

Molecular Biology of Archaea, 2022

Edited by

Marleen van Wolferen, Solenne Ithurbide,
Michel Geovanni Santiago-Martínez and Arthur Charles-Orszag

Published in

Frontiers in Microbiology



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 978-2-8325-4806-6
DOI 10.3389/978-2-8325-4806-6

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: frontiersin.org/about/contact

Molecular Biology of Archaea - 2022

Topic editors

Marleen van Wolferen — University of Freiburg, Germany

Solenne Ithurbide — University of Montreal, Canada

Michel Geovanni Santiago-Martínez — University of Connecticut, United States

Arthur Charles-Orszag — University of California, San Francisco, United States

Citation

van Wolferen, M., Ithurbide, S., Santiago-Martínez, M. G., Charles-Orszag, A., eds. (2024). *Molecular Biology of Archaea - 2022*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-4806-6

Table of contents

- 05 **Editorial: Molecular Biology of Archaea - 2022**
Marleen van Wolferen, Solenne Ithurbide,
Michel Geovanni Santiago-Martínez and Arthur Charles-Orszag
- 08 **Integrated conjugative plasmid drives high frequency
chromosomal gene transfer in *Sulfolobus islandicus***
Ruben L. Sanchez-Nieves, Changyi Zhang and Rachel J. Whitaker
- 18 **Biochemical and genetic examination of two
aminotransferases from the hyperthermophilic archaeon
*Thermococcus kodakarensis***
Yu Su, Yuta Michimori and Haruyuki Atomi
- 34 **The extremophile *Picrophilus torridus* carries a DNA adenine
methylase M.PtoI that is part of a Type I
restriction-modification system**
Pallavi Gulati, Ashish Singh, Manisha Goel and Swati Saha
- 50 **Analysis of the external signals driving the transcriptional
regulation of the main genes involved in denitrification in
*Haloferax mediterranei***
Jose María Miralles-Robledillo, Rosa María Martínez-Espinosa and
Carmen Pire
- 65 **Improving the genetic system for *Halorubrum lacusprofundi*
to allow in-frame deletions**
L. Johanna Gebhard, Iain G. Duggin and Susanne Erdmann
- 75 **Complete genome assembly and methylome dissection of
Methanococcus aeolicus PL15/H^p**
Alexey Fomenkov, Peter Weigle, Colleen McClung, Casey Madinger
and Richard J. Roberts
- 85 **Structural and biochemical elucidation of class I hybrid
cluster protein natively extracted from a marine
methanogenic archaeon**
Olivier N. Lemaire, Mélissa Belhamri and Tristan Wagner
- 101 **Functional diversity of nanohaloarchaea within
xylan-degrading consortia**
Oleg Reva, Enzo Messina, Violetta La Cono, Francesca Crisafi,
Francesco Smedile, Gina La Spada, Laura Marturano,
Elena A. Selivanova, Manfred Rohde, Mart Krupovic and
Michail M. Yakimov
- 118 **Transcriptome profiling of Nudix hydrolase gene deletions in
the thermoacidophilic archaeon *Sulfolobus acidocaldarius***
Ruth Breuer, José Vicente Gomes-Filho, Jing Yuan and
Lennart Randau
- 128 **The universal Sua5/TsaC family evolved different mechanisms
for the synthesis of a key tRNA modification**
Adeline Pichard-Kostuch, Violette Da Cunha, Jacques Oberto,
Ludovic Sauguet and Tamara Basta

- 142 **A sweet new set of inducible and constitutive promoters in *Haloferax volcanii***
Theopi Rados, Katherine Andre, Micaela Cerletti and Alex Bisson
- 151 **Membrane lipid and expression responses of *Saccharolobus islandicus* REY15A to acid and cold stress**
Beverly K. Chiu, Jacob Waldbauer, Felix J. Elling, Öykü Z. Mete, Lichun Zhang, Ann Pearson, Erin M. Eggleston and William D. Leavitt
- 170 **Probing archaeal cell biology: exploring the use of dyes in the imaging of *Sulfolobus* cells**
Alice Cezanne, Baukje Hoogenberg and Buzz Baum
- 180 **"Influence of plasmids, selection markers and auxotrophic mutations on *Haloferax volcanii* cell shape plasticity"**
Megha Patro, Iain G. Duggin, Sonja-Verena Albers and Solenne Ithurbide
- 196 **Application of the endogenous CRISPR-Cas type I-D system for genetic engineering in the thermoacidophilic archaeon *Sulfolobus acidocaldarius***
Jan Bost, Alejandra Recalde, Bianca Waßmer, Alexander Wagner, Bettina Siebers and Sonja-Verena Albers
- 208 **The transcriptional regulator EarA and intergenic terminator sequences modulate archaeellation in *Pyrococcus furiosus***
Richard Stöckl, Laura Nißl, Robert Reichelt, Reinhard Rachel, Dina Grohmann and Felix Grünberger
- 219 **Diversity and taxonomic revision of methanogens and other archaea in the intestinal tract of terrestrial arthropods**
Evgenii Protasov, James O. Nonoh, Joana M. Kästle Silva, Undine S. Mies, Vincent Hervé, Carsten Dietrich, Kristina Lang, Lena Mikulski, Katja Platt, Anja Poehlein, Tim Köhler-Ramm, Edouard Miambi, Hamadi I. Boga, Christopher Feldewert, David K. Ngugi, Rudy Plarre, David Sillam-Dussès, Jan Šobotník, Rolf Daniel and Andreas Brune
- 240 **Characterization of the zinc finger μ -protein HVO_0758 from *Haloferax volcanii*: biological roles, zinc binding, and NMR solution structure**
Deniz Üresin, Dennis J. Pyper, Andreas Borst, Lydia Hadjeras, Rick Gelhausen, Rolf Backofen, Cynthia Sharma, Harald Schwalbe and Jörg Soppa



OPEN ACCESS

EDITED AND REVIEWED BY

Chuanlun Zhang,
Southern University of Science and
Technology, China

*CORRESPONDENCE

Marleen van Wolferen

✉ marleen.van.wolferen@biologie.uni-freiburg.de

Solenne Ithurbide

✉ solenne.ithurbide@umontreal.ca

Michel Geovanni Santiago-Martínez

✉ geo_santiagom@uconn.edu

Arthur Charles-Orszag

✉ arthur.charles-orszag@ucsf.edu

†These authors have contributed equally to
this work

RECEIVED 29 February 2024

ACCEPTED 28 March 2024

PUBLISHED 09 April 2024

CITATION

van Wolferen M, Ithurbide S,
Santiago-Martínez MG and Charles-Orszag A
(2024) Editorial: Molecular Biology of Archaea
- 2022. *Front. Microbiol.* 15:1393932.
doi: 10.3389/fmicb.2024.1393932

COPYRIGHT

© 2024 van Wolferen, Ithurbide,
Santiago-Martínez and Charles-Orszag. 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.

