 3:e00389-15. doi: 10.1128/genomeA.00389-15
- Jeong, H., Lee, D. H., Kim, S. H., Kim, H. J., Lee, K., Song, J. Y., et al. (2012). Genome sequence of the thermotolerant yeast *Kluyveromyces marxianus* var. *marxianus* KCTC 17555. *Eukaryot. Cell* 11, 1584–1585. doi: 10.1128/EC.00260-12
- Jiang, H., Lei, R., Ding, S. W., and Zhu, S. (2014). Skewer: a fast and accurate adapter trimmer for next-generation sequencing paired-end reads. *BMC Bioinformatics* 15:182. doi: 10.1186/1471-2105-15-182
- Kobayashi, Y., Sahara, T., Suzuki, T., Kamachi, S., Matsushika, A., Hoshino, T., et al. (2017). Genetic improvement of xylose metabolism by enhancing the expression of pentose phosphate pathway genes in *Saccharomyces cerevisiae* IR-2 for high-temperature ethanol production. *J. Ind. Microbiol. Biotechnol.* 44, 879–891. doi: 10.1007/s10295-017-1912-5
- Lachance, M.-A. (2011). “Chapter 35-*Kluyveromyces* van der Walt (1971) A2-Kurtzman, Cletus P” in *The Yeasts 5th Edn*, eds J. W. Fell and T. Boekhout (London: Elsevier), 471–481.
- Lane, M. M., Burke, N., Karreman, R., Wolfe, K. H., O’Byrne, C. P., and Morrissey, J. P. (2011). Physiological and metabolic diversity in the yeast *Kluyveromyces marxianus*. *Antonie Van Leeuwenhoek* 100, 507–519. doi: 10.1007/s10482-011-9606-x
- Lane, M. M., and Morrissey, J. P. (2010). *Kluyveromyces marxianus*: a yeast emerging from its sister’s shadow. *Fungal Biol. Rev.* 24, 17–26. doi: 10.1016/j.fbr.2010.01.001
- Lappe-Oliveras, P., Moreno-Terrazas, R., Arrizón-Gaviño, J., Herrera-Suárez, T., García-Mendoza, A., and Gschaedler-Mathis, A. (2008). Yeasts associated with the production of Mexican alcoholic nondistilled and distilled Agave beverages. *FEMS Yeast Res.* 8, 1037–1052. doi: 10.1111/j.1567-1364.2008.00430.x



- Lee, J. W., In, J. H., Park, J. B., Shin, J., Park, J. H., Sung, B. H., et al. (2017). Co-expression of two heterologous lactate dehydrogenases genes in *Kluyveromyces marxianus* for L-lactic acid production. *J. Biotechnol.* 241, 81–86. doi: 10.1016/j.jbiotec.2016.11.015
- Lertwattanasakul, N., Kosaka, T., Hosoyama, A., Suzuki, Y., Rodrussamee, N., Matsutani, M., et al. (2015). Genetic basis of the highly efficient yeast *Kluyveromyces marxianus*: complete genome sequence and transcriptome analyses. *Biotechnol. Biofuels* 8:47. doi: 10.1186/s13068-015-0227-x
- Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760. doi: 10.1093/bioinformatics/btp324
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352
- Lin, Y. J., Chang, J. J., Lin, H. Y., Thia, C., Kao, Y. Y., Huang, C. C., et al. (2017). Metabolic engineering a yeast to produce astaxanthin. *Bioresour. Technol.* 245(Pt A), 899–905. doi: 10.1016/j.biortech.2017.07.116
- Lischer, H. E., Excoffier, L., and Heckel, G. (2014). Ignoring heterozygous sites biases phylogenomic estimates of divergence times: implications for the evolutionary history of microtus voles. *Mol. Biol. Evol.* 31, 817–831. doi: 10.1093/molbev/mst271
- Liti, G., Barton, D. B., and Louis, E. J. (2006). Sequence diversity, reproductive isolation and species concepts in *Saccharomyces*. *Genetics* 174, 839–850. doi: 10.1534/genetics.106.062166
- Liti, G., Carter, D. M., Moses, A. M., Warringer, J., Parts, L., James, S. A., et al. (2009). Population genomics of domestic and wild yeasts. *Nature* 458, 337–341. doi: 10.1038/nature07743
- Morrissey, J. P., Etschmann, M. M., Schrader, J., and de Billerbeck, G. M. (2015). Cell factory applications of the yeast *Kluyveromyces marxianus* for the biotechnological production of natural flavour and fragrance molecules. *Yeast* 32, 3–16. doi: 10.1002/yea.3054
- Nambu-Nishida, Y., Nishida, K., Hasunuma, T., and Kondo, A. (2017). Development of a comprehensive set of tools for genome engineering in a cold- and thermo-tolerant *Kluyveromyces marxianus* yeast strain. *Sci. Rep.* 7:8993. doi: 10.1038/s41598-017-08356-5
- Nonklang, S., Abdel-Banat, B. M., Cha-aim, K., Moonjai, N., Hoshida, H., Limtong, S., et al. (2008). High-temperature ethanol fermentation and transformation with linear DNA in the thermotolerant yeast *Kluyveromyces marxianus* DMKU3-1042. *Appl. Environ. Microbiol.* 74, 7514–7521. doi: 10.1128/AEM.01854-08
- Nonklang, S., Ano, A., Abdel-Banat, B. M., Saito, Y., Hoshida, H., and Akada, R. (2009). Construction of flocculent *Kluyveromyces marxianus* strains suitable for high-temperature ethanol fermentation. *Biosci. Biotechnol. Biochem.* 73, 1090–1095. doi: 10.1271/bbb.80853
- Oliveira, S. S. S., Bello, M. L., Rodrigues, C. R., Azevedo, P. L., Ramos, M. C. K. V., Aquino-Neto, F. R., et al. (2017). Asymmetric bioreduction of beta-ketoesters derivatives by *Kluyveromyces marxianus*: influence of molecular structure on the conversion and enantiomeric excess. *An. Acad. Bras. Cienc.* 89, 1403–1415. doi: 10.1590/0001-3765201720170118
- Ortiz-Merino, R. A., Kuanyshv, N., Braun-Galleani, S., Byrne, K. P., Porro, D., Branduardi, P., et al. (2017). Evolutionary restoration of fertility in an interspecies hybrid yeast, by whole-genome duplication after a failed mating-type switch. *PLoS Biol.* 15:e2002128. doi: 10.1371/journal.pbio.2002128
- Peter, J., and Schacherer, J. (2016). Population genomics of yeasts: towards a comprehensive view across a broad evolutionary scale. *Yeast* 33, 73–81. doi: 10.1002/yea.3142
- Quarella, S., Lovrovich, P., Scalabrini, S., Campedelli, I., Backovic, A., Gatto, V., et al. (2016). Draft genome sequence of the probiotic yeast *Kluyveromyces marxianus* fragilis B0399. *Genome Announc.* 4:e00923-16. doi: 10.1128/genomeA.00923-16
- Rajaei, N., Chiruvella, K. K., Lin, F., and Aström, S. U. (2014). Domesticated transposase Kat1 and its fossil imprints induce sexual differentiation in yeast. *Proc. Natl. Acad. Sci. U.S.A.* 111, 15491–15496. doi: 10.1073/pnas.1406027111
- Ricci, A., Allende, A., Bolton, D., Chemaly, M., Davies, R., Girones, R., et al. (2017). Update of the list of QPS-recommended biological agents intentionally added to food or feed as notified to EFSA 5: suitability of taxonomic units notified to EFSA until September 2016. *EFSA J.* 15:4663. doi: 10.2903/j.efsa.2017.4663
- Rocha, S. N., Abrahão-Neto, J., and Gombert, A. K. (2011). Physiological diversity within the *Kluyveromyces marxianus* species [corrected]. *Antonie Van Leeuwenhoek* 100, 619–630. doi: 10.1007/s10482-011-9617-7
- Rodicio, R., and Heinisch, J. J. (2013). Yeast on the milky way: genetics, physiology and biotechnology of *Kluyveromyces lactis*. *Yeast* 30, 165–177. doi: 10.1002/yea.2954
- Schabort D. T. W. P., Letebele, P. K., Steyn, L., Kilian, S. G., and du Preez, J. C. (2016). Differential RNA-seq, multi-network analysis and metabolic regulation analysis of *Kluyveromyces marxianus* reveals a compartmentalised response to xylose. *PLoS ONE* 11:e0156242. doi: 10.1371/journal.pone.0156242
- Schacherer, J., Shapiro, J. A., Ruderfer, D. M., and Kruglyak, L. (2009). Comprehensive polymorphism survey elucidates population structure of *Saccharomyces cerevisiae*. *Nature* 458, 342–345. doi: 10.1038/nature07670
- Schaffrath, R., and Breunig, K. D. (2000). Genetics and molecular physiology of the yeast *Kluyveromyces lactis*. *Fungal Genet. Biol.* 30, 173–190. doi: 10.1006/fgbi.2000.1221
- Schröder, M. S., Martínez de San Vicente, K., Prandini, T. H., Hammel, S., Higgins, D. G., Bagagli, E., et al. (2016). Multiple origins of the pathogenic yeast *Candida orthopsilosis* by separate hybridizations between two parental species. *PLoS Genet.* 12:e1006404. doi: 10.1371/journal.pgen.1006404
- Signori, L., Passolunghi, S., Ruohonen, L., Porro, D., and Branduardi, P. (2014). Effect of oxygenation and temperature on glucose-xylose fermentation in *Kluyveromyces marxianus* CBS712 strain. *Microb. Cell Fact.* 13:51. doi: 10.1186/1475-2859-13-51
- Silveira, W. B., Diniz, R. H., Cerdán, M. E., González-Siso, M. I., Souza Rde, A., Vidigal, P. M., et al. (2014). Genomic sequence of the yeast *Kluyveromyces marxianus* CCT 7735 (UFV-3), a highly lactose-fermenting yeast isolated from the Brazilian Dairy Industry. *Genome Announc.* 2:e01136-14. doi: 10.1128/genomeA.01136-14
- Souciet, J., Aigle, M., Artiguenave, F., Blandin, G., Bolotin-Fukuhara, M., Bon, E., et al. (2000). Genomic exploration of the hemiascomycetous yeasts: 1. A set of yeast species for molecular evolution studies. *FEBS Lett.* 487, 3–12. doi: 10.1016/S0014-5793(00)02272-9
- Strope, P. K., Skelly, D. A., Kozmin, S. G., Mahadevan, G., Stone, E. A., Magwene, P. M., et al. (2015). The 100-genomes strains, an *S. cerevisiae* resource that illuminates its natural phenotypic and genotypic variation and emergence as an opportunistic pathogen. *Genome Res.* 25, 762–774. doi: 10.1101/gr.185538.114
- Suh, S. O., Gujjari, P., Beres, C., Beck, B., and Zhou, J. (2013). Proposal of *Zygosaccharomyces parabailii* sp. nov. and *Zygosaccharomyces pseudobailii* sp. nov., novel species closely related to *Zygosaccharomyces bailii*. *Int. J. Syst. Evol. Microbiol.* 63(Pt 5), 1922–1929. doi: 10.1099/ijs.0.048058-0
- Tittarelli, F., Varela, J. A., Gethins, J., Stanton, C., Ross, C., Suzzi, P., et al. (2018). Development and implementation of multilocus sequence typing to study diversity of the yeast *Kluyveromyces marxianus* in Italian cheeses. *Microb. Genomics* doi: 10.1099/mgen.0.000153. [Epub ahead of print].
- Van der Auwera, G. A., Carneiro, M. O., Hartl, C., Poplin, R., Del Angel, G., Levy-Moonshine, A., et al. (2013). From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr. Protoc. Bioinformatics* 43, 11.10.1–11.10.33. doi: 10.1002/0471250953.bi1110s43
- Varela, J. A., Montini, N., Scully, D., Van der Ploeg, R., Oreb, M., Boles, E., et al. (2017). Polymorphisms in the LAC12 gene explain lactose utilisation variability in *Kluyveromyces marxianus* strains. *FEMS Yeast Res.* 17:fox021. doi: 10.1093/femsyr/fox021
- Verdugo Valdez, A., Segura Garcia, L., Kirchmayr, M., Ramírez Rodríguez, P., González Esquinca, A., Coria, R., et al. (2011). Yeast communities associated with artisanal mezcals fermentations from *Agave salmiana*. *Antonie Van Leeuwenhoek* 100, 497–506. doi: 10.1007/s10482-011-9605-y
- Wang, Y. J., Ying, B. B., Shen, W., Zheng, R. C., and Zheng, Y. G. (2017). Rational design of *Kluyveromyces marxianus* ZJB14056 aldo-keto reductase KmAKR to enhance diastereoselectivity and activity. *Enzyme Microb. Technol.* 107, 32–40. doi: 10.1016/j.enzmictec.2017.07.012
- Wendland, J. (2014). Lager yeast comes of age. *Eukaryot. Cell* 13, 1256–1265. doi: 10.1128/EC.00134-14
- Wu, W. H., Hung, W. C., Lo, K. Y., Chen, Y. H., Wan, H. P., and Cheng, K. C. (2016). Bioethanol production from taro waste using thermo-tolerant yeast *Kluyveromyces marxianus* K21. *Bioresour. Technol.* 201, 27–32. doi: 10.1016/j.biortech.2015.11.015



- Yarimizu, T., Nonklang, S., Nakamura, J., Tokuda, S., Nakagawa, T., Llorreungsil, S., et al. (2013). Identification of auxotrophic mutants of the yeast *Kluyveromyces marxianus* by non-homologous end joining-mediated integrative transformation with genes from *Saccharomyces cerevisiae*. *Yeast* 30, 485–500. doi: 10.1002/yea.2985
- Zhu, Y. O., Sherlock, G., and Petrov, D. A. (2016). Whole genome analysis of 132 clinical *Saccharomyces cerevisiae* strains reveals extensive ploidy variation. *G3 (Bethesda)*. 6, 2421–2434. doi: 10.1534/g3.116.029397
- Zonneveld, B. J. M., and Steensma, H. Y. (2003). “Mating, sporulation and tetrad analysis in *Kluyveromyces lactis*,” in *Non-Conventional Yeasts in Genetics, Biochemistry and Biotechnology: Practical Protocols*, eds K. Wolf, K. Breunig, and G. Barth (Berlin; Heidelberg: Springer), 151–154.