Editorial: Molecular Biology of Archaea - 2022

Marleen van Wolferen^{1*†}, Solenne Ithurbide^{2*†},
Michel Geovanni Santiago-Martínez^{3*†} and
Arthur Charles-Orszag^{4*†}

¹Molecular Biology of Archaea, Institute of Biology II - Microbiology, University of Freiburg, Freiburg, Germany, ²Département de Microbiologie, Infectiologie et Immunologie, Université de Montréal, Montréal, QC, Canada, ³The Microbial Ecophysiology Laboratory, Department of Molecular and Cell Biology, University of Connecticut, Storrs, CT, United States, ⁴Department of Cellular and Molecular Pharmacology, Howard Hughes Medical Institute, University of California, San Francisco, San Francisco, CA, United States

KEYWORDS

archaea, archaeal biology, genetics, omics, molecular biology, microscopy, extremophile, evolution

Editorial on the Research Topic
[Molecular Biology of Archaea - 2022](#)

Introduction

Since their identification and reclassification four decades ago, the study of microorganisms from the Domain Archaea has proven to be a continuous source of exciting discoveries, contributing to the characterization of their unique molecular mechanisms, metabolisms, ecophysiology, phylogeny, and cell biology (Spang et al., 2017; Lyu et al., 2018; Baker et al., 2020; van Wolferen et al., 2022). These findings have revealed the impact that archaea play in nutrient cycles, biotechnology advancements, and One-Health microbiomes (Pfeifer et al., 2021; Hoegenauer et al., 2022). In addition, the advancements in archaeal biology have highlighted the key position that archaea occupy in the evolution and diversification of the Tree of Life (Spang et al., 2017; Baker et al., 2020). The study of archaea has thereby brought us closer to elucidating the origin and early forms of microbial life, while also gaining insights into the limits of life detection.

Despite these important findings and the larger audience that archaea have consequently gained, many aspects of their biology remain unexplored. However, recent and ongoing developments in the field are addressing the technical limitations related to the isolation and cultivation. These advancements are allowing archaeal researchers to tackle open and upcoming questions. This promises exciting new discoveries in the near future that will continue to build on our understanding of the biology of archaea and allow us to uncover their unique features.

We are a team of early career scientists coming from different laboratories around the world. We are involved in the initiatives Archaea Power Hour as well as Archaea.bio, both aiming at promoting the diversity of researchers, supporting the exchange of ideas, and fostering collaborations in the field of archaeal biology. In collaboration with the organizers, Jörg Soppa, Sonja-Verena Albers, and Anita Marchfelder of the EMBO workshop for Molecular Biology of Archaea (August 1–4, 2022, Frankfurt, Germany),

we have curated a collection of publications highlighting the current research trends in the Molecular Biology of Archaea and related topics. Our aim was to showcase the latest research findings and to encourage discussion about the advances, remaining challenges and future directions for researchers studying archaea.

With this goal in mind, we invited all participants of the workshop, as well as other archaeal researchers around the world, to submit their manuscripts on the molecular biology of archaea. We are pleased to have accepted 18 manuscripts that collectively contribute to our understanding of the biology of archaea. In this editorial, we provide a brief overview of these manuscripts, highlighting their key contributions to the field.

Advances on Molecular Biology of Archaea

Organisms from the domain of Archaea were once thought to thrive exclusively in extreme environments, such as hot springs, hydrothermal vents, or salt flats. They however turned out to be ubiquitous and can actually be found in very diverse environments, from oceans to soils and even the host-associated microbiomes of plants, animals and the human gut. Archaea exhibit remarkable genetic diversity and metabolic versatility, exceeding that of bacteria. They thereby represent an immense -yet largely untapped- source of new microbial biology with potential applications in various fields. Evolutionarily, it is now commonly accepted that the eukaryotic lineage roots form within the archaea. From an ecological standpoint, archaea play crucial roles in global biogeochemical cycles and climate regulation. They are the main producers of Earth's methane, and their biomass (in gigatons of carbon) is thought to rival or even surpass that of animals, making them key players in global biogeochemical cycles and the Earth's climate. Yet, compared to bacteria and eukaryotes, their significance is often overlooked.

Since 2008, the international conference Molecular Biology of Archaea (MBoA) has been instrumental in advancing the field of archaeal research. By providing a platform for the exchange of new ideas, spurring scientific collaborations and, most importantly, welcoming new researchers to the field of archaea, the MBoA has significantly contributed to the rapidly evolving field of archaeal biology. This Research Topic showcases a selection of discoveries presented at the MBoA conference held in August 2022 in Frankfurt, Germany as well as contributions from researchers around the world. Covering a wide variety of topics across different biological scales, from environmental studies to controlled laboratory experiments, in both model and non-model organisms, the research showcased here illustrates the breadth typical of research endeavors in the biology of archaea.

In the past decade, advances in metagenomics and genomics have significantly enhanced our knowledge of archaeal biology. These techniques have enabled the identification of numerous new species and lineages without the need for traditional culture-based methods. This has had widespread implications for our understanding of archaeal phylogeny, ecology, physiology, and genetic diversity. The central and multifaceted role of genomics and other -omics approaches in archaeal research is well illustrated

in this Research Topic. It includes the complete genome assembly of *Methanococcus aeolicus* PL15/Hp, relevant to biotechnology (Fomenkov et al.), the characterization of two novel extremely halophilic symbiotic nanohaloarchaea (Reva et al.), and the drawing up of an exhaustive taxonomy of methanogens associated with terrestrial arthropods intestinal tract (Protasov et al.).

Historically, the study of archaea is almost indiscernible from the study of microorganisms living in extreme environments (extremophiles). Ever since the discovery of the first hyperthermophiles in the early 1970s, various extremophilic archaeal species have successfully been cultivated. These organisms allowed the characterization of biological processes and molecular adaptations to extreme environments, such as high salt, high pressure, extreme temperatures and pH levels, oxygen-free environments, and others. More recently, the development of genetic tools in extremophilic archaea has revolutionized our ability to address functional questions in these organisms. This boom is reflected in a series of articles that dig into the physiology of thermophilic archaea.

Several of these studies cover transcriptional regulation and response, including: the regulation of archaeal cell division in *Pyrococcus furiosus* by the highly conserved transcriptional regulator EarA (Stöckl et al.); the transcriptional regulation of denitrification in *Haloferax mediterranei* (Miralles-Robledillo et al.); the transcriptional response to the deletion of Nudix hydrolases in *Sulfolobus acidocaldarius* (Breuer et al.); and lipid modifications and differential expression in response to cold stress in *Saccharolobus islandicus* (Chiu et al.) and the role of zinc finger μ -protein HVO_0758 in *Haloferax volcanii* (Üresin et al.). Other studies focused on: tRNA modification in Archaeoglobi (Pichard-Kostuch et al.); metabolism of amino acids in *Thermococcus kodakarensis* (Su et al.); high frequency chromosomal gene transfer induced by an integrated conjugative plasmid in *Sulfolobus islandicus* (Sanchez-Nieves et al.). Notably, certain archaeal species, despite being unamenable to genetic manipulations, serve as key models to characterize enzymes involved in highly conserved processes such as DNA methylation in the extremophile *Picrophilus torridus* (Gulati et al.) and methanogenesis in *Methanothermococcus thermolithotrophicus* (Lemaire et al.).

The field of archaeal biology is now in a phase of rapid growth, with continuous development of new genetic methods, even in model species where genetic tools are already available. For example, an alternative method for gene deletion based on the CRISPR-Cas system has now been introduced in the model *Sulfolobus acidocaldarius* (Bost et al.). In halophiles, a whole new set of xylose-inducible promoters has been introduced, providing valuable tools for research model haloarchaea (Rados et al.). Additionally, an improved system for in-frame gene deletion has been developed for the model species *Haloarubrum lacusprofundi* (Gebhard et al.). These advancements represent significant progress in expanding the genetic toolkit available for studying archaeal biology.

Exploring the unique cell biology of archaea presents a vibrant and exciting new avenue of research, albeit with many challenges. For example, the study of archaeal cell shape maintenance is particularly complex due to pleomorphic nature of some haloarchaeal species. This complexity requires a rigorous

survey of cell shapes changes that may solely arise from the experimental conditions, as shown in *Haloferax volcanii* (Patro et al.). Furthermore, the specific membrane lipid composition, growth temperature and pH characteristics typical of archaeal thermophiles, such as *Sulfolobus*, have posed significant challenges to imaging techniques. To address this, a vast repertoire of membrane and DNA dyes, commonly used to image bacteria and eukaryotes was tested in *Sulfolobus acidocaldarius*. This study serves as a crucial reference for future imaging studies in archaea, providing valuable insights into the selection of appropriate imaging tools (Cezanne et al.).

To finalize this editorial, we encourage all readers to explore the exciting discoveries and advances in the Molecular Biology of Archaea that we have collected in this Research Topic. The archaeal community is expanding, and this Research Topic underscores how new hypotheses are continuously being proposed and ongoing research is set to reveal the unique biology of archaea. After the success of this Research Topic linked to the *Molecular Biology of Archaea 2022*, we are eagerly anticipating the upcoming Molecular Biology of Archaea meeting, which will be held in Paris this coming June.

Author contributions

MvW: Writing – review & editing, Writing – original draft, Conceptualization. SI: Writing – review & editing, Writing – original draft, Conceptualization. MS-M: Writing – review & editing, Writing – original draft, Conceptualization. AC-O: Writing – review & editing, Writing – original draft, Conceptualization.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. MvW was supported by a Momentum grant (VW Foundation grant

number 94933 to Sonja-Verena Albers). MS-M was supported by MCB StartUp funds (4171510) from CLAS-UConn. AC-O was funded by the National Institute of General Medical Sciences of the National Institutes of Health (R01-GM061010 to Dyche Mullins) and by the Howard Hughes Medical Institute Investigator Program (Dyche Mullins). SI was supported by a Canada 150 Research Chair in bacterial cell biology to Yves Brun.

Acknowledgments

We thank the organizers of The Archaea Power Hour (<https://www.archaea.page/>) and Archaea.bio (<https://www.archaea.bio/>) for their efforts in increasing the visibility of this Research Topic. Additionally we would like to thank all organizers of the *Molecular Biology of Archaea – 2022* meeting: Prof. Jörg Soppa, Prof. Sonja-Verena Albers, and Prof. Anita Marchfelder, for their support and trust in enabling us to edit this Research Topic linked to their meeting. This Research Topic is dedicated to all scientists who have contributed for many years to the advancements in archaea research.

Conflict of interest

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.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Baker, B. J., De Anda, V., Seitz, K. W., Dombrowski, N., Santoro, A. E., and Lloyd, K. G. (2020). Diversity, ecology and evolution of Archaea. *Nat. Microbiol.* 5, 887–900. doi: 10.1038/s41564-020-0715-z
- Hoegenauer, C., Hammer, H. F., Mahnert, A., and Moissl-Eichinger, C. (2022). Methanogenic archaea in the human gastrointestinal tract. *Nat. Rev. Gastroenterol. Hepatol.* 19, 805–813. doi: 10.1038/s41575-022-00673-z
- Lyu, Z., Shao, N., Akinyemi, T., and Whitman, W. B. (2018). Methanogenesis. *Curr. Biol.* 28, R727–R732. doi: 10.1016/j.cub.2018.05.021
- Pfeifer, K., Ergal, I., Koller, M., Basen, M., Schuster, B., and Rittmann, S. K.-M. R. (2021). Archaea Biotechnology. *Biotechnol. Adv.* 47:107668. doi: 10.1016/j.biotechadv.2020.107668
- Spang, A., Caceres, E. F., and Ettema, T. J. G. (2017). Genomic exploration of the diversity, ecology, and evolution of the archaeal domain of life. *Science* 357:eaaf3883. doi: 10.1126/science.aaf3883
- van Wolferen, M., Pulschen, A. A., Baum, B., Gribaldo, S., and Albers, S.-V. (2022). The cell biology of archaea. *Nat. Microbiol.* 7, 1744–1755. doi: 10.1038/s41564-022-01215-8