**Conflict of Interest Statement:** CW was employed by company Lallemand Inc and J-MG and NK by Heineken.

The other authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Ortiz-Merino, Varela, Coughlan, Hoshida, Silveira, Wilde, Kuijpers, Geertman, Wolfe and Morrissey. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Interplay of Chimeric Mating-Type Loci Impairs Fertility Rescue and Accounts for Intra-Strain Variability in *Zygosaccharomyces rouxii* Interspecies Hybrid ATCC42981

Melissa Bizzarri<sup>1</sup>, Stefano Cassanelli<sup>1</sup>, Laura Bartolini<sup>1</sup>, Leszek P. Pryszcz<sup>2</sup>, Michala Dušková<sup>3</sup>, Hana Sychrová<sup>3</sup> and Lisa Solieri<sup>1\*</sup>

## OPEN ACCESS

### Edited by:

Isabel Sá-Correia,  
University of Lisbon, Portugal

### Reviewed by:

Paola Branduardi,  
Università degli Studi di  
Milano-Bicocca, Italy  
Geraldine Butler,  
University College Dublin, Ireland

### \*Correspondence:

Lisa Solieri  
lisa.solieri@unimore.it

### Specialty section:

This article was submitted to  
Evolutionary and Genomic  
Microbiology,  
a section of the journal  
Frontiers in Genetics

Received: 01 October 2018

Accepted: 11 February 2019

Published: 01 March 2019

### Citation:

Bizzarri M, Cassanelli S,  
Bartolini L, Pryszcz LP, Dušková M,  
Sychrová H and Solieri L (2019)  
Interplay of Chimeric Mating-Type  
Loci Impairs Fertility Rescue  
and Accounts for Intra-Strain  
Variability in *Zygosaccharomyces*  
*rouxii* Interspecies Hybrid  
ATCC42981. *Front. Genet.* 10:137.  
doi: 10.3389/fgene.2019.00137

<sup>1</sup> Department of Life Sciences, University of Modena and Reggio Emilia, Reggio Emilia, Italy, <sup>2</sup> Laboratory of Zebrafish Developmental Genomics, International Institute of Molecular and Cell Biology, Warsaw, Poland, <sup>3</sup> Department of Membrane Transport, Institute of Physiology, Czech Academy of Sciences, Prague, Czechia

The pre-whole genome duplication (WGD) *Zygosaccharomyces* clade comprises several allodiploid strain/species with industrially interesting traits. The salt-tolerant yeast ATCC42981 is a sterile and allodiploid strain which contains two subgenomes, one of them resembling the haploid parental species *Z. rouxii*. Recently, different mating-type-like (*MTL*) loci repertoires were reported for ATCC42981 and the Japanese strain JCM22060, which are considered two stocks of the same strain. *MTL* reconstruction by direct sequencing approach is challenging due to gene redundancy, structure complexities, and allodiploid nature of ATCC42981. Here, DBG2OLC and MaSuRCA hybrid *de novo* assemblies of ONT and Illumina reads were combined with *in vitro* long PCR to definitively solve these incongruences. ATCC42981 exhibits several chimeric *MTL* loci resulting from reciprocal translocation between parental haplotypes and retains two *MATa*/*MATα* expression loci, in contrast to *MATα* in JCM22060. Consistently to these reconstructions, JCM22060, but not ATCC42981, undergoes mating and meiosis. To ascertain whether the damage of one allele at the *MAT* locus regains the complete sexual cycle in ATCC42981, we removed the *MATα* expressed locus by gene deletion. The resulting *MATa*- hemizygous mutants did not show any evidence of sporulation, as well as of self- and out-crossing fertility, probably because incomplete silencing at the chimeric *HMLα* cassette masks the loss of heterozygosity at the *MAT* locus. We also found that *MATα* deletion switched off *a2* transcription, an activator of *a*-specific genes in pre-WGD species. These findings suggest that regulatory scheme of cell identity needs to be further investigated in *Z. rouxii* protoploid yeast.

**Keywords:** mating-type, MinION, sexual cycle, *Zygosaccharomyces*, chimeric loci, interspecies hybridization, yeast

## INTRODUCTION

Polyploidization, a state resulting from doubling of a genome within a species (autopolyploidy) or the merging between different species (allopolyploidy) (Campbell et al., 2016), is an important evolutionary force which shapes eukaryotic genomes (Albertin and Marullo, 2012), triggers speciation, and can result in phenotypic changes driving adaptation (Ohno, 1970). A whole-genome duplication (WGD) event occurred approximately 100–200 Mya in the common ancestor of six yeast genera in the family Saccharomycetaceae, including *Saccharomyces cerevisiae* (as reviewed by Wolfe et al., 2015). WGD was recently proposed to be a direct consequence of an ancient hybridization between two ancestral species (Marcet-Houben and Gabaldón, 2015), followed by genome doubling of initially sterile hybrid to regain fertility, *i.e.*, the ability to undergo meiosis and produce viable spore (Wolfe, 2015).

Different mechanisms can contribute to hybrid infertility, such as chromosomal missegregation caused by meiosis I non-disjunction (Boynton et al., 2018), chromosomal rearrangements (Liti et al., 2006; Rajeh et al., 2018), and Dobzhansky–Muller gene incompatibilities either between nuclear genes (Bizzarri et al., 2016) or between mitochondrial and nuclear genes (Lee et al., 2008). Specialized loci, called the mating-type (*MAT*)-like (*MTL*) cassettes, regulate mating between haploid cells with opposite *MATa* and *MATα* idiomorphs, as well as meiosis in diploid *a/α* cells. In diplontic yeast *S. cerevisiae* *MAT* locus on chromosome III contains either the *a1* or the *α1* and *α2* genes in *Ya* and *Yα* segments, respectively, surrounded by *X* and *Z* regions at the left and right sides. In haploid *α* cells, *α1* activates the *α*-specific genes (*αsgs*), while *α2* represses a cohort of *a*-specific genes (*asgs*), which a cells transcribe by default (Haber, 2012). Finally, diploid *a/α* cells are meiosis but not mating-competent, because the *a1-α2* heterodimer positively regulates *IME1* (Inducer of Meiosis) gene expression and represses the transcription of *RME1*, a haploid-specific gene (*hsg*) that inhibits entry into meiosis, and of other *hsgs* required for mating responses. *S. cerevisiae* cells also have extra copies of *MAT* genes at the *HMRa* and *HMLα* loci located close to telomeres of chromosome III and silenced by a combination of the Sir1–4 proteins (Hickman et al., 2011). These extra copies serve as donors during the mating-type switching which enables *MATa* cells to convert into *MATα* cells, or vice versa, and to mate each other. This autodiploidization event is triggered by a site-specific endonuclease called HO which induces double-strand break at *Z* region of the *MAT* locus. In *Saccharomyces* interspecies hybrids, experimental deletion of one *MAT* locus or elimination of the entire chromosome carrying one *MAT* locus yielded fertile allotetraploids (Greig et al., 2002; Pfliegler et al., 2012; Karanyicz et al., 2017). More recently, the *MAT* locus damage was proposed to be the most plausible evolutionary route which enables natural interspecies hybrids of the *Zygosaccharomyces bailii* complex to rescue mating and meiosis (Ortiz-Merino et al., 2017; Braun-Galleani et al., 2018).

In the Saccharomycetaceae lineage, *Z. rouxii* stands on the crossroad where different and relevant evolutionary events take their way (Dujon and Louis, 2017). This evolutionary route involves ancient allopolyploidization between two parental

lineages, one of which was close to *Z. rouxii* and *Torulaspora delbrueckii* (ZT) clade (Marcet-Houben and Gabaldón, 2015). *Z. rouxii* represents the early branching species before WGD that recruits HO from a LAGLIDADG intein to catalyze the first step of mating-type switching (Fabre et al., 2005). Furthermore, *Z. rouxii* exhibits the triplication of *MTL* loci, which is a genomic landmark of the Saccharomycetaceae family, but, in contrast to *S. cerevisiae*, it lacks of *MAT-HMR* linkage. Whereas the route of *αsg* regulation appears to be conserved, the regulatory circuit of *asgs* has been extensively rewired across the Saccharomycotina clade. Instead of the negative regulatory circuit widespread in post-WGD species, several pre-WGD species activate *asgs* by an HMG-domain protein (*a2*) that is encoded by *MATa* (Tsong et al., 2003). Conventionally, *Z. rouxii* displays haplontic life style, where heterothallic haploid cells with opposite mating-type mate each other or, alternatively, homothallic haploid cells switch mating-type and subsequently undergo mating between mother and daughter cells. In both cases, the transient diploid zygote should sporulate to restore the haploid state. Alternatively, stable allodiploid strains arose from mating between divergent haploid parents. One parental haplotype (called T-subgenome) resembles *Z. rouxii* and was 15% different from the other parental haplotype (called P-subgenome) (Gordon and Wolfe, 2008; Bizzarri et al., 2016, 2018; Watanabe et al., 2017).

Both haploid and allodiploid strains show highly variable gene arrangements around *MTL*, suggesting that these loci are recombination hotspot during error-prone mating-type switching events (Watanabe et al., 2013; Solieri et al., 2014). Structural rearrangements are so rampant in these regions that different stock cultures of the same haploid (Watanabe et al., 2013) or allodiploid (Bizzarri et al., 2016; Watanabe et al., 2017) strains can display distinct *MTL* repertoires. For instance, differences in *MTL* loci were recently found between two sub-cultures of the allodiploid strain ATCC42981. In our previous work, we found 7 *MTL* loci in in-house stock of ATCC42981 (termed ATCC42981\_R for convenience) (Bizzarri et al., 2016), while Watanabe et al. (2017) detected 6 *MTL* loci in strain JCM22060, the Japanese stock of ATCC42981. Ectopic recombination between *MTL*-flanking regions from divergent parental haplotypes yields chimeric arrangements hardly to resolve both by targeted long PCR approaches (Bizzarri et al., 2016) and by genome sequencing technologies based on short reads (Watanabe et al., 2017).

In 2014, the MinION sequencing device (Oxford Nanopore Technology, ONT) was released and initially exploited to sequence and assemble PCR products or microbial genomes (Jain et al., 2016). Recent improvements in protein pore (a laboratory-evolved *Escherichia coli* CsgG mutant named R9.4), library preparation techniques (1D ligation and 1D rapid), sequencing speed (450 bases/s), and control software enabled the usage of Nanopore sequence data, in combination with other sequencing technologies, for assembling eukaryotic genomes including yeasts, nematodes and human (Istace et al., 2017; Jansen et al., 2017; Yue et al., 2017; Jain et al., 2018). The main advantage of ONT is that reads can reach tens of kilobases (Jain et al., 2016), making more easy to resolve repeat regions and to detect structural variation. Recently, the genome of allodiploid

strain ATCC42981\_R was sequenced and assembled through a *de novo* hybrid strategy which combined MinION long and Illumina short reads (Bizzarri et al., 2018).

Here, we took advantage from the newly released genome of ATCC42981\_R (Bizzarri et al., 2018), in order to resolve incongruences in the highly dynamic *MTL* loci. Furthermore, we deleted the expressed *MAT* $\alpha^P$  locus in ATCC42981\_R to test whether the loss of *MAT* heterozygosity can induce genome doubling and rescue fertility in allodiploid cells of the ZT clade.

## MATERIALS AND METHODS

### Strains, Plasmids, and Culture Conditions

Yeast strains and plasmids used in this study are listed in **Table 1**. Yeast cells were routinely propagated at 28°C in YPD (1% yeast extract, 2% peptone, 2% glucose) medium with 1.5% agar when necessary. Stock cultures were stored at –80°C with glycerol at final concentration of 25% (v/v) for long-term preservation. For sporulation and mating assays, MEA (5% malt extract, 2% agar) with and without 6% NaCl and YM (0.3% yeast extract, 0.5% peptone, 0.3% malt extract, 1% dextrose, 1.5% agar) media were used. *Z. parabolii* strain G21C was used as control for conjugated asci formation after growth on MEA medium. When required, YPD medium was supplemented with G418 (100 mg mL<sup>–1</sup>; MP Biomedicals, Germany) to the final concentration of 200  $\mu$ g mL<sup>–1</sup>.

### DNA Manipulations

DNA manipulations were performed according to standard protocols (Sambrook et al., 1989). Genomic DNA from yeast cells was isolated according to Hoffman and Winston (1987), while plasmid DNA from *E. coli* was isolated using the GenElute™ Plasmid Miniprep Kit (Sigma). DNA quantity and quality were evaluated electrophoretically and spectrophotometrically using a NanoDrop ND-1000 device (Thermo Scientific, Waltham, MA, United States). Zymoclean™ Gel DNA Recovery and DNA Clean & Concentrator™-5 Kits (Zymo Research, Orange, CA, United States) were used for the isolation of DNA fragments from agarose gels and for PCR amplicons purification, respectively. Long PCR amplifications were carried out with rTaq DNA polymerase (Takara Bio, Shiga, Japan) according to manufacturer's instructions. For colony PCR 1  $\mu$ l of DNA extracted with lithium acetate-SDS method (Löoke et al., 2011) was amplified with DreamTaq polymerase (Thermo Scientific, Waltham, MA, United States) according to the manufacturer's instructions in 20  $\mu$ l reaction volume. All PCR amplifications were carried out in a T100 Thermal cycler (Bio-Rad, Hercules, CA, United States). All primers used in this study are listed in **Supplementary Table S1**.

### Genome Re-assembly

Hybrid assembly of ATCC42981\_R genome from Oxford Nanopore and Illumina reads was released to the European Nucleotide Archive under accession number PRJEB26771

(Bizzarri et al., 2018). In the deposited assembly Platanus contigs were scaffolded into 33 scaffolds with corrected MinION reads using DBG2OLC (Ye et al., 2016). These scaffolds were submitted to two-step polishing with long reads using Racon v1.2.0 (Vaser et al., 2017) and with short reads using Pilon v1.22 (Walker et al., 2014), and, finally, reduced using Redundans v.014 (Pryszcz and Gabaldón, 2016). Here, both long and short reads were assembled jointly with the alternative assembly algorithm Maryland Super-Read Celera Assembler v.3.2.2 (MaSuRCA) (Zimin et al., 2017) with default settings. Gene identification and annotation were carried out through the Yeast Genome Annotation Pipeline (YGAP)<sup>1</sup> without frameshift correction (Proux-Wéra et al., 2012). MaSuRCA assembly completeness was assessed by Benchmarking Universal Single-Copy Orthologs (BUSCO) v3.0.2 (Simão et al., 2015) using saccharomycetales\_odb9 data set.