OPEN ACCESS

EDITED BY

Marleen van Wolferen,
University of Freiburg, Germany

REVIEWED BY

Xu Feng,
Shandong University, China
Issay Narumi,
Toyo University, Japan
Susanne Erdmann,
Max Planck Society, Germany

*CORRESPONDENCE

Rachel J. Whitaker
✉ rwhitakr@illinois.edu

SPECIALTY SECTION

This article was submitted to
Biology of Archaea,
a section of the journal
Frontiers in Microbiology

RECEIVED 02 December 2022

ACCEPTED 09 January 2023

PUBLISHED 23 January 2023

CITATION

Sanchez-Nieves RL, Zhang C and Whitaker RJ
(2023) Integrated conjugative plasmid drives
high frequency chromosomal gene transfer
in *Sulfolobus islandicus*.
Front. Microbiol. 14:1114574.
doi: 10.3389/fmicb.2023.1114574

COPYRIGHT

© 2023 Sanchez-Nieves, Zhang and Whitaker.
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.

Integrated conjugative plasmid drives high frequency chromosomal gene transfer in *Sulfolobus islandicus*

Ruben L. Sanchez-Nieves^{1,2}, Changyi Zhang¹ and
Rachel J. Whitaker^{1,2*}

¹Carl R. Woese Institute for Genomic Biology, University of Illinois Urbana-Champaign, Urbana, IL,
United States, ²Department of Microbiology, University of Illinois Urbana-Champaign, Urbana, IL,
United States

Gene transfer in crenarchaea has been observed within natural and experimental populations of *Sulfolobus*. However, the molecular factors that govern how gene transfer and recombination manifest themselves in these populations is still unknown. In this study, we examine a plasmid-mediated mechanism of gene transfer in *S. islandicus* that results in localized high frequency recombination within the chromosome. Through chromosomal marker exchange assays with defined donors and recipients, we find that while bidirectional exchange occurs among all cells, those possessing the integrated conjugative plasmid, pM164, mobilize a nearby locus at a significantly higher frequency when compared to a more distal marker. We establish that *traG* is essential for this phenotype and that high frequency recombination can be replicated in transconjugants after plasmid transfer. Mapping recombinants through genomic analysis, we establish the distribution of recombinant tracts with decreasing frequency at increasing distance from pM164. We suggest the bias in transfer is a result of an Hfr (high frequency recombination)-like conjugation mechanism in this strain. In addition, we find recombinants containing distal non-selected recombination events, potentially mediated by a different host-encoded marker exchange (ME) mechanism.

KEYWORDS

recombination, gene transfer, conjugative plasmid, *Sulfolobus islandicus*, archaea, conjugation frequency

1. Introduction

The hyperthermophilic *Sulfolobales* are the genetically tractable model system of the crenarchaea, and the model archaeon that shares a most recent common ancestor with eukaryotes. Mechanisms of genetic exchange and recombination are not well understood in the Crenarchaea; however, they are important molecular components that may hold a key to understanding the origin of meiosis and eukaryogenesis (Bernstein et al., 1985; Gross and Bhattacharya, 2010). The *Sulfolobales* have been shown to exchange genetic markers when complementary auxotrophic mutants are co-incubated and plated on selective media. Previous studies showed that selected markers were transferred during marker exchange (ME) at relatively equal frequencies for both auxotrophs (Grogan, 1996). Further work showed that selected marker transfer occurs in small discontinuous patches that are dependent on minimal homology (Hansen et al., 2005). In most studies, ME is up-regulated with DNA damaging agent exposure

leading to the current hypothesis that ME is a mechanism for DNA repair (Schmidt et al., 1999; Ajon et al., 2011; van Wolferen et al., 2015, 2016, 2020; Feng et al., 2018; Schult et al., 2018; Sun et al., 2018). It has been experimentally shown that in *Sulfolobus acidocaldarius*, exposure to DNA damaging agents results in induced type IV pilin-mediated cell aggregation (van Wolferen et al., 2016; Sun et al., 2018). Recently, it has been shown that S-layer glycosylation patterns and a specific sequence of the UpsA proteins are responsible for species-specific aggregation and ME (van Wolferen et al., 2020). In addition, a novel DNA uptake system (Ced system) that is also UV inducible has been shown to be essential for this process (van Wolferen et al., 2015). Still, the mechanism by which DNA is moved from donor to recipient cell is unknown. It has also been hypothesized that integrated conjugative elements and their transfer systems may be responsible for chromosomal DNA mobilization (She et al., 2004; Chen et al., 2005).

Sulfolobaceae possess the only experimentally characterized conjugative plasmids in Archaea (Schleper et al., 1995; Prangishvili et al., 1998; Lipps, 2006). Most of the plasmid-containing strains are from the *S. islandicus* species (Prangishvili et al., 1998; Greve et al., 2004). Conjugative plasmids and other mobile genetic elements (MGEs) integrate at tRNA sites and can be stably maintained within the chromosome (Lipps, 2006; Cadillo-Quiroz et al., 2012). Transfer of episomal plasmids, through conjugation between *Sulfolobus* spp., can be readily achieved in the lab through co-incubation in liquid media under shaking conditions (Prangishvili et al., 1998). Conjugative plasmids in *Sulfolobus*, described as pNOB8-like plasmids, usually possess an origin of replication, a *repA* homolog (implicated in replication), *parAB* partitioning homologs, a highly conserved *plra* gene (implicated in plasmid regulation), a site-specific integrase, and transfer gene homologs *trbE* and *traG*, among other variable open reading frames (ORFs) (Stedman et al., 2000; Greve et al., 2004; She et al., 2004). Sequence alignments between pNOB8-like transfer genes, such as *trbE* and *traG*, have also shown that these genes are among the most conserved between plasmids (She et al., 2001). pNOB8-like plasmids can sometimes produce defective smaller plasmids that lack their transfer genes and can only be maintained in the presence of an intact copy, denoting the importance of these genes for mobilization (Stedman et al., 2000). Although molecular studies into the components of *Sulfolobus* plasmid transfer machinery have not been undertaken, studies in bacterial plasmids, such as the F plasmid in *Escherichia coli* and the Ti plasmid of *Agrobacterium* spp., have shown that TraG and TrbE, respectively are type IV secretion system conjugation components which form an essential function of the mate-pair formation (mpf) machinery and thus conjugation itself (Firth and Skurray, 1992; Li et al., 1999). The F plasmid in *E. coli*, specifically, has shown the ability to recombine into the chromosome and transfer chromosomal genes in a linear fashion from the origin of transfer (Hfr) (Lloyd and Buckman, 1995).