### *MTL* Loci Search and Sanger-Based Validation

Search for *MTL* loci on scaffolds generated by DBG2OLC and MaSuRCA hybrid assemblies was carried out with a custom BLAST server built using the Sequenceserver software package (Priyam et al., 2015).  $\gamma$  and  $\alpha$  sequences and *MTL* flanking genes from the haploid reference genome of *Z. rouxii* CBS732<sup>T</sup> (Souciet et al., 2009) were used as queries.

The *in silico* *MTL* arrangements were *in vitro* validated by PCR and Sanger sequencing. Specific primer sets were built on *MTL*-flanking regions outside the X and Z regions (**Supplementary Table S1**). For putatively active *MAT* $\alpha^P$  cassette, walking strategy was adopted to cover ~1 Kb downstream and upstream  $\gamma$  (Wang et al., 2011). According Watanabe et al. (2017), *MTL* and flanking genes were marked with T and P superscripts when they shared >99% identity with *Z. rouxii* CBS732<sup>T</sup> or with P-subgenome from allodiploid NBRC110957 (Watanabe et al., 2017), respectively. N superscript was used to identify gene variants divergent from both T and P counterparts (identity lower than 99%). The 5' *MTL*-flanking gene ZYRO0F18524g was named as *CHA1<sub>L</sub>* for brevity. Sequences were aligned with Clustal Omega (Sievers and Higgins, 2014) and viewed using Jalview (Waterhouse et al., 2009). Neighbor-joining (NJ) tree was built using MEGA v.6 software (Tamura et al., 2013).

### Deletion Cassettes Construction and Yeast Transformation

Deletion of the active *MAT* $\alpha$  locus from P-subgenome (abbreviated as *MAT* $\alpha^P$ ) was performed with the reusable *loxP*-*kanMX*-*loxP* cassette as described previously (Güldener et al., 1996). The *MAT* $\alpha$ 1/2cp2-*kanMX*-F-80nt and *MAT* $\alpha$ 1/2cp2-*kanMX*-R-80nt primers contained ~80 bp homology sequences outside the X and Z regions of *MAT* $\alpha^P$  locus, respectively, and were used to amplify the *kanMX* deletion cassette from pUG6. After purification, the resulting PCR product was used to transform *Z. rouxii* cells by electroporation with

<sup>1</sup><http://wolfe.ucd.ie/annotation/>



**TABLE 1** | Yeast strains and plasmid used in this work.

Strains	Other codes	Relevant characteristics	Reference
<b><i>Z. parabailii</i></b> G21C	Na	Isolated from balsamic glaze	This work
<b><i>Z. rouxii</i> strains</b>			
ATCC42981_R	JCM22060	<i>MATa<sup>T</sup>/MATa<sup>P</sup></i>	Solieri et al., 2008; Bizzarri et al., 2016, 2017
CBS4837	NCYC1682; NBRC1876	<i>MATa<sup>P</sup></i>	Solieri et al., 2008; Sato et al., 2017; Watanabe et al., 2017
CBS4838	NBRC1877	<i>MATa<sup>P</sup></i>	Solieri et al., 2008; Watanabe et al., 2017
<b>Transformants</b>			
ATCC42981 <i>MATaΔ</i> clone_6	Na	<i>MATa<sup>P</sup> Δ::loxP-KanMX-loxP; MATa<sup>T</sup></i>	This work
ATCC42981 <i>MATaΔ</i> clone_65	Na	<i>MATa<sup>P</sup> Δ::loxP-KanMX-loxP; MATa<sup>T</sup></i>	This work
ATCC42981 <i>MATaΔ</i> clone_74	Na	<i>MATa<sup>P</sup> Δ::loxP-KanMX-loxP; MATa<sup>T</sup></i>	This work
ATCC42981 <i>MATaΔ</i> clone_177	Na	<i>MATa<sup>P</sup> Δ::loxP-KanMX-loxP; MATa<sup>T</sup></i>	This work
<b>Plasmids</b>			
pUG6		<i>loxP-KanMX-loxP</i> module	Güldener et al., 1996

ATCC42981\_R represents in-house stock culture of strain ATCC42981. Other codes indicate the name of strains in other culture collections. Genotype reports *Y* sequence from the putatively expression active mating-type locus (*MAT*). *T* and *P* superscripts indicate *Ya* or *Yα* sequences from *T*- or *P*-subgenomes, respectively. Na, not available.

a modified protocol from Pribylova and Sychrova (2003). Briefly, cells were grown (28°C; 180 rpm) in 80 ml of YPD medium supplemented with 300 mM NaCl until the exponential phase (corresponding to OD<sub>600nm</sub> of 0.7–0.8). After washing with ddH<sub>2</sub>O, cells were resuspended into 16 ml of TE buffer (Tris-hydrochloride buffer, pH 8.0, containing 1.0 mM EDTA) supplemented with 25 mM dithiothreitol and 20 mM LiAc, and incubated at 30°C for 30 min with gently shaking. Cells were centrifuged at 6,000 g for 5 min at 4°C, and washed twice by resuspension in 20 mL of ice-cold 1 M sorbitol. Finally, cells were washed in 5 ml of ice-cold 1M sorbitol and resuspended in 800 μl of ice-cold 1 M sorbitol. One hundred microliter of competent cell suspension were transferred into a pre-chilled 2-mm electroporation cuvette (Molecular Bioproducts Inc., San Diego, CA, United States) and 1 μg of *loxP-kanMX-loxP* deletion cassette was added before the electroporation at 2250 V/cm for 5 ms (Eporator, Eppendorf, Germany). Immediately after electroporation, 100 μl of ice-cold 1 M sorbitol were added to electroporation mixture. Before plating on selective YPDA medium supplemented with G418, the transformation mixtures were incubated for 2 h in 5 ml of YPD at 30°C. In G418<sup>R</sup> clones, targeted gene disruption was confirmed by full-length, 5′-, and 3′-end diagnostic PCRs (Supplementary Figure S1).

## RNA Extraction, cDNA Synthesis and RT-PCR

RNA was extracted from ATCC42981 wild type and deletion mutants cultured in YPD and harvested at stationary phase, as previously reported (Solieri et al., 2016). RNAs were reverse transcribed using 0.5 μM oligo (dT) and RevertAid H Minus Reverse Transcriptase (Thermo Scientific, Waltham, MA, United States) according to the manufacturer's instructions. cDNAs (25 ng) were amplified using DreamTaq polymerase with primers specific for different variants of *MATa1*, *MATα1*, and *MATα2* genes, as well as for *T* and *P* variants of asgs *AGA2*, *STE2*, and *STE6* (Supplementary Table S1).

## RESULTS

### Inventory of ATCC42981\_R *MTL* Cassettes

To unambiguously characterize *MTL* loci in our stock culture, we exploited the new available ATCC42981\_R draft genome (Bizzarri et al., 2018). This draft genome relies on the hybrid DBG2OLC assembly of MinION ultra-long and Illumina MiSeq short reads to resolve high heterozygosity and span repetitive regions, which represent the greatest technical challenges during the assembly of complex non-haploid genomes (Treangen and Salzberg, 2012; Del Angel et al., 2018).

Custom BLAST searches using Sequenceserver identified six scaffolds harboring 8 *MTL* loci (2 *MTLα<sup>T</sup>*, 4 *MTLα<sup>P</sup>*, and 2 *MTLa*) mainly at the scaffold edge (Table 2). As this pattern matched only partially either with our previous results (Bizzarri et al., 2016) or with the JCM22060 set of *MTL* loci (Watanabe et al., 2017), we took into account the possibility of misassembled segments, mainly considering that reference *P*-type genome is not available. Misassemblies could be more burdensome at the *MTL* loci which contain the long non-tandem repeated *X* and *Z* sequences enriched in homopolymeric stretches. To circumvent these caveats, we validated the *MTL* cassettes found in DBG2OLC assembly *in silico* by using the alternative assembler MaSuRCA, as well as *in vitro* by direct PCR and Sanger sequencing. With appropriate caution, agreement between these assemblies – which are completely independent in assembly algorithms – and among assemblies and Sanger sequencing can confirm the integrity of *MTL* cassettes.

MaSuRCA assembly resulted in an assembled genome size of 21.09 Mb distributed across 59 scaffolds with *N*<sub>50</sub> of 1.34 Mb (Table 3). In our previous analysis, 10,524 predicted genes were estimated by Exonerate (Slater and Birney, 2005; Bizzarri et al., 2018). Here, gene number was re-calculated for both DBG2OLC and MaSuRCA assemblies using YGAP software. Based on this analysis, DBG2OLC and MaSuRCA displayed roughly the same number of predicted genes

**TABLE 2** | Overview of the *MTL* cassettes confirmed by hybrid *de novo* genome assemblies and PCR approach.

Cassette	DBG2OLC Scaffolds	Coordinates	PCR	MaSuRCA	JCM22060
<b>Y<math>\alpha</math><sup>T</sup></b>					
<i>DIC1</i> <sup>P</sup> - <i>MTL</i> $\alpha$ <sup>T</sup> -ZYRO0F18634g <sup>T</sup>	UEMZ01000028.1	45,980...56,093	+	+	4B
<i>CHA1</i> <sub>L</sub> <sup>T</sup> - <i>MTL</i> $\alpha$ <sup>T</sup> -ZYRO0F18634g <sup>T</sup>	UEMZ01000013.1	263,261...275,557	+	—	—
<b>Y<math>\alpha</math><sup>P</sup></b>					
<i>DIC1</i> <sup>T</sup> - <i>MTL</i> $\alpha$ <sup>P</sup> - <i>SLA2</i> <sup>P</sup>	UEMZ01000013.1	35,683...40,522	+	+	1D
<i>CHA1</i> <sub>L</sub> <sup>T</sup> - <i>MTL</i> $\alpha$ <sup>P</sup> - <i>SLA2</i> <sup>P</sup>	UEMZ01000003.1	11,848...18,890 (r.c.)	+	+	2D
<i>CHA1</i> <sub>L</sub> <sup>P</sup> - <i>MTL</i> $\alpha$ <sup>P</sup> -ZYRO0F18634g <sup>P</sup>	UEMZ01000003.1	241,988...250,941 (r.c.)	+	+	5E
<i>DIC1</i> <sup>T</sup> - <i>MTL</i> $\alpha$ <sup>P</sup> - <i>SLA2</i> <sup>N</sup>	UEMZ01000007.1	1,444,839...1,449,671 (r.c.)	—	—	—
<b>Y<math>\alpha</math></b>					
<i>DIC1</i> <sup>N</sup> - <i>MTL</i> $\alpha$ <sup>N</sup> - <i>SLA2</i> <sup>T</sup>	UEMZ01000008.1	1,427,380...1,431,846	+	—	—
<i>CHA1</i> <sup>T</sup> - <i>MTL</i> $\alpha$ <sup>T</sup> -ZYRO0C18392g <sup>T</sup>	UEMZ01000015.1	1,296,432...1,304,606 (r.c.)	+	+	3C
<i>CHA1</i> <sup>P</sup> - <i>MTL</i> $\alpha$ <sup>P</sup> -ZYRO0C18392g <sup>P</sup>	n.r.	n.r.	+	+	6F

*MTL* cassettes were found by BLAST searching *Y $\alpha$*  and *Y $\alpha$*  coding DNA sequences from *Z. rouxii* CBS732<sup>T</sup> reference genome against DBG2OLC and MaSuRCA assemblies and then they were validated by long PCR and Sanger sequencing. JCM66020 *MTL* cassettes were described based on flanking genes according to nomenclature reported by Watanabe et al. (2017). Briefly, numbers 1–6 indicate 5'-flanking genes *DIC1*<sup>T</sup>, *CHA1*<sub>L</sub><sup>T</sup>, *CHA1*<sup>T</sup>, *DIC1*<sup>P</sup>, *CHA1*<sub>L</sub><sup>P</sup>, and *CHA1*<sup>P</sup>, respectively. Capital letters A–F indicate 3'-flanking genes *SLA2*<sup>T</sup>, ZYRO0F18634g<sup>T</sup>, ZYRO0C18392g<sup>T</sup>, *SLA2*<sup>P</sup>, ZYRO0F18634g<sup>P</sup> and ZYRO0C18392g<sup>P</sup>, respectively. r.c., reverse complement; n.r., not reported.

(Table 3). Single-copy orthologs analysis by BUSCO 3.0 revealed a high degree of completeness in both assemblies (>98.0%), even if MaSuRCA retrieved more duplicated orthologs than DBG2OLC.

MaSuRCA validated five out of eight DBG2OLC *MTL* cassettes, while one was MaSuRCA assembly specific (Table 2 and Supplementary Table S2). All six MaSuRCA cassettes were consistent with JCM22060. Like in DBG2OLC, MaSuRCA-derived *MTL* cassettes especially laid at the scaffold edges, confirming difficulties in scaffolding over repeated X and Z sequences shared by multiple and partially divergent *MTL*-flanking regions. Figure 1 showed that direct *in vitro* PCR validated eight *MTL* arrangements. Moreover, MaSuRCA consensus sequences were often more consistent with Sanger sequencing compared with DBG2OLC. Probably, this discrepancy resulted from a more aggressive DBG2OLC approach enabled to reduce the genome fragmentation, but at the price of local assembling accuracy.