At a population level, there is evidence for chromosomal gene exchange and episomal plasmid transfer in *Sulfolobus* spp. (Schleper et al., 1995; Cadillo-Quiroz et al., 2012). Previous studies from isolate genomes of *S. islandicus* from a single hot spring in Kamchatka, Russia, have shown that closely related groups can maintain a level of gene flow which is higher within each group than it is between them, fitting the biological species definition. There is populational genomics evidence for differences in recombination between two “species”, with the highest exchange rates among strains in the “Red” group between isolates M.16.4 and M.16.40, although analysis

of proximal regions to recombinant tracts suggests this has no relationship to regional homology (Cadillo-Quiroz et al., 2012). Genome analysis shows there are differences in recombination and structural changes in certain parts of the chromosome (Krause and Whitaker, 2015). It is not known whether this chromosomal architecture is a function of mechanisms of gene transfer and recombination or a function of differences in selective forces removing or maintaining variation in different regions.

In this study we show that the presence of an integrated conjugative plasmid, pM164, increases the frequency of marker exchange near its integration site in a manner that is dependent on the plasmid encoded TraG. Because high frequency of recombination occurs surrounding the plasmid integration site, we hypothesize that an Hfr-like (high frequency recombination) mechanism led to the bias in recombination frequency near the integrated element while background levels of recombination occur in other locations.

2. Materials and methods

2.1. *Sulfolobus* strains and growth conditions

S. islandicus M.16.4 and its derivatives described in Table 1, were grown in tissue culture flasks (Falcon) at 73°C to 76°C without shaking. DY (dextrin-tryptone) media (pH 3.45) was used in all cases and contained the following components (per 1 L of Milli-Q H₂O): basal salts (K₂SO₄, 3.0 g; NaH₂PO₄, 0.5 g; MgSO₄, 0.145 g; CaCl₂ · 2H₂O, 0.1 g), 20 µl trace mineral stock solution (3.0% FeCl₃, 0.5% CoCl₂ · 6H₂O, 0.5% MnCl₂ · 4H₂O, 0.5% ZnCl₂, and 0.5% CuCl₂ · 2H₂O), 0.2% (wt/vol) dextrin, and 0.1% (wt/vol) tryptone. Plate media was prepared by pre-warming 2x DY media, supplemented with 20 mM of MgSO₄ and 7 mM CaCl₂ · H₂O, and mixing an equal volume of a freshly boiling 1.7% (wt/vol) Gelrite solution. To grow uracil and agmatine auxotrophic strains, DY media was supplemented with a final concentration of 20 µg/ml of uracil and 50 µg/ml of agmatine. For 5-FOA counterselections, a final concentration of 50 µg/ml was also added to DY plate media. For general molecular cloning manipulations *Escherichia coli* (NEB 5-alpha competent *E. coli*) was utilized on LB media. Ampicillin (100 µg/ml) was added to media when required.

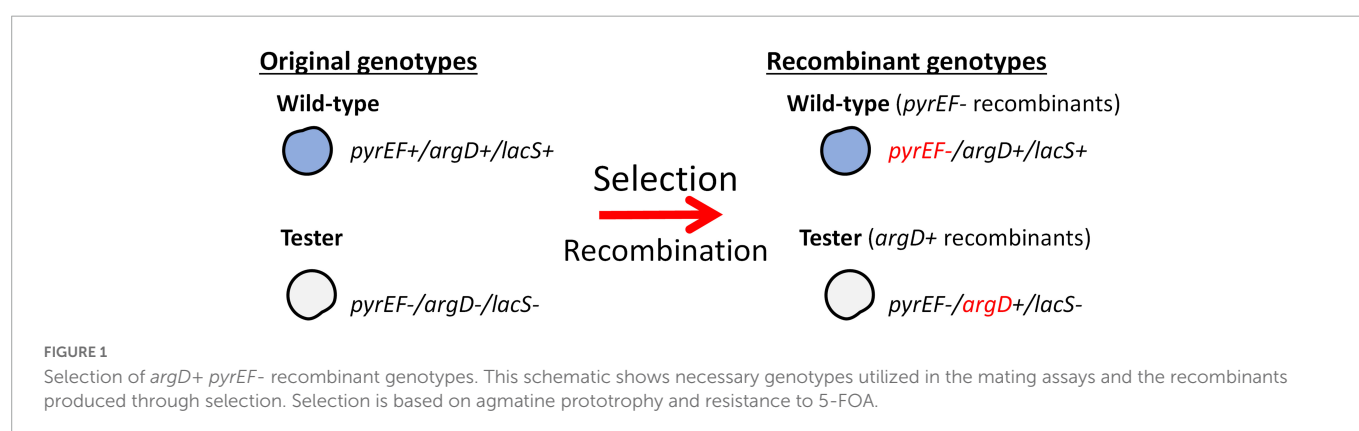
2.2. Mating assay

The mating assays for *S. islandicus* strains were performed through selection of agmatine prototrophy and 5-FOA resistance. Two genotypes are needed in order to perform the mating assay: a strain with functional *lacS*, *argD*, and *pyrEF* loci, and a strain with deletions in these three markers (Figure 1).

In all mating assays, both partner strains were grown to a mid-log phase, as detailed above, in either DY media or DY media supplemented with uracil and agmatine. Once strains reached their mid-log phase, their optical density (600 nm) was measured and equalized using warm DY media to an OD₆₀₀ of 0.3. The two strains were then mixed in equal parts and incubated at room temperature for 15 min. The co-incubates were then incubated under shaking conditions (180 rpm) at 76°C overnight. Following incubation, the mixed cultures were washed to remove agmatine using centrifugation

TABLE 1 Strains and plasmids used in this study.

Strain	Genotype/Feature	References
<i>S. islandicus</i> M.16.4 (Red group)	Wild type (<i>argD</i> ⁺ <i>pyrEF</i> ⁺ <i>lacS</i> ⁺)	Reno et al., 2009 PNAS
<i>S. islandicus</i> M.16.40 (Red group)	Wild type (<i>argD</i> ⁺ <i>pyrEF</i> ⁺ <i>lacS</i> ⁺)	Cadillo-Quiroz et al., 2012
<i>S. islandicus</i> M.16.2 (Red group)	Wild type (<i>argD</i> ⁺ <i>pyrEF</i> ⁺ <i>lacS</i> ⁺)	Cadillo-Quiroz et al., 2012
<i>S. islandicus</i> M.16.27 (Blue Group)	Wild type (<i>argD</i> ⁺ <i>pyrEF</i> ⁺ <i>lacS</i> ⁺)	Reno et al., 2009 PNAS
RJW004	M.16.4Δ <i>argD</i> Δ <i>pyrEF</i> Δ <i>lacS</i>	Zhang et al., 2013a
Δ <i>traG</i> -RJW004	Δ <i>argD</i> :: <i>argD</i> Δ <i>pyrEF</i> :: <i>pyrEF</i> Δ <i>lacS</i> Δ <i>traG</i>	This Study
RJW007	M.16.4Δ <i>argD</i> Δ <i>pyrEF</i> <i>lacS</i> ⁺	Zhang and Whitaker, 2018
<i>S. islandicus</i> M.16.2p	M.16.2 harboring conjugative plasmid pM164	This study
Plasmid		
pSe-RP-StoargD- <i>traG</i>	pSeSd-StoargD with <i>S. islandicus</i> M.16.4 CRISPR repeats and <i>traG</i> spacer + HR template; created through NEB Hifi assembly	This Study



and equalized to an OD₆₀₀ of 0.5. Finally, cultures underwent serial dilutions and were plated in either selective DY media containing 5-FOA and uracil (10⁰ and 10⁻¹ dilutions) or non-selective DY media containing agmatine and uracil (10⁻⁴ and 10⁻⁵ dilutions). After cells are plated, plates are incubated for 7–14 days at 76°C. Once colonies are observed on plates, an X-gal solution (2 mg/ml) was sprayed on both selective and control plates to determine original background. After the genotype is confirmed through PCR, the colonies are counted and analyzed to determine the recombination frequency of both the tester strain and either the wild-type or mutants with all three genetic markers.

Recombinant frequencies are calculated through the following formula: $\text{recombinant frequency} = \frac{\text{recombinantsG1}}{\text{controlG1}}$, where *recombinantsG1* represents the cfu/ml of colonies on selective media of a particular genotype. *ControlG1* represents the cfu/ml of colonies counted on non-selective media of the same genotype, G1. Colonies are verified through PCR amplification using primers: *argD_chk-F/R* & *M16_pyrEFII-F/R* (Supplementary Table 1).

2.3. Construction of *traG* in-frame deletion plasmids

Gene knockout shuttle vectors were constructed using NEBuilder Hifi assembly kit (NEB, USA) and according to manufacturer's protocol. IDT G-Blocks were synthesized for the *traG* spacer and recombination template. The shuttle vectors were cloned in NEB 5-alpha competent *E. coli* (High Efficiency; NEB, USA) cells and plated

on LB with ampicillin (100 μg/ml). Shuttle vectors were subsequently isolated using the QIAprep Spin Miniprep Kit (Qiagen, USA) as per manufacturer's instructions.

2.4. Genetic manipulation of *S. islandicus* strains

Once knockout plasmids were isolated, plasmids were electroporated into competent RJW004 cells and incubated in incubation buffer for 1 hr (Zhang et al., 2013a). Transformations were then plated onto DY plates and incubated for 10–14 days. After colonies appeared on media, X-gal staining and PCR amplification was performed on colonies to confirm vector presence in the strain. Once colonies were confirmed, they were grown in liquid DY media to mid-log phase.

These strains were subsequently plated onto counter-selective DY media containing uracil, agmatine, and 5-FOA at concentrations mentioned above. Plates were incubated for 10–14 days. After incubation, colonies were stained with X-gal and picked for PCR amplification of target genes. Subsequently verified colonies were streaked for purification and subsequent DNA extraction and whole-genome sequencing.

Further manipulation of RJW004-Δ*traG* was needed to obtain the proper genotype for mating assays (Figure 4). The *argD*, *pyrE*, and *pyrF* genes were inserted through purification and electroporation of M.16.4-derived amplicons (Zhang and Whitaker, 2018). Because the source is isogenic to RJW004, no homology

modifications were introduced into the amplification of the genes. Enrichment for 14 days was necessary for efficient isolation of *pyrEF*⁺ colonies. Genomic DNA was extracted and sent to sequencing for further verification.

2.5. Conjugation of pM164

Conjugation of pM164 into *S. islandicus* M.16.2 was performed according to Prangishvili et al. (1998) with minor modifications and a selection scheme (Supplementary Figure 1). In general, *S. islandicus* RJW004 was co-incubated with *S. islandicus* M.16.2 at a donor:recipient ratio of 1:10,000 for 48 h under shaking conditions in liquid DY media containing uracil and agmatine at 76°C. After incubation, the co-culture was diluted using serial dilutions to 10⁻³ of the original incubation and plated on DY media to select against RJW004. After incubation for 10 days, colonies were picked and verified for pM164 in the chromosome and original background through PCR amplification of pM164 region and the strain verification, respectively. M.16.2p was then purified through streaking and sent to sequencing using Illumina short reads. Additionally, circularization of pM164 was assessed through amplification of plasmid distal regions (Supplementary Figure 3).