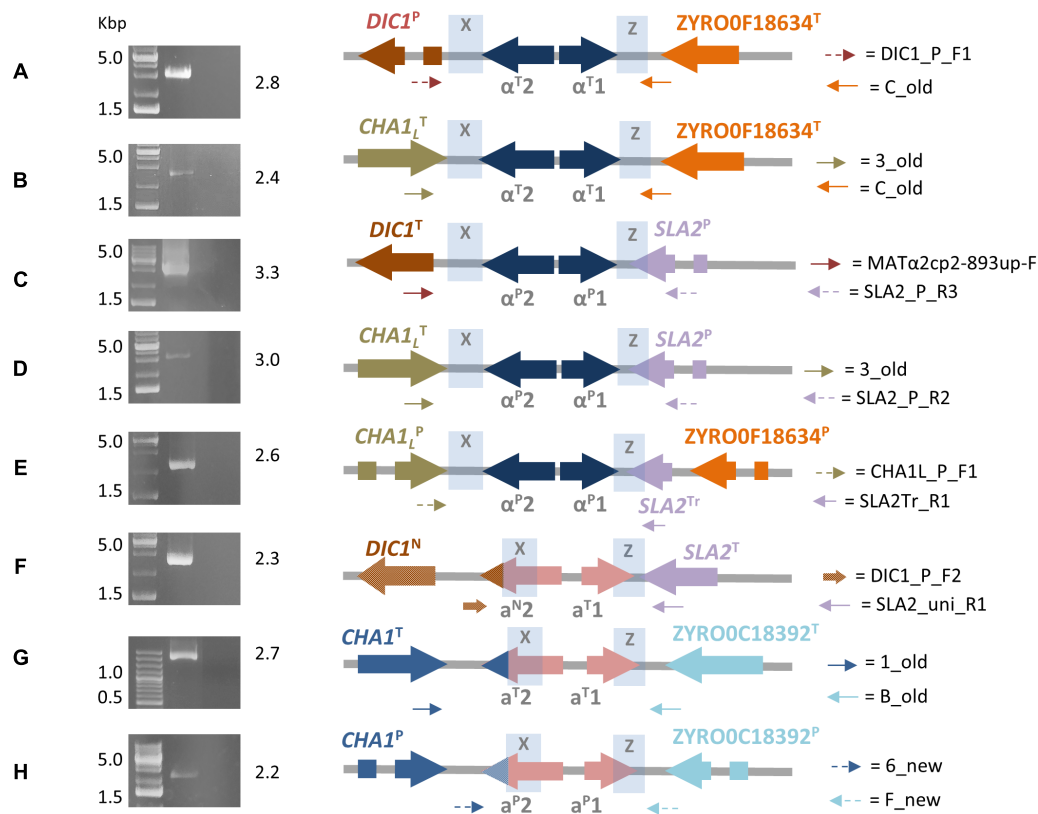
**TABLE 3** | Assembly metrics and annotation completeness obtained by using BUSCO universal fungal genes (saccharomycetales\_odb9) data set.

Feature	Assembler	
	DBG2OLC	MaSuRCA
Assembly size (bp)	20,910,059	21,093,102
Number of scaffolds	33	59
G+C content (%)	39.65	39.95
<i>N</i> <sub>50</sub> contig size (bp)	1,393,912	1,337,761
<i>N</i> <sub>90</sub> contig size (bp)	400,395	638,558
Gaps	0	0
Longest scaffold (bp)	1,903,919	2,966,114
Number of genes	10,678	10,362
BUSCO complete genes	1,687 (98.6%)	1,692 (98.9%)
BUSCO duplicated genes	1,491 (87.1%)	1,582 (92.5%)

### *MTL* $\alpha$ <sup>P</sup> Cassettes

Congruently with our previous data (Bizzarri et al., 2016), the DBG2OLC and MaSuRCA assemblies supported the cassettes *DIC1*<sup>T</sup>-*MTL* $\alpha$ <sup>P</sup>-*SLA2*<sup>P</sup> and *CHA1*<sub>L</sub><sup>T</sup>-*MTL* $\alpha$ <sup>P</sup>-*SLA2*<sup>P</sup> (Table 2). PCR approach confirmed these arrangements (Figure 1). Pairwise comparisons showed that *DIC1*<sup>T</sup> and *CHA1*<sub>L</sub><sup>T</sup> were 100% identical to the *Z. rouxii* CBS732<sup>T</sup> counterparts. In cassette *DIC1*<sup>T</sup>-*MTL* $\alpha$ <sup>P</sup>-*SLA2*<sup>P</sup>, the 3'-flanking gene *SLA2*<sup>P</sup> diverged from CBS732<sup>T</sup> counterpart (83.65% identity), and resembled *SLA2* found in allopolyploid NBRC110957 and NBRC1876 (99.58% identity) (Sato et al., 2017; Watanabe et al., 2017). In *CHA1*<sub>L</sub><sup>T</sup>-*MTL* $\alpha$ <sup>P</sup>-*SLA2*<sup>P</sup> cassette, DBG2OLC assembly reported mismatches compared to *SLA2*<sup>P</sup> in NBRC110957 (93.12% identity), which were not supported by MaSuRCA. Sanger sequencing confirmed the accuracy of MaSuRCA assembling (Supplementary Figure S2).

According to the model of T- and P-subgenomes, *DIC1*<sup>T</sup>-*MTL* $\alpha$ <sup>P</sup>-*SLA2*<sup>P</sup> and *CHA1*<sub>L</sub><sup>T</sup>-*MTL* $\alpha$ <sup>P</sup>-*SLA2*<sup>P</sup> should be chimeric cassettes arisen from rearrangements involving the X regions. NBRC110957 also contains the *DIC1*<sup>T</sup>-*MTL* $\alpha$ <sup>P</sup>-*SLA2*<sup>P</sup> chimeric arrangement (Watanabe et al., 2017; Supplementary Table S2), suggesting that recombination is frequent upstream the Y sequence. Recombinant sites at the *MAT* locus were also documented in several *Saccharomyces* lager yeasts (Bond et al., 2004; Hewitt et al., 2014). Breakpoints frequently occurred at the right of the *MAT* locus resulting in hybrid *S. cerevisiae*–*S. eubayanus* chromosomes III. These chromosomes contain *S. eubayanus* sequences in the W region and *S. cerevisiae* in the Y region hitch-hiking downstream genes or *vice versa* (Monerawela and Bond, 2017). In lager yeast Ws34/70 a possible location for the recombination event is a 9-bp insertion in the *S. eubayanus* X region compared to *S. cerevisiae*. We found a similar indel between X regions of ATCC42981\_R *DIC1* variants (Supplementary Figure S3), confirming that X region could represent a specific 'fragile' chromosomal location susceptible to double strand breakage (DSB).



**FIGURE 1 |** Final ATCC42981\_R *MAT*-like cassettes organization resulting from DBG2OLC and MaSuRCA assemblies. Picture depicts full-length PCR results (left) and the corresponding inferred *MTL* arrangements (right). Genes from T- and P-subgenomes are marked with T and P superscripts, respectively. *SLA2* truncated variant is marked with Tr superscript, while *DIC1* and *MATa2* new variants with N superscript. Blue shading denotes the X and Z regions. Dot arrows indicate P-variants of *MAT* flanking genes, while full arrows stand for T-variants. Variable *MATa2* 3'-end tags are indicated as colored tips. Capital letters from A to H refer to primer pairs listed in **Supplementary Table S1**. Numbers on the right of gel electrophoresis images indicate PCR product length in Kbp.

Novel sets of P-subgenome-specific primers confirmed an additional *MTLα<sup>P</sup>* locus (*CHA1<sub>L</sub><sup>P</sup>-MTLα<sup>P</sup>-ZYRO0F18634g<sup>P</sup>*) which escaped our previous reconstruction (Bizzarri et al., 2016) (Figure 1). Based on Watanabe et al. (2017), this locus should be a cryptic *HML* cassette, which did not affect the true cell identity. This cassette had a truncated *SLA2* sequence downstream the Z region, confirming DNA erosion on the right side of *MAT* locus (Gordon et al., 2011). Interestingly, in both DBG2OLC and MaSuRCA assemblies this cassette is linked to *CHA1<sub>L</sub><sup>T</sup>-MTLα<sup>P</sup>-SLA2<sup>P</sup>* on the same scaffold (Supplementary Figure S4).

### *MTLα<sup>T</sup>* Cassettes

DBG2OLC and MaSuRCA assemblies failed to congruently reconstruct *MTLα<sup>T</sup>* loci (Table 2). DBG2OLC scaffold UEMZ01000013.1 contains *CHA1<sub>L</sub><sup>T</sup>-MTLα<sup>T</sup>-ZYRO0F18634g<sup>T</sup>* linked to the chimeric cassette *DIC1<sup>T</sup>-MTLα<sup>P</sup>-SLA2<sup>P</sup>*, while another *MTLα<sup>T</sup>* locus (*DIC1<sup>P</sup>-MTLα<sup>T</sup>-ZYRO0F18634g<sup>T</sup>*) lies on the scaffold UEMZ01000028.1. MaSuRCA assembly reported only the *DIC1<sup>P</sup>-MTLα<sup>T</sup>-ZYRO0F18634g<sup>T</sup>* cassette. Moreover, *MTL* cassette linkage differed between DBG2OLC and MaSuRCA: *DIC1<sup>T</sup>-MTLα<sup>P</sup>-SLA2<sup>P</sup>* was linked to *CHA1<sub>L</sub><sup>T</sup>-MTLα<sup>T</sup>-ZYRO0F18634g<sup>T</sup>* in DBG2OLC, while it was linked to *CHA1<sub>L</sub><sup>T</sup>-MTLα<sup>P</sup>-SLA2<sup>P</sup>* and *CHA1<sub>L</sub><sup>T</sup>-MTLα<sup>P</sup>-ZYRO0F18634g<sup>P</sup>*

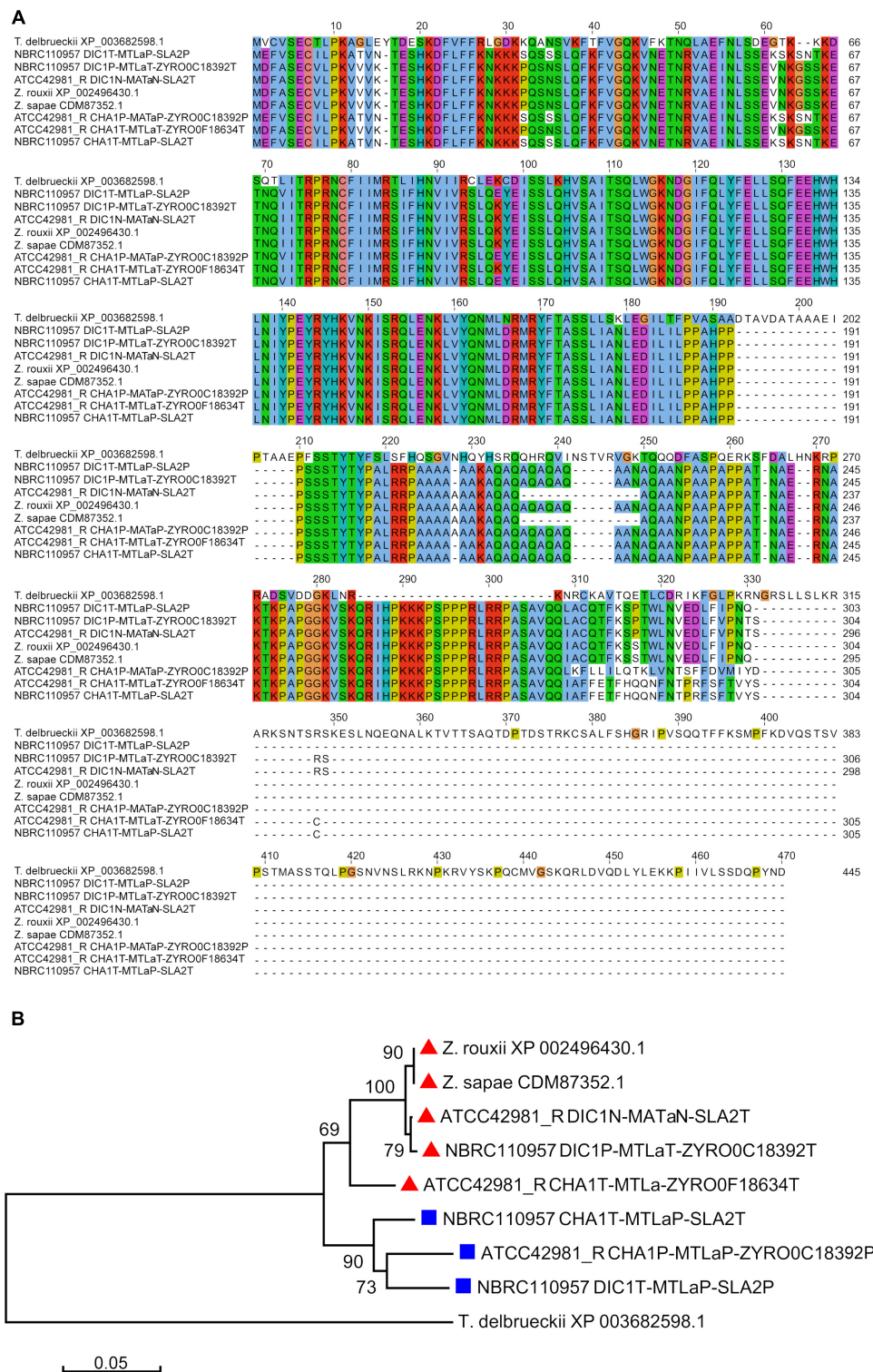
in MaSuRCA (Supplementary Figure S4). PCR approach supported both *MTLα<sup>T</sup>* cassettes from DBG2OLC assembly (Figure 1), while scaffold comparison suggests that MaSuRCA collapsed the *CHA1<sub>L</sub><sup>T</sup>* flanking regions into a single locus (Supplementary Figure S4).

### *MTLa* Cassettes

Blast search against the DBG2OLC assembly revealed two *MTLa* cassettes (Table 2). The arrangement *CHA1<sub>L</sub><sup>T</sup>-MTLa<sup>T</sup>-ZYRO0C18392g<sup>T</sup>* was also supported by MaSuRCA and PCR approach, and was congruent with our previous reconstruction (Bizzarri et al., 2016) and with JCM22060 (Watanabe et al., 2017) (Supplementary Table S2).

The second *MTLa* locus resolved by DBG2OLC, *DIC1<sup>N</sup>-MTLa<sup>N</sup>-SLA2<sup>T</sup>*, contained *a<sup>T</sup>1* and a novel *a<sup>N</sup>2* gene variant (indicated with N superscript) which was 97.99% identical to *MATa2* from NBRC110957 *DIC1<sup>P</sup>-MTLa<sup>T</sup>-ZYRO0C18392<sup>T</sup>* cassette (Figure 2). PCR approach demonstrated that this cassette really exists in ATCC42981\_R genome, even if it was missing both in MaSuRCA assembly and in JCM22060 (Figure 1). Like in case of *SLA2<sup>P</sup>* from *CHA1<sub>L</sub><sup>T</sup>-MTLα<sup>P</sup>-SLA2<sup>P</sup>*, DBG2OLC *MATa2* sequence showed some indels in homopolymeric stretches compared to the Sanger-sequence data





**FIGURE 2 |** Multiple sequence alignment and phylogenetic analysis of MATa2 proteins. **(A)** Depicts the alignment involving 9 MATa2 amino acid sequences. The amino acid identities were colored according to Clustal Omega color scheme (Sievers and Higgins, 2014). In **(B)** dendrogram was inferred using the Neighbor-Joining method. The percentages of replicate trees in which the associated taxa clustered together in the bootstrap test (1,000 replicates) are shown next to the branches, when  $\geq 50\%$ . The evolutionary distances were computed using the p-distance method and are in the units of the number of amino acid differences per site. All positions containing gaps and missing data were eliminated. Red triangles and blue squares marked T and P variants.



(98.54% pairwise identity), resulting in a prematurely interrupted ORF (data not shown). The neighbor genes at the 5' and 3' sides were a novel *DIC1* variant (named *DIC1<sup>N</sup>*) and the *SLA2<sup>T</sup>* gene, respectively. Noteworthy, the *DIC1*-*MAT*-*SLA2* arrangement is retained around the transcriptionally active *MAT* loci in almost all the pre-WGD species (Gordon et al., 2011). Therefore *DIC1<sup>N</sup>*-*MTLa<sup>N</sup>*-*SLA2<sup>T</sup>* cassette could be a good candidate to be the active *MATa* cassette in ATCC42981\_R.

Finally, PCR approach with haplotype P-specific primers identified a third *MTLa* locus (*CHAI<sup>P</sup>*-*MTLa<sup>P</sup>*-ZYRO0C18392g<sup>P</sup>) which was present in JCM22060 and in MaSuRCA assembly (Table 2). Blast search for *CHAI<sup>P</sup>* gene revealed that DBG2OLC assembler did not extend scaffold UEMZ01000005.1 beyond this gene.

## Reconstruction of *MTL* Structure

Analysis of regions around *MTL* loci assisted us to reconstruct the putative *MTL* structure in ATCC42981\_R. NBRC1130<sup>T</sup> culture retains ancestral *MTL* arrangement compared with CBS732<sup>T</sup> (Watanabe et al., 2013) and was used as reference strain. In this strain, chromosome C contains *MAT* and *HML* loci flanked by sets of genes which were also conserved around ATCC42981\_R *MTL* cassettes (Supplementary Figure S5). In particular, *MAT* locus was flanked on the left by *PEX2* and *CBP1* and on the right by *SUI1* and *CWC25*, while *HML* cassette was flanked by *VAC17* at the left side and by *FET4* and *COS12* at the right side (Figure 3). Blast analysis indicated that DBG2OLC scaffold UEMZ01000008.1 was almost collinear to NBRC1130<sup>T</sup> chromosome C in the first 1,427,380 bp. Genes upstream and downstream the *MATa<sup>N</sup>* cassette were P and T-type, respectively. Congruently, *MATa<sup>N</sup>* cassette retained the synteny with *PEX2<sup>P</sup>* and *CBP1<sup>P</sup>* at 5'- and *SUI1<sup>T</sup>* and *CWC25<sup>T</sup>* at 3'-end. However, 3'-end side was interrupted at *RAD50<sup>T</sup>*. Scaffold UEMZ01000003.1 (rc) linked *CHAI<sub>L</sub><sup>T</sup>*-*MTLa<sup>P</sup>*-*SLA2<sup>P</sup>* and *CHAI<sub>L</sub><sup>P</sup>*-*MTLa<sup>P</sup>*-ZYRO0F18634g<sup>P</sup> cassettes (Figure 3). Reciprocal translocation between chromosomes C from T and P haplotypes led to a similar arrangement in CBS4837 (Watanabe et al., 2017). As result, in CBS4837 the *MATa<sup>P</sup>* expression cassette is linked to *CHAI<sub>L</sub><sup>T</sup>*-*MTLa<sup>P</sup>*-*SLA2<sup>P</sup>* and *CHAI<sub>L</sub><sup>P</sup>*-*MTLa<sup>P</sup>*-ZYRO0F18634g<sup>P</sup>. In ATCC42981\_R, flanking gene analysis also supported a linkage between *MATa<sup>N</sup>* and *CHAI<sub>L</sub><sup>T</sup>*-*MTLa<sup>P</sup>*-*SLA2<sup>P</sup>*/*CHAI<sub>L</sub><sup>P</sup>*-*MTLa<sup>P</sup>*-ZYRO0F18634g<sup>P</sup> cassettes, suggesting that scaffolds UEMZ01000008.1 and UEMZ01000013.1 contributed to the chimeric chromosome C. Like in CBS4837 (Watanabe et al., 2017), this chromosome C could arise from a reciprocal translocation between two ancestral T and P chromosomes C.

Scaffold UEMZ01000028.1 was chimeric with P-type (*PEX2* and *CBP1*) and T-type (*FET4* and *COS12*) genes upstream and downstream the cassette *DIC1<sup>P</sup>*-*MTLa<sup>T</sup>*-ZYRO0F18634g<sup>T</sup>, respectively (Figure 3). The loss of gene block between *MAT* and *HML* cassettes suggested that a deletion between *MAT* and *HML* cassettes led to this arrangement, similar to that described in strain NBRC0686 (Watanabe et al., 2013; Supplementary Figure S5). Alternatively, in CBS4837 a similar arrangement resulted from reciprocal translocation leading to chimeric chromosome C (Watanabe et al., 2017).

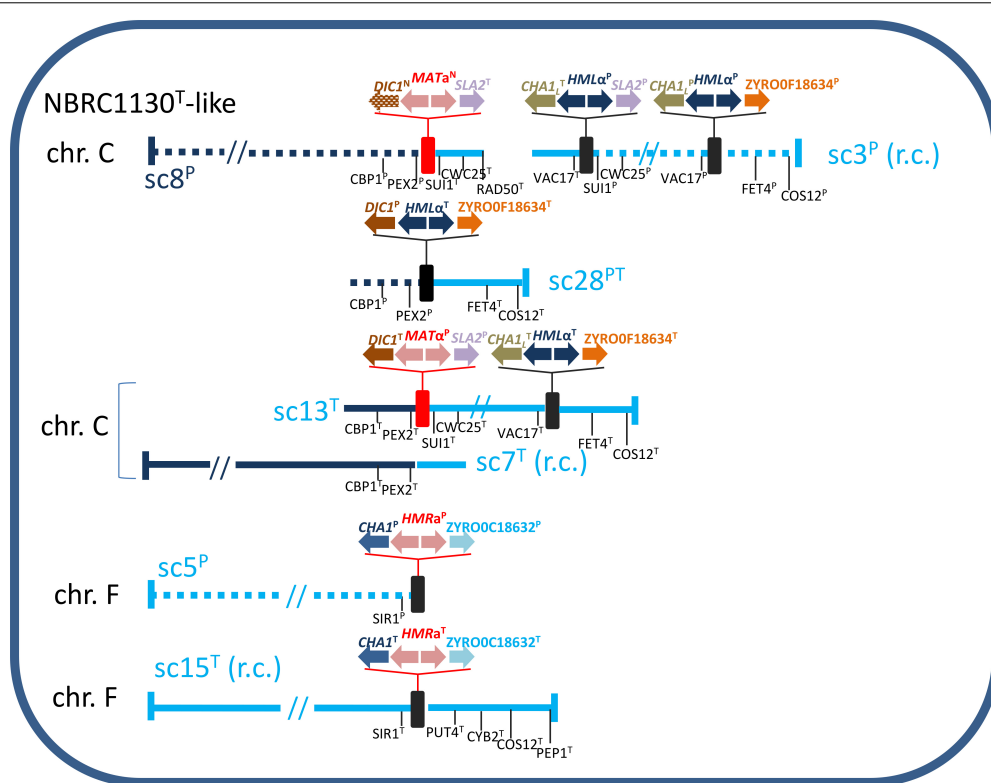
DBG2OLC scaffold UEMZ01000013.1 exhibited T-type flanking genes around *DIC1<sup>T</sup>*-*MTLa<sup>P</sup>*-*SLA2<sup>P</sup>* and *CHAI<sub>L</sub><sup>T</sup>*-*MTLa<sup>T</sup>*-ZYRO0F18634g<sup>T</sup>. Overlapping region with scaffold UEMZ01000007.1 suggested that scaffolds UEMZ01000013.1 and UEMZ01000007.1 could contribute to the T-type chromosome C in ATCC42981\_R (Figure 3).

NBRC1130<sup>T</sup> strain has the *HMRa* locus on chromosome F. *SIR1* and a set of genes including *PUT4*, *CYB2*, *COS12*, and *PEP1* are upstream and downstream to *HMRa*, respectively (Supplementary Figure S5). ATCC42981\_R DBG2OLC assembly exhibited two scaffolds retaining this synteny, namely 5 and 15 (rc). Scaffold UEMZ01000005.1 contained P-type genes, including *SIR1<sup>P</sup>* (Figure 3). Unfortunately, DBG2OLC assembler interrupted this scaffold after *CHAI<sup>P</sup>*. However, MaSuRCA assembly retained *PUT4<sup>P</sup>*, *CYB2<sup>P</sup>*, *COS12<sup>P</sup>*, and *PEP1<sup>P</sup>* downstream of *HMRa<sup>P</sup>*, suggesting that ATCC42981\_R has a P-type chromosome F collinear to NBRC1130 chromosome F. Syntenic relationships and Blast analysis supported scaffold UEMZ01000015.1 as the T-type version of NBRC1130<sup>T</sup> chromosome F (Supplementary Figure S5).

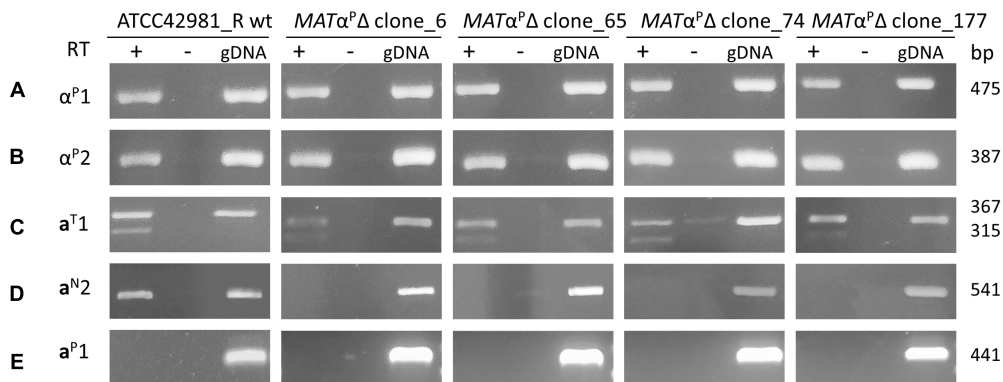
## Disclosing the True Cell Identity

Watanabe et al. (2017) identified two *MTL* patterns: strains with pattern A, such as NBRC110957, exhibit two active *MAT* loci, namely *DIC1<sup>T</sup>*-*MAT<sup>P</sup>*-*SLA2<sup>P</sup>* and *CHAI<sub>L</sub><sup>T</sup>*-*MTL<sup>P</sup>*-*SLA2<sup>T</sup>*, while strains with pattern B have *DIC1<sup>T</sup>*-*MAT<sup>P</sup>*-*SLA2<sup>P</sup>* as active *MAT* locus, even if they also actively transcribed genes from *CHAI<sub>L</sub><sup>T</sup>*-*MTL<sup>P</sup>*-*SLA2<sup>P</sup>*. JCM66020 belongs to this last group, exhibits a *MATa<sup>P</sup>* idiomorph and, congruently, mates only the tester strain a (CBS4838). Conversely, ATCC42981\_R displays another pattern of putatively active *MAT* loci, namely, *DIC1<sup>T</sup>*-*MATa<sup>P</sup>*-*SLA2<sup>P</sup>* and *DIC1<sup>N</sup>*-*MATa<sup>N</sup>*-*SLA2<sup>T</sup>*, in addition to the *CHAI<sub>L</sub><sup>T</sup>*-*MTL<sup>P</sup>*-*SLA2<sup>P</sup>* cassette. RT-PCR analysis confirmed that  $\alpha^1$ ,  $\alpha^2$ ,  $a^N2$  and  $a^T1$  genes were expressed, while  $a^P1$  gene encoded by *CHAI<sup>P</sup>*-*MTL<sup>P</sup>*-ZYRO0C18392g<sup>P</sup> cassette was silent (Figure 4). Interestingly,  $a^T1$ -specific RT-PCR resulted in two PCR amplicons compatible with alternative spliced intronic sequence.

Genome comparison with other pre-WGD yeasts indicates that *HMLa* silent cassettes are generally 5'-flanked by *CHAI<sub>L</sub>* (Gordon et al., 2011). Conversely, strains with pattern B actively transcribed *MTL* genes from *CHAI<sub>L</sub><sup>T</sup>*-*MTL*-*SLA2<sup>P</sup>* cassette without that these transcripts affect cell identity (Watanabe et al., 2017). This is evident for strain CBS4837, where genes encoding opposite  $\alpha^P$  and  $a^P$  idiomorphs are both expressed by *DIC1<sup>T</sup>*-*MAT<sup>P</sup>*-*SLA2<sup>P</sup>* and *CHAI<sub>L</sub><sup>T</sup>*-*MTL<sup>P</sup>*-*SLA2<sup>P</sup>* cassettes, respectively. In JCM22060 (encoding  $\alpha^P$  genes at both these loci), outcross experiment with CBS4837 and gamete segregation support that cell identity was determined by *DIC1<sup>T</sup>*-*MAT<sup>P</sup>*-*SLA2<sup>P</sup>* cassette. To establish which cassette contributes to cell identity in ATCC42981\_R, we deleted  $\alpha^P$  idiomorph genes by replacing the entire segment including  $\alpha^P1$ ,  $\alpha^P2$  encoding genes and the intergenic region from *DIC1<sup>T</sup>*-*MATa<sup>P</sup>*-*SLA2<sup>P</sup>* with *loxP*-*kanMX*-*loxP* module. From approximately 300 screened colonies we obtained four G418<sup>R</sup> clones. PCR genotyping showed that these



**FIGURE 3 |** Inferred gene organization around the *MTL* loci in ATCC42981\_R. Scaffold (sc) numbers referred to the DBG2OLC genome assembly deposited in European Nucleotide Archive under accession number PRJEB26771 (Bizzarri et al., 2018); for brevity each scaffold is identified by the last number of ENA code (i.e., UEMZ01000028.1 in short sc28). Solid and dotted lines referred to T- and P-subgenomes, respectively. Genes from T- and P-subgenomes are marked with T and P superscripts, respectively, while *DIC1* and *MATa2* new variants with N superscript. Red and black rectangles defined *MAT* and *HML/HMR* loci, respectively. Scaffold lengths are not in scale. r.c., reverse complement.