2.6. Sequencing SNP analysis and bioinformatics

SNP analyses were performed on recombinant isolates from three independent tester/M.16.2p mating assays. Colonies were isolated and transferred to liquid DY media until mid-log phase. Subsequently, DNA extraction and Illumina short read sequencing were performed on 30 RJW004 recombinants. Sequencing files were then analyzed for M.16.2p-specific donor SNPs through Breseq analysis (Deatherage and Barrick, 2014). SNPs were then mapped onto an M.16.4 reference for analysis. Alignment for Figure 3 was performed using Clinker (Gilchrist and Chooi, 2021). Variant positions in core genome were assessed using Spine and Nucmer (Kurtz et al., 2004; Ozer et al., 2014).

2.7. Statistical analysis

Statistical analysis was performed on mating assays in all cases. Comparisons were done utilizing ANOVA when experiments involved more than two assays. In cases where two assays were compared a *t*-test was performed to assess significant differences. In all cases *p*-value-0.05 is utilized.

3. Results

3.1. Mutnovsky Red group show varying levels of recombination frequencies

The natural isolates utilized in this study were isolated from the Mutnovsky region in Kamchatka Russia in 2010. We focus on one of two closely related but distinct groups (the Red group) observed to be

diverging over time in a manner that represents incipient speciation from within the same hot spring (Cadillo-Quiroz et al., 2012). The Red group is composed of three natural isolates which are closely related and have shown evidence of high levels of recombination between them. At the nucleotide level, the Red group possess 6,052 variant positions across core genome and an ANI (average nucleotide identity) of above 99.75% in all cases. We include one Blue group strain, M.16.27, for comparison of transfer among incipient species.

To test recombination frequencies between these strains we utilized a tester strain, RJW004, which is an M.16.4 background strain possessing three gene deletions ($\Delta argD$, $\Delta pyrEF$ and $\Delta lacS$). This strain is crossed with wild-type strains in mating assays described in Figure 1 (Zhang et al., 2013a,b). The *lacS* gene is not selected in this assay but used to differentiate donor and recipient genome backgrounds on selective plates through X-gal staining. Selection on uracil/5-FOA (5-Fluoroorotic acid) containing plates without agmatine selects for strains in wild type background (blue) that receive the $\Delta pyrEF$ deletion locus from the tester (*argD*⁺*pyrEF*⁻*lacS*⁺), and the tester strain (white) that receives the *argD* locus from the WT (*argD*⁺*pyrEF*⁻*lacS*⁻) (Figure 1). In the M.16.4 strain the selected markers ($\Delta pyrEF$, *argD*) are separated by 62Kb. To differentiate deletion from spontaneous mutations in each experiment WT and tester strains are plated on selective plates and the number of spontaneous mutations is subtracted from total.

Figure 1 shows bi-directional recombination when the tester strain is crossed with WT strains from the same population. Figure 2 shows the recombination profiles of the above-mentioned natural isolates for our two selective markers. The recombination frequency of the $\Delta pyrEF$ locus from the tester to WT strains is consistent, independent of the WT background, yet slightly higher in the cross with M.16.27. In contrast, *argD*⁺ recombinants showed substantial variation with M.16.4 having the highest recombination frequency, M.16.27 having the lowest frequency, and the M.16.2 assay possessing the lowest donor frequency of the red isolates (Figure 2).

3.2. pM164 conjugation and transfer play a role in high frequency recombination

Interestingly, an integrated conjugative plasmid, named pM164, was found 8086bp from the highly transferred *argD* marker in M.16.4 and M.16.40 (Figure 3). PCR amplification has also shown evidence of excision in M.16.4 along with the integrated version, providing evidence that pM164 is an active pNOB8-like element. Because the *argD* recombination frequency was associated with pM164, in both M.16.4 and M.16.40, we hypothesized that pM164 may well be driving the transfer of proximal genes like *argD*. We note that M.16.27 has a divergent plasmid integrated in a different position relative to markers and relatively closer (~72 kb) to *pyrEF* than to *argD*. This plasmid, however, lacks important plasmid components such as *repA*, *parA*, and *trbE*.

In bacterial conjugation models *traG* (M164_1618), encoding for a VirD4-type conjugation component, is essential for plasmid transfer as it mediates the interactions between the mating pore and the DNA transfer system (Arends et al., 2013; Kohler et al., 2013). The *traG* gene was identified through comparisons to pNOB8 and its genomic characterization (She et al., 1998). *TraG* in *Sulfolobus* spp. contains a conjugation-specific type IV secretion system (T4SS) domain found in bacterial conjugation proteins such as VirD4, TrwB,

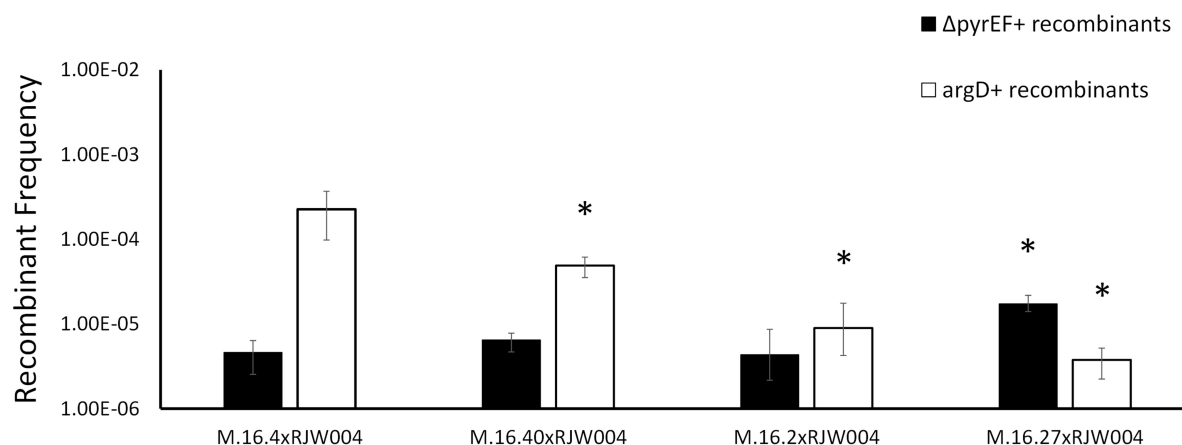


FIGURE 2

Recombination frequencies of Red species isolates and Blue species representative (M.16.27) in mating assays with the tester strain, RJW004. White bars represent the recombinant frequency of *argD* recombinants, while black bars represent the recombinant frequency of $\Delta pyrEF$ recombinants. Error bars represent highest and lowest replicate for each assay, *represents a *p*-value of <0.05 when compared to M.16.4 cross.