**FIGURE 4 |** Gene expression at the *MAT* loci in ATCC42981\_R wild type and *MATα<sup>P</sup>Δ* deletion mutants. cDNA was amplified from total RNA extracted from stationary growing cells. Plus or minus indicates with or without reverse transcriptase in cDNA synthesis reaction, respectively. Gene variants from T- and P-subgenomes are indicated with T and P superscripts, while divergent *MATa2* gene variant from *DIC1<sup>N</sup>-MATa<sup>N</sup>-SLA2<sup>T</sup>* cassette with N superscript. Capital letters from A to E refer to primer sets listed in **Supplementary Table S1**. wt, wild type; gDNA, genomic DNA.

clones are *MATα<sup>P</sup>Δ* deletants containing *loxP-kanMX-loxP* surrounded by *DIC1<sup>T</sup>* and *SLA2<sup>P</sup>* instead of *MATα<sup>P</sup>* locus (**Supplementary Figure S1**).

Gene deletion of *DIC1<sup>T</sup>-MATα<sup>P</sup>-SLA2<sup>P</sup>* cassette should abolish the heterozygosity at the *MATa/α* active loci and results in

an allodiploid partially resembling a haploid cell with a mating-type. Conversely, ATCC42981\_R *MATα<sup>P</sup>Δ* still showed α<sup>P</sup>1 and α<sup>P</sup>2 gene expression (**Figure 4**). These mRNAs could be only transcribed by the not completely silenced cassettes *CHA1<sub>L</sub><sup>T</sup>-MTLα<sup>P</sup>-SLA2<sup>P</sup>* or by *CHA1<sub>L</sub><sup>P</sup>-MTLα<sup>P</sup>-ZYRO0F18634g<sup>P</sup>*.

Since allodiploid lacking one *MAT* active locus should behave like haploid with opposite mating-type, we expected to detect both *a1* and *a2* transcripts in ATCC42981\_R *MAT $\alpha^P$*   $\Delta$  mutants. In some haploid pre-WGD species, *a2* gene encodes a transcription activator of *asgs*, while *a1* should not affect *asgs* in a cells (Tsong et al., 2003, 2006; Baker et al., 2012). Unexpectedly, RT-PCR showed that *MAT $\alpha^P$*  deletion switched off *a2* but not *a1* gene expression (Figure 4). By contrast, ATCC42981\_R wild type both transcribed *a1* and *a2* genes. Preliminary end-point RT-PCRs showed that the *asgs* *AGA2*, *STE6*, and *STE2* are transcriptionally active in both wild type and *MAT $\alpha^P$*   $\Delta$  cells (data not shown).

## Mating and Sporulation Competence Assays

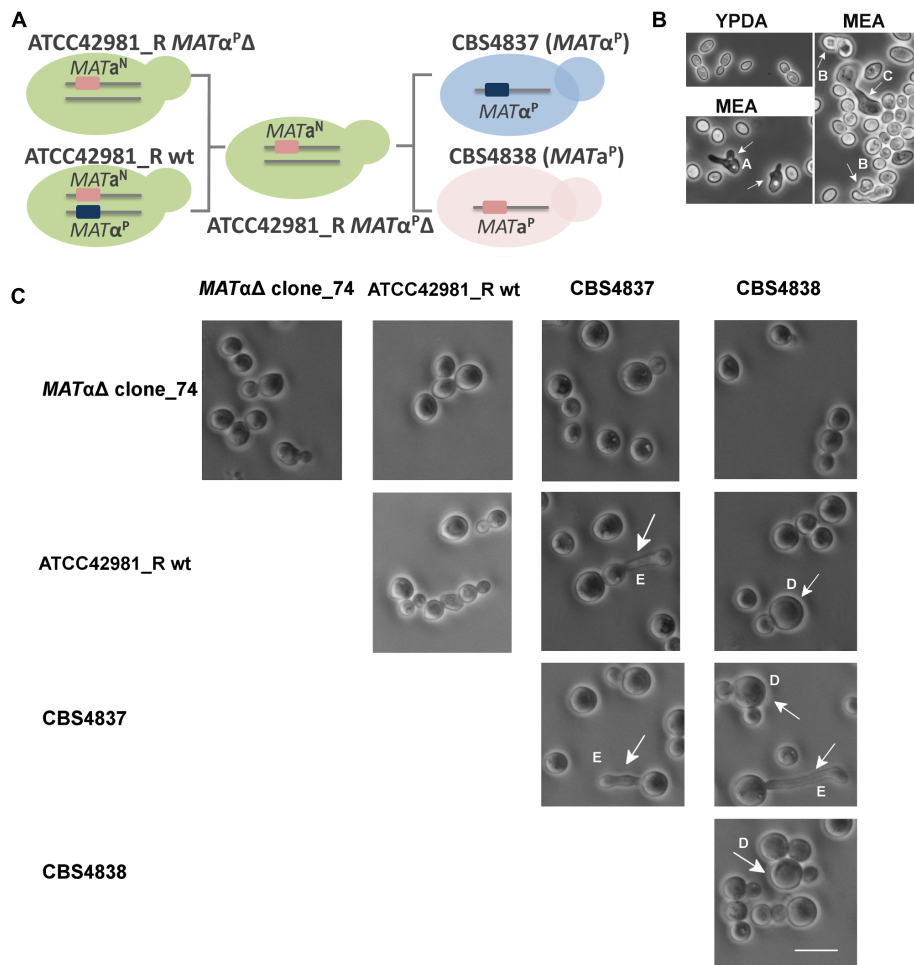
To test whether the *MAT $\alpha^P$*  deletion rescues the mating competence in ATCC42981\_R, we carried out self- and out-cross fertility assays of the wild type strain and the *MAT $\alpha^P$*   $\Delta$  transformants as monoculture or in mixture with CBS4837 ( $\alpha$ ) or CBS4838 (*a*) mating testers, respectively. If *MAT $\alpha^P$*   $\Delta$  transformants behave as homothallic haploids, they should produce shmoo and conjugated asci as monoculture, while, if they are like heterothallic haploids, they should mate and sporulate in mixture either with CBS4837 or CBS4838. We used three media reported in literature to promote zygote formation and conjugated asci of *Zygosaccharomyces* cells, as proved for *Z. parvulus* G21C (Figure 5). In particular, 5–6% NaCl addition was reported to increase sporulation occurrence (Mori and Onishi, 1967). Like the wild type strain, *MAT $\alpha^P$*   $\Delta$  mutants did not show any evidence of conjugative bridge and/or conjugative asci either as monoculture or in mixture with the mating testers (Figure 5). The composition of three test media did not affect the inability to mate or to undergo meiosis. Overall, these evidences indicate that the deletion of active *MAT $\alpha^P$*  locus did not make ATCC42981\_R cells phenotypically heterothallic or homothallic haploids.

## DISCUSSION

Our study is the first to combine the Nanopore whole-genome sequencing to conventional PCR-based methods in order to survey *MTL* loci in a *Z. rouxii* allodiploid genome. This yeast is particularly prone to outbreeding and provides a particularly appealing platform to study genome re-shaping after the merger of two parental subgenomes. Recombination and introgression between subgenomes have been rampant in hybrid yeasts, resulting in loss of heterozygosity and gradual genome reduction (Sipiczki, 2008). In *Z. rouxii* *MTL* loci markedly contribute to this genomic plasticity (Watanabe et al., 2013; Solieri et al., 2014). As consequence, this species frequently undergoes chromosomal translocations at the *MTL* loci, which make hard the understanding of true cell identity by simple *MTL* genotyping. For example, haploid *Z. rouxii* strain CBS732<sup>T</sup> switched mating-type at the *CHA1-MAT-SLA2* locus (Bizzarri et al., 2018), suggesting that *CHA1* gene flanks the actively transcribed *MAT* locus instead of *DIC1*. Several assortments of

different flanking gene variants and distinct idiomorph encoding genes make challenging and laborious to resolve the complex genetic *MTL* architecture by PCR targeted approaches. For these reasons, we generated a high-quality genome assembly in order to dissect complex rearrangements at the *MTL* loci that were not fully resolvable from the earlier survey based only on long-range PCR amplification (Bizzarri et al., 2016). One of the major advantages of the ONT is the possibility of sequencing very long DNA fragments, which span the entire *MTL* cassettes. This strategy assures to accurately reconstruct gene order around different *MTL*s. On the other hand, using noisy ultra-long reads for self-correction and assembling of highly heterozygous genomes can affect the consensus sequence accuracy and the parental haplotypes sorting. In case of ATCC42981\_R, distinguishing between homeologous sequences is further challenging as only the *Z. rouxii* parental genome is available to guide homeologous scaffold assembly. Error rate made necessary to polish MinION reads with Illumina-derived reads, resulting into DBG2OLC-driven hybrid *de novo* genome assembly (Bizzarri et al., 2018). However, our result showed that a single “best assembler” does not exist to resolve highly heterozygous and highly repeated *MTL* regions. DBG2OLC assembly suffers from poor performance in certain sequence contexts, such as in regions with low coverage or regions that contain short repeats. Besides, the new assembly generated with MaSuRCA showed higher sequencing accuracy compared to DBG2OLC, but loses some *MTL* cassettes. As bottom-end validation step, PCR approach was used to discard artificial *MTL* arrangements arisen from flawed contig assemblies. This strategy resolves controversies over *MTL* loci in ATCC42981\_R genome derived from the analysis of the Japanese stock JCM22060 (Watanabe et al., 2017).

Reconstruction of *MTL* structure indicates that ATCC42981\_R resembles CBS4837 for the exception of an additional scaffold containing *DIC1<sup>T</sup>-MTL $\alpha^P$ -SLA2<sup>P</sup>* linked to *CHA1<sup>L</sup>-MTL $\alpha^T$ -ZYRO0F18634g<sup>T</sup>* (Figure 3). This assessment was congruent with previous PFGE-Southern blotting which showed two signals for *MAT $\alpha$* -specific probe (Bizzarri et al., 2016). The most significant difference between ATCC42981\_R and JCM22060 is that ATCC42981\_R harbors the transcriptionally active *MAT $\alpha^N$*  cassette in addition to the expected *MAT $\alpha^P$* . Differently from *Z. parvulus* (Ortiz-Merino et al., 2017), *MAT $\alpha^N$*  cassette of ATCC42981\_R contains *MAT $\alpha 1$*  gene. This means that *Z. rouxii* retains the ancestral regulatory circuit based on *a1*–*a2* heterodimer as diploid cell sensor (Booth et al., 2010). Watanabe et al. (2017) showed that strain JCM22060, which contains only *MAT $\alpha^P$* , mates the tester strain *a* in a medium containing Shoyu-koji extract. By contrast, we did not find any evidence of meiosis or mating in ATCC42981\_R (Bizzarri et al., 2016), when grown on the media reported in literature to promote *Z. rouxii* mating and sporulation (James and Stratford, 2011). Watanabe et al. (2017) argued that difference in medium composition could account for the phenotypic discrepancy between ATCC42981\_R and the sister stock JCM22060. As the Shoyu-koji extract is difficult to gain in western countries, we cannot rule out



**FIGURE 5 |** Self- and out-cross fertility assays of ATCC42981\_R wild type and *MAT* $\alpha^P\Delta$  deletion mutant. **(A)** Shows the scheme of self- and out-cross fertility assays involving ATCC42981\_R wild type and *MAT* $\alpha\Delta$  clone\_74, while **(B)** depicts representative phase contrast microscopic images of control strain *Zygosaccharomyces parvibailii* G21C grown in YPDA and MEA media. **(C)** Shows selected phase contrast microscopic images of representative ATCC42981\_R *MAT* $\alpha\Delta$  clone\_74 grown as monoculture and in mixture with ATCC42981\_R wild type, CBS4837 ( $\alpha$ ) and CBS4838 (a). Phase contrast microscopic images of ATCC42981\_R wild type, CBS4837 and CBS4838 monocultures were reported for comparative purposes. The scale bar represents 11  $\mu$ m and was reported in only one picture for brevity. Capital letters from A to E indicate abortive shmoo, conjugated asci with ascospores, conjugative bridge, giant cells, and abnormal conjugative tube, respectively. All cells were photographed after 7 days on incubation on MEA medium at 27°C. wt, wild type.

this hypothesis. Otherwise, heterozygosity at the *MAT* locus could significantly contribute to the allodiploid infertility. In particular, the hybrid heterodimer with divergent  $\alpha 1$  and  $\alpha 2$  subunits brings the cell in an ‘haploid-diploid intermediate’ functional state which hamper both the meiosis commitment and the responsiveness to mating stimuli (Bizzarri et al., 2016).

In *Saccharomyces* clade, experimental deletion of one *MAT* locus leads to allotetraploids suitable to undergo meiosis (Greig et al., 2002; Pfliegler et al., 2012). Similarly, *Z. parvibailii* and *Z. pseudobailii* hybrid strains ATCC60483 and MT15 were recently supposed to be fertile due to the accidental breakage of 1 of the 2 homeologous copies of the *MAT* locus (Ortiz-Merino et al., 2017; Braun-Galleani et al., 2018). A prediction of this model is that artificial deletion of one *MAT* locus in *Zygosaccharomyces* cells

should override the arrest in mating commitment. In our model, ATCC42981\_R cells did not behave as haploids with idiomorph a, when the *MAT* $\alpha^P$  locus was deleted. This suggests that mechanism underpinning the cell identity in *Z. rouxii* hybrids could be different from those involved in cell identity regulation of the sister species *Z. parvibailii* and *Z. pseudobailii*.

Gene deletion of transcriptionally active *MAT* $\alpha^P$  locus did not rescue the ability to produce conjugated asci in ATCC42981\_R, while the persistence of  $\alpha 1$  and  $\alpha 2$  transcripts suggests that *HML* $\alpha$  silencing was leaky in ATCC42981\_R. Consequently,  $\alpha^P$  genes either from *CHA1<sub>L</sub><sup>T</sup>-MTL $\alpha^P$ -SLA2<sup>P</sup>* or *CHA1<sub>L</sub><sup>P</sup>-MTL $\alpha^P$ -ZYRO0F18634g<sup>P</sup>* are transcriptionally active in *MAT* $\alpha^P\Delta$  mutants. Strain NBRC110957, which does not have the *CHA1<sub>L</sub><sup>T</sup>-MTL $\alpha^P$ -SLA2<sup>P</sup>* cassette, uses *CHA1<sub>L</sub><sup>P</sup>-MTL $\alpha^P$ -ZYRO0F18634g<sup>P</sup>* as donor during switching from  $\alpha^P$  to  $\alpha^P$



(Watanabe et al., 2017). This suggests that *CHAI<sub>L</sub><sup>P</sup>-MTL $\alpha$ <sup>P</sup>-ZYRO0F18634g<sup>P</sup>* cassette is most likely silenced and that  $\alpha^P$  could be expressed by the *CHAI<sub>L</sub><sup>T</sup>-MTL $\alpha$ <sup>P</sup>-SLA2<sup>P</sup>* in ATCC42981\_R. Congruently, strain CBS4837 actively transcribed genes from *CHAI<sub>L</sub><sup>T</sup>-MTL $\alpha$ <sup>P</sup>-SLA2<sup>P</sup>* cassette. These findings make less probable the alternative hypothesis that *MAT $\alpha$ <sup>P</sup>* deletion induces *HML $\alpha$*  cassette de-silencing. Abnormal expression of cryptic *HMR/HML* loci has been described in *Vanderwaltozyma polyspora*, the *Z. rouxii* closest relative that branched after WGD (Roberts and Van der Walt, 1959). Consequently, *V. polyspora* haploid cells behave as a/ $\alpha$  diploid and appear mating-incompetent for many generations only to subsequently restore silencing. Significantly, *V. polyspora* lacks of Sir1 histone deacetylase, which mediates the *HM* loci silencing in *S. cerevisiae* together with the SIR complex (Sir2/Sir3/Sir4). In *S. cerevisiae* failure to recruit Sir1 is thought to account for the instability of subtelomeric silencing relative to *HM* loci (Chien et al., 1993). Like *V. polyspora*, *Candida glabrata* is another species close to *Z. rouxii*, which lacks of a SIR1 ortholog (Gabaldón et al., 2013). A defective silencing system leads to the expression of *MAT $\alpha$*  genes in *C. glabrata MAT $\alpha$*  cells (Muller et al., 2008) and makes *HML* more prone to HO cleavage at the Y/Z junctions (Boisnard et al., 2015). *Z. rouxii* has the archetypal member of the SIR1 family, KOS3 (Kin of Sir1 3) (Gallagher et al., 2009). In pre-WGD species *Torulaspora delbrueckii* KOS3 located ~1 kb away from *HMR* and plays a key role in *HML/HMR* silencing (Ellahi and Rine, 2016). Strikingly, in ATCC42981\_R we also found two KOS3 copies, KOS3<sup>T</sup> and KOS3<sup>P</sup>, upstream of *HMR $\alpha$ <sup>T</sup>* and *HMR $\alpha$ <sup>P</sup>* loci, respectively. In addition, Sir1 and the components of SIR complex have been reported to rapidly evolve in the Saccharomycetaceae family. This could potentially jeopardize the efficiency of the silencing machinery in interspecific hybrids. For example, Sir1, Sir4 and the *cis*-acting silencer sequences are incompatible in *S. cerevisiae*  $\times$  *S. uvarum* hybrids (Zill et al., 2010, 2012). In ATCC42981\_R, heterochromatin formation across silent loci could be less effective due to the incompatibility in the silencing machinery between the T- and P-subgenomes. Watanabe et al. (2017) suggest that chimeric *MTL* cassettes could display epigenetic expression control when only E silencer sequence is maintained around *MTL* locus. This could produce allodiploid single cells which undergo epigenetic silencing at one of *MAT* loci and restore fertility. In ATCC42981\_R two *DIC1-MAT-SLA2* cassettes assure active transcription of opposite idiomorphs, while the presence of E silencer only at the right side of *HML $\alpha$ <sup>P</sup>* locus could unlock the silencing and mask the loss of heterozygosity at the *MAT* locus induced by *MAT $\alpha$*  locus deletion.

Strikingly, the depletion of  $\alpha^P1$  and  $\alpha^P2$  genes switched off the  $\alpha2$  but not the  $\alpha1$  gene transcription. Moreover in both deleted and wild type strains two  $\alpha1$  alternative spliced isoforms are present, one of them compatible with the retention of first intron. In *S. cerevisiae* exon-intron structure is conserved and the retention of first intron resulted in a functional  $\alpha1$  transcriptional factor that prevents mating (Ner and Smith, 1989). Since  $\alpha1$  activates the  $\alpha$ sgs in the ancestral circuit of

yeast cell identity (Baker et al., 2011), we rule out the possibility that  $\alpha1$  is involved in  $\alpha2$  gene repression. In *S. cerevisiae*,  $\alpha2$  represses  $\alpha$ sgs by binding  $\alpha$ sgs *cis*-regulatory sequences cooperatively with a MADS-box transcription regulator, Mcm1 (Tsong et al., 2003). *Z. rouxii*, which branched from the *S. cerevisiae* lineage prior to the loss of  $\alpha2$  gene, should maintain both the  $\alpha2$  activation and the  $\alpha2$  repression of  $\alpha$ sgs (Tsong et al., 2006; Baker et al., 2012). In *Lachancea kluyveri* haploid cells,  $\alpha2$  deletion induces the transcription of the  $\alpha$ sgs *AGA1* and *AGA2*, while  $\alpha2$  deletion decreases the  $\alpha$ sgs transcript levels (Baker et al., 2012). However, to the best of our knowledge, no evidence has been provided until now about the consequences of  $\alpha2$  gene deletion in diploid cells which retain  $\alpha2$  gene. As  $\alpha1$  is still expressed in *Z. rouxii MAT $\alpha$  $\Delta$ /MAT $\alpha$*  hemizygous cells, we speculate that  $\alpha2$  silencing could be a promoter-driven event directly or indirectly regulated by  $\alpha2$ . Furthermore, in our *MAT $\alpha$  $\Delta$ /MAT $\alpha$*  model, the  $\alpha$ sgs were expressed even when  $\alpha2$  was switched off by the *MAT $\alpha$  $\Delta$*  deletion, suggesting the existence of a different  $\alpha$ sgs regulatory network in the ATCC42981\_R hybrid compared to *Z. rouxii*.

## CONCLUSION

This study revised the pattern of *MTL* loci in allodiploid strain ATCC42981\_R. By taking advantage from ONT technology, we captured a novel *MAT $\alpha$*  cassette which did not correspond to the expected T and P counterparts, providing preliminary evidences that a third haplotype contributes to this genome. The differences between ATCC42981\_R and JCM22060 support that *MTL*s are a root source of genetic variation, leading to novel chimeric *MTL* cassettes, different cell identities and, consequently, distinct phenotypic behaviors. While further researches are required to investigate mechanisms responsible of this extensive *MTL* reshaping, our results confirm that these yeast stocks are genetically unstable (Watanabe et al., 2013; Bizzarri et al., 2018). We also demonstrated how *HMR/HML* silencing is crucial to establish the cell identity, as leakage in *HML* silencing prevents allodiploid *MAT $\alpha$ <sup>P</sup> $\Delta$*  cells to behave like haploids. How allodiploid cell modulates  $\alpha2$  expression via  $\alpha2$  transcriptional factor represents an unexplored regulatory circuit that has to be investigated in future.

## DATA AVAILABILITY

The whole genome sequence datasets generated for this study can be found under the NCBI BioProject number PRJEB26771.

## AUTHOR CONTRIBUTIONS

SC and LS contributed conception and design of the study. MB conducted the experiments described in this study. LB contributed to *in vitro* PCR validation and  $\alpha$ sg expression.

HS and MD contributed to deletion mutant construction. SC and LP performed bioinformatic analysis of the whole genome sequence data. LS wrote the manuscript. SC and MB contributed to draft revision. All authors read and approved the final manuscript.

## FUNDING

LS was partially supported by the Italian Ministry of Education, University and Research (MIUR), within the framework of the Italian National Grant for Fundamental Research (FFABR 2017). The work of HS group was supported by the Ministry of Education, Youth and Sports of CR (MEYS) within the LQ1604 National Sustainability Program

## REFERENCES

- Albertin, W., and Marullo, P. (2012). Polyploidy in fungi: evolution after whole-genome duplication. *Proc. R. Soc. B. Biol. Sci.* 279, 2497–2509. doi: 10.1098/rspb.2012.0434
- Baker, C. R., Booth, L. N., Sorrells, T. R., and Johnson, A. D. (2012). Protein modularity, cooperative binding, and hybrid regulatory states underlie transcriptional network diversification. *Cell* 151, 80–95. doi: 10.1016/j.cell.2012.08.018
- Baker, C. R., Tuch, B. B., and Johnson, A. D. (2011). Extensive DNA-binding specificity divergence of a conserved transcription regulator. *Proc. Natl. Acad. Sci. U.S.A.* 108, 7493–7498. doi: 10.1073/pnas.1019177108
- Bizzarri, M., Cassanelli, S., Pryszcz, L. P., Gawor, J., Gromadka, R., and Solieri, L. (2018). Draft genome sequences of the highly halotolerant strain *Zygosaccharomyces rouxii* ATCC 42981 and the novel allopolyploid strain *Zygosaccharomyces sapae* ATB301T obtained using the MinION platform. *Microbiol. Res. Announc.* 7:e874–18. doi: 10.1128/MRA.00874-18
- Bizzarri, M., Cassanelli, S., and Solieri, L. (2017). Mating-type switching in CBS 732T derived subcultures unveils potential genetic and phenotypic novelties in haploid *Zygosaccharomyces rouxii*. *FEMS Microbiol. Lett.* 365, 1–8. doi: 10.1093/femsle/fnx263
- Bizzarri, M., Giudici, P., Cassanelli, S., and Solieri, L. (2016). Chimeric sex-determining chromosomal regions and dysregulation of cell-type identity in a sterile *Zygosaccharomyces* allopolyploid yeast. *PLoS One* 11:e0152558. doi: 10.1371/journal.pone.0152558
- Boisnard, S., Li, Y. Z., Arnaise, S., Sequeira, G., Raffoux, X., Enache-Angoulvant, A., et al. (2015). Efficient mating-type switching in *Candida glabrata* induces cell death. *PLoS One* 10:e0140990. doi: 10.1371/journal.pone.0140990
- Bond, U., Neal, C., Donnelly, D., and James, T. C. (2004). Aneuploidy and copy number breakpoints in the genome of lager yeasts mapped by microarray hybridisation. *Curr. Genet.* 45, 360–370. doi: 10.1007/s00294-004-0504-x
- Booth, L. N., Tuch, B. B., and Johnson, A. D. (2010). Intercalation of a new tier of transcription regulation into an ancient circuit. *Nature* 468, 959–963. doi: 10.1038/nature09560
- Boynton, P. J., Janzen, T., and Greig, D. (2018). Modeling the contributions of chromosome segregation errors and aneuploidy to *Saccharomyces* hybrid sterility. *Yeast* 35, 85–98. doi: 10.1002/yea.3282
- Braun-Galleani, S., Ortiz-Merino, R. A., Wu, Q., Xu, Y., and Wolfe, K. H. (2018). *Zygosaccharomyces pseudobailii*, another yeast interspecies hybrid that regained fertility by damaging one of its MAT loci. *FEMS Yeast Res.* 18:foy079. doi: 10.1093/femsyr/foy079
- Campbell, M. A., Ganley, A. R., Gabaldon, T., and Cox, M. P. (2016). The case of the missing ancient fungal polyploids. *Am. Nat.* 188, 602–614. doi: 10.1086/688763
- Chien, C. T., Buck, S., Sternglanz, R., and Shore, D. (1993). Targeting of SIR1 protein establishes transcriptional silencing at HM loci and telomeres in yeast. *Cell* 75, 531–534. doi: 10.1016/0092-8674(93)90387-6
- II (Project BIOCEV-FAR) and by the project “BIOCEV” (CZ.1.05/1.1.00/02.0109).

## ACKNOWLEDGMENTS

We are grateful to Prof. Paolo Giudici for his valuable comments and to Marcello Benevelli for help in mating assays.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2019.00137/full#supplementary-material>

- Del Angel, D. V., Hjerde, E., Sterck, L., Capella-Gutierrez, S., Notredame, C., Pettersson, O. V., et al. (2018). Ten steps to get started in genome assembly and annotation. *F1000Research* 7:ELIXIR-148. doi: 10.12688/f1000research.13598.1
- Dujon, B. A., and Louis, E. J. (2017). Genome diversity and evolution in the budding yeasts (*Saccharomycotina*). *Genetics* 206, 717–750. doi: 10.1534/genetics.116.199216
- Ellahi, A., and Rine, J. (2016). Evolution and functional trajectory of Sir1 in gene silencing. *Mol. Cell Biol.* 36, 1164–1179. doi: 10.1128/MCB.01013-15
- Fabre, E., Muller, H., Therizols, P., Lafontaine, I., Dujon, B., and Fairhead, C. (2005). Comparative genomics in hemiascomycete yeasts: evolution of sex, silencing, and subtelomeres. *Mol. Biol. Evol.* 22, 856–873. doi: 10.1093/molbev/msi070
- Gabaldón, T., Martin, T., Marcet-Houben, M., Durrens, P., Bolotin-Fukuhara, M., Lespinet, O., et al. (2013). Comparative genomics of emerging pathogens in the *Candida glabrata* clade. *BMC Genomics* 14:623. doi: 10.1186/1471-2164-14-623
- Gallagher, J. E. G., Babiarz, J. E., Teytelman, L., Wolfe, K. H., and Rine, J. (2009). Elaboration, diversification and regulation of the Sir1 family of silencing proteins in *Saccharomyces*. *Genetics* 181, 1477–1491. doi: 10.1534/genetics.108.099663
- Gordon, J. L., Armisen, D., Proux-Wéra, E., ÔhÉigeartaigh, S. S., Byrne, K. P., and Wolfe, K. H. (2011). Evolutionary erosion of yeast sex chromosomes by mating-type switching accidents. *Proc. Natl. Acad. Sci. U.S.A.* 108, 20024–20029. doi: 10.1073/pnas.1112808108
- Gordon, J. L., and Wolfe, K. H. (2008). Recent allopolyploid origin of *Zygosaccharomyces rouxii* strain ATCC 42981. *Yeast* 25, 449–456. doi: 10.1002/yea.1598
- Greig, D., Borts, R. H., Louis, E. J., and Travisano, M. (2002). Epistasis and hybrid sterility in *Saccharomyces*. *Proc. Biol. Sci.* 269, 1167–1171. doi: 10.1098/rspb.2002.1989
- Guldener, U., Heck, S., Fiedler, T., Beinhauer, J., and Hegemann, J. H. (1996). A new efficient gene disruption cassette for repeated use in budding yeast. *Nucleic Acids Res.* 24, 2519–2524. doi: 10.1093/nar/24.13.2519
- Haber, J. E. (2012). Mating-type genes and MAT switching in *Saccharomyces cerevisiae*. *Genetics* 191, 33–64. doi: 10.1534/genetics.111.134577
- Hewitt, S. K., Donaldson, I. J., Lovell, S. C., and Delneri, D. (2014). Sequencing and characterisation of rearrangements in three *S. pastorianus* strains reveals the presence of chimeric genes and gives evidence of breakpoint reuse. *PLoS One* 9:e92203. doi: 10.1371/journal.pone.0092203
- Hickman, M. A., Froyd, C. A., and Rusche, L. N. (2011). Reinventing heterochromatin in budding yeasts: Sir2 and the origin recognition complex take center stage. *Eukaryot. Cell* 10, 1183–1192. doi: 10.1128/EC.05123-11
- Hoffman, C. S., and Winston, F. (1987). A ten-minute DNA preparation from yeast efficiently releases autonomous plasmids for transformation of *Escherichia coli*. *Gene* 51, 267–272. doi: 10.1016/0378-1119(87)90131-4
- Istace, B., Friedrich, A., d'Agata, L., Faye, S., Payen, E., Beluche, O., et al. (2017). De novo assembly and population genomic survey of natural yeast isolates with the oxford nanopore MinION sequencer. *Gigascience* 6, 1–13. doi: 10.1093/gigascience/giw018

- Jain, M., Koren, S., Miga, K. H., Quick, J., Rand, A. C., and Sasani, T. A. (2018). Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nat. Biotechnol.* 36, 338–345. doi: 10.1038/nbt.4060
- Jain, M., Olsen, H. E., Paten, B., and Akeson, M. (2016). The oxford nanopore MinION: delivery of nanopore sequencing to the genomics community. *Genome Biol.* 17:239. doi: 10.1186/s13059-016-1103-0
- James, S. A., and Stratford, M. (2011). “*Zygosaccharomyces* Barker (1901),” in *The Yeasts, a Taxonomic Study*, eds C. P. Kurtzman, J. W. Fell, and T. Boekhout (London: Elsevier), 937–947.
- Jansen, H. J., Liem, M., Jong-Raadsen, S. A., Dufour, S., Weltzien, F. A., Swinkels, W., et al. (2017). Rapid de novo assembly of the European eel genome from nanopore sequencing reads. *Sci. Rep.* 7:7213. doi: 10.1038/s41598-017-07650-6
- Karanyicz, E., Antunovics, Z., Kallai, Z., and Sipiczki, M. (2017). Non-introgressive genome chimerisation by malsegregation in autotetraploidised allotetraploids during meiosis of *Saccharomyces kudriavzevii* x *Saccharomyces uvarum* hybrids. *Appl. Microbiol. Biotechnol.* 2017, 4617–4633. doi: 10.1007/s00253-017-8274-9
- Lee, H., Chou, J., Cheong, L., Chang, N. H., Yang, S. Y., and Leu, J. Y. (2008). Incompatibility of nuclear and mitochondrial genomes causes hybrid sterility between two yeast species. *Cell* 135, 1065–1073. doi: 10.1016/j.cell.2008.10.047
- Liti, G., Barton, D. B. H., and Louis, E. J. (2006). Sequence diversity, reproductive isolation and species concepts in *Saccharomyces*. *Genetics* 174, 839–850. doi: 10.1534/genetics.106.062166
- Löke, M., Kristjuhan, K., and Kristjuhan, A. (2011). Extraction of genomic dna from yeasts for pcr-based applications. *BioTechniques* 50, 325–328. doi: 10.2144/000113672
- Marcet-Houben, M., and Gabaldón, T. (2015). Beyond the whole-genome duplication: phylogenetic evidence for an ancient interspecies hybridization in the baker's yeast lineage. *PLoS Biol.* 13:e1002220. doi: 10.1371/journal.pbio.1002220
- Monerawela, C., and Bond, U. (2017). Recombination sites on hybrid chromosomes in *Saccharomyces pastorianus* share common sequence motifs and define a complex evolutionary relationship between group I and II lager yeasts. *FEMS Yeast Res.* 17:fox047. doi: 10.1093/femsyr/fox047
- Mori, H., and Onishi, H. (1967). Diploid hybridization in a heterothallic haploid yeast, *Saccharomyces rouxii*. *Appl. Microbiol.* 15, 928–934.
- Muller, H., Hennequin, C., Gallaud, J., Dujon, B., and Fairhead, C. (2008). The asexual yeast *Candida glabrata* maintains distinct a and  $\alpha$  haploid mating-types. *Eukaryot. Cell* 7, 848–858. doi: 10.1128/EC.00456-07
- Ner, S. S., and Smith, M. (1989). Role of intron splicing in the function of the MATa1 gene of *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* 9, 4613–4620. doi: 10.1128/mcb.9.11.4613
- Ohno, S. (1970). *Evolution by Gene Duplication*. New York, NY: Springer-Verlag. doi: 10.1007/978-3-642-86659-3
- Ortiz-Merino, R. A., Kuanyshv, N., Braun-Galleani, S., Byrne, K. P., Porro, D., and Branduardi, P. (2017). Evolutionary restoration of fertility in an interspecies hybrid yeast, by whole-genome duplication after a failed mating-type switch. *PLoS Biol.* 15:e2002128. doi: 10.1371/journal.pbio.2002128
- Pflieger, P. W., Antunovics, Z., and Sipiczki, M. (2012). Double sterility barrier between *Saccharomyces* species and its breakdown in allopolyploid hybrids by chromosome loss. *FEMS Yeast Res.* 12, 703–718. doi: 10.1111/j.1567-1364.2012.00820.x
- Pribylova, L., and Sychrova, H. (2003). Efficient transformation of the osmotolerant yeast *Zygosaccharomyces rouxii* by electroporation. *J. Microbiol. Methods* 55, 481–484. doi: 10.1016/S0167-7012(03)00197-0
- Priyam, A., Woodcroft, B. J., Rai, V., Munagala, A., Moghul, I., Ter, F., et al. (2015). Sequenceserver: a modern graphical user interface for custom BLAST databases. *bioRxiv* [Preprint]. doi: 10.1101/033142
- Proux-Wéra, E., Armisen, D., Byrne, K. P., and Wolfe, K. H. (2012). A pipeline for automated annotation of yeast genome sequences by a conserved-synteny approach. *BMC Bioinformatics* 13:237. doi: 10.1186/1471-2105-13-237
- Pryszcz, L., and Gabaldón, T. (2016). Redundans: an assembly pipeline for highly heterozygous genomes. *Nucleic Acids Res.* 44:e113. doi: 10.1093/nar/gkw294
- Rajeh, A., Lv, J., and Lin, Z. (2018). Heterogeneous rates of genome rearrangement contributed to the disparity of species richness in Ascomycota. *BMC Genomics* 19:282. doi: 10.1186/s12864-018-4683-0
- Roberts, C., and Van der Walt, J. P. (1959). The life cycle of *Kluyveromyces polysporus*. *C. R. Trav. Lab. Carlsberg Chim.* 31, 129–148.
- Sambrook, J., Maniatis, T., and Fritsch, E. (1989). *Molecular Cloning: A Laboratory Manual*. New York, NY: Cold Spring Harbor Laboratory.
- Sato, A., Matsushima, K., Oshima, K., Hattori, M., and Koyama, Y. (2017). Draft genome sequencing of the highly halotolerant and allopolyploid yeast *Zygosaccharomyces rouxii* NBRC 1876. *Genome Announc.* 5:e1610-16. doi: 10.1128/genomeA.01610-16
- Sievers, F., and Higgins, D. G. (2014). “Clustal omega, accurate alignment of very large numbers of sequences,” in *Multiple Sequence Alignment Methods*, ed. D. J. Russell (Totowa, NJ: Humana Press). doi: 10.1007/978-1-62703-646-7\_6
- Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., and Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31, 3210–3212. doi: 10.1093/bioinformatics/btv35
- Sipiczki, M. (2008). Interspecies hybridization and recombination in *Saccharomyces* wine yeasts. *FEMS Yeast Res.* 8, 996–1007. doi: 10.1111/j.1567-1364.2008.00369.x
- Slater, G. S. C., and Birney, E. (2005). Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics* 6:31. doi: 10.1186/1471-2105-6-31
- Solieri, L., Cassanelli, S., Croce, M. A., and Giudici, P. (2008). Genome size and ploidy level: new insights for elucidating relationships in *Zygosaccharomyces* species. *Fungal Genet. Biol.* 45, 1582–1590. doi: 10.1016/j.fgb.2008.10.001
- Solieri, L., Dakal, T. C., Giudici, P., and Cassanelli, S. (2014). Sex-determination system in the diploid yeast *Zygosaccharomyces sapae*. *G3* 4, 1011–1025. doi: 10.1534/g3.114.010405
- Solieri, L., Vezzani, V., Cassanelli, S., Dakal, T. C., Pazzini, J., and Giudici, P. (2016). Differential hypersaline stress response in *Zygosaccharomyces rouxii* complex yeasts: a physiological and transcriptional study. *FEMS Yeast Res.* 16:fow063. doi: 10.1093/femsyr/fow063
- Souciet, J. L., Dujon, B., and Gaillardin, C. (2009). Comparative genomics of protoploid *Saccharomycetaceae*. *Genome Res.* 19, 1696–1709. doi: 10.1101/gr.091546.109
- Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013). MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* 30, 2725–2729. doi: 10.1093/molbev/mst197
- Treangen, T. J., and Salzberg, S. L. (2012). Repetitive DNA and next-generation sequencing: computational challenges and solutions. *Nat. Rev. Genet.* 13:36. doi: 10.1038/nrg3117
- Tsong, A. E., Miller, M. G., Raisner, R. M., and Johnson, A. D. (2003). Evolution of a combinatorial transcriptional circuit: a case study in yeasts. *Cell* 115, 389–399. doi: 10.1016/S0092-8674(03)00885-7
- Tsong, A. E., Tuch, B. B., Li, H., and Johnson, A. D. (2006). Evolution of alternative transcriptional circuits with identical logic. *Nature* 443, 415–420. doi: 10.1038/nature05099
- Vaser, R., Sovic, I., Nagarajan, N., and Šikic, M. (2017). Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Res.* 27, 737–746. doi: 10.1101/gr.214270.116
- Walker, B. J., Abee, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., et al. (2014). Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 9:e112963. doi: 10.1371/journal.pone.0112963
- Wang, Z., Ye, S., Li, J., Zheng, B., Bao, M., and Ning, G. (2011). Fusion primer and nested integrated PCR (FPNI-PCR): a new high-efficiency strategy for rapid chromosome walking or flanking sequence cloning. *BMC Biotechnol.* 11:109. doi: 10.1186/1472-6750-11-109
- Watanabe, J., Uehara, K., and Mogi, Y. (2013). Diversity of mating-type chromosome structures in the yeast *Zygosaccharomyces rouxii* caused by ectopic exchanges between MAT-Like loci. *PLoS One* 8:e62121. doi: 10.1371/journal.pone.0062121
- Watanabe, J., Uehara, K., Mogi, Y., and Tsukioka, Y. (2017). Mechanism for restoration of fertility in hybrid *Zygosaccharomyces rouxii* generated by interspecies hybridization. *Appl. Environ. Microbiol.* 83:AEM.1187-17. doi: 10.1128/AEM.01187-17

- Waterhouse, A. M., Procter, J. B., Martin, D. M., Clamp, M., and Barton, G. J. (2009). Jalview version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25, 1189–1191. doi: 10.1093/bioinformatics/btp033
- Wolfe, K. H. (2015). Origin of the yeast whole-genome duplication. *PLoS Biol.* 13:e1002221. doi: 10.1371/journal.pbio.1002221
- Wolfe, K. H., Armisen, D., Proux-Wera, E., OhEigeartaigh, S. S., Azam, H., Gordon, J. L., et al. (2015). Clade- and species-specific features of genome evolution in the *Saccharomycetaceae*. *FEMS Yeast Res.* 15:fov035. doi: 10.1093/femsyr/fov035
- Ye, C., Hill, C. M., Wu, S., Ruan, J., and Ma, Z. S. (2016). DBG2OLC: efficient assembly of large genomes using long erroneous reads of the third generation sequencing technologies. *Sci. Rep.* 6:31900. doi: 10.1038/srep31900
- Yue, J. X., Li, J., Aigrain, L., Hallin, J., Persson, K., Oliver, K., et al. (2017). Contrasting evolutionary genome dynamics between domesticated and wild yeasts. *Nat. Genet.* 49, 913–924. doi: 10.1093/bioinformatics/bty614
- Zill, O. A., Scannell, D., Teytelman, L., and Rine, J. (2010). Co-evolution of transcriptional silencing proteins and the DNA elements specifying their assembly. *PLoS Biol.* 8:e1000550. doi: 10.1371/journal.pbio.1000550
- Zill, O. A., Scannell, D. R., Kuei, J., Sadhu, M., and Rine, J. (2012). Evolutionary analysis of heterochromatin protein compatibility by interspecies complementation in *Saccharomyces*. *Genetics* 192, 1001–1014. doi: 10.1534/genetics.112.141549
- Zimin, A. V., Puiu, D., Luo, M. C., Zhu, T., Koren, S., Marçais, G., et al. (2017). Hybrid assembly of the large and highly repetitive genome of *Aegilops tauschii*, a progenitor of bread wheat, with the MaSuRCA mega-reads algorithm. *Genome Res.* 27, 787–792. doi: 10.1101/gr.213405.116

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Bizzarri, Cassanelli, Bartolini, Pryszcz, Dušková, Sychrová and Solieri. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Advantages of publishing in Frontiers



## OPEN ACCESS

Articles are free to read  
for greatest visibility  
and readership



## FAST PUBLICATION

Around 90 days  
from submission  
to decision



## HIGH QUALITY PEER-REVIEW

Rigorous, collaborative,  
and constructive  
peer-review



## TRANSPARENT PEER-REVIEW

Editors and reviewers  
acknowledged by name  
on published articles

## Frontiers

Avenue du Tribunal-Fédéral 34  
1005 Lausanne | Switzerland

Visit us: [www.frontiersin.org](http://www.frontiersin.org)

Contact us: [info@frontiersin.org](mailto:info@frontiersin.org) | +41 21 510 17 00



## REPRODUCIBILITY OF RESEARCH

Support open data  
and methods to enhance  
research reproducibility



## DIGITAL PUBLISHING

Articles designed  
for optimal readership  
across devices



## FOLLOW US

@frontiersin



## IMPACT METRICS

Advanced article metrics  
track visibility across  
digital media



## EXTENSIVE PROMOTION

Marketing  
and promotion  
of impactful research



## LOOP RESEARCH NETWORK

Our network  
increases your  
article's readership