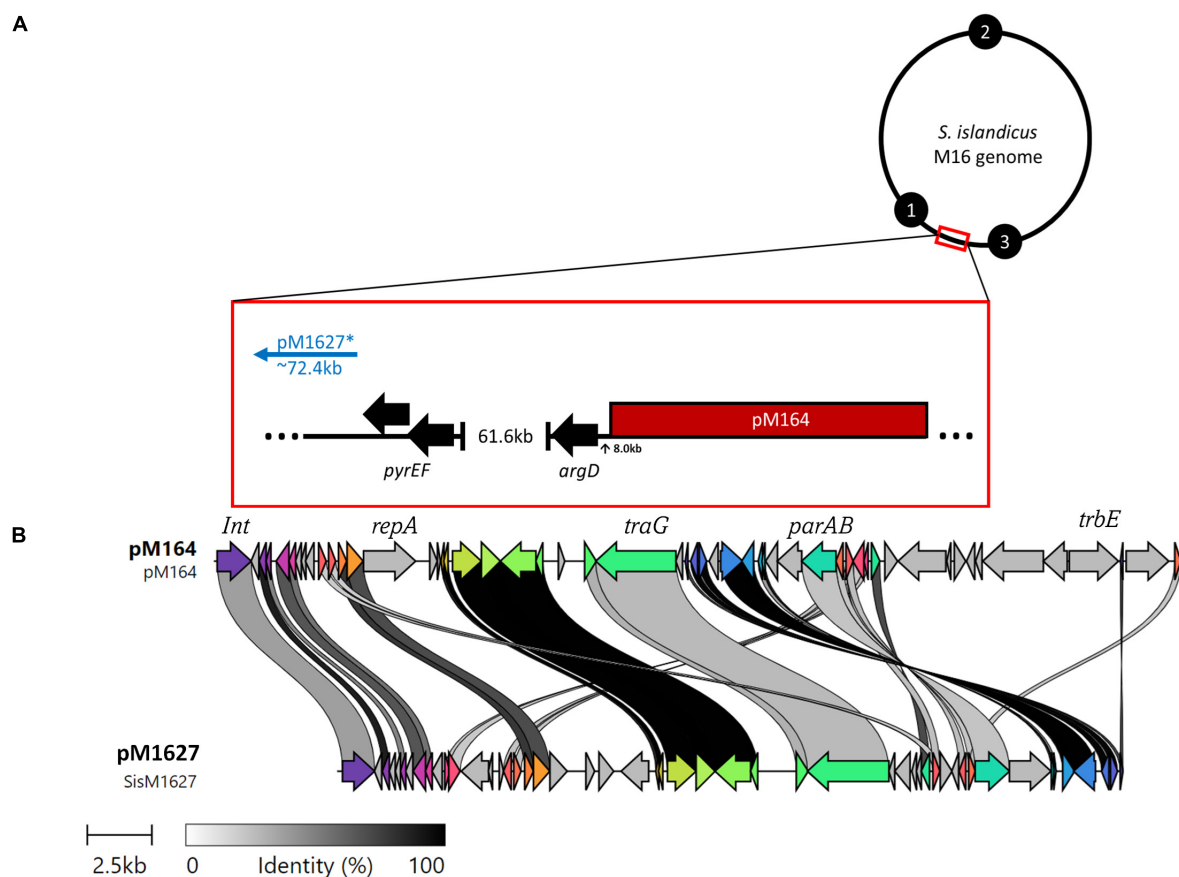


FIGURE 3

(A) M16 genomes possess broad similarities, thus the genomic position of pM164 and pM1627 (*not a complete plasmid) can be resolved by a general model in reference to the selective markers *argD* and *pyrEF*. Black numbered circles represent replication origins. (B) A clinker nucleotide alignment of pM164 and pM1627, which is missing essential plasmid components such as *repA*, *parA*, and *trbE*.

and TraG (She et al., 1998). When compared, three motifs are present that correspond to bacterial conjugation proteins such as TrwB in *E. coli* and with the archaeal HerA helicase (Supplementary Figure 2). Thus, we hypothesized that a *traG* deletion in pM164 would abrogate transfer and thus the associated high frequency

recombination of *argD*. We created a marker-less in-frame deletion of the pM164 *traG* (M164_1618) in RJW004, along with the appropriate marker insertions, to create an *argD*⁺ *pyrEF*⁺ *lacS*⁻ *traG*⁻ strain by using the endogenous CRISPR-Cas system-based genome editing (Li et al., 2015). Because this strain is *lacS*⁻, it was paired

with RJW007, an *argD*⁻*pyrEF*⁻*lacS*⁺ tester, to yield selectable recombinant genotypes (Table 1). An M.16.4 (wt)/RJW004 (tester) mating assay was performed for comparison. We found a significant decrease in *argD* recombination frequencies when *traG* was not present in the *argD* donor strain (Figure 4). Background levels of marker exchange were maintained in these crosses, however, the bias for the *argD* marker was absent.

3.3. pM164 transfer increases recombination frequency in transconjugants

We tested whether pM164 is mobilizable and can independently transfer its high recombination phenotype to a naïve strain. The red isolate, M.16.2, did not possess a conjugative plasmid but has an identical *att* site and selective marker locations, as such we chose to conjugate pM164 into M.16.2. The conjugation protocol was adapted from Prangishvili et al. (1998) and is diagrammed in Supplementary Figure 1 (see “3. Materials and methods”). First, Δ *traG*-RJW004 was utilized as a pM164 donor to test the importance of *traG* to plasmid transfer in *S. islandicus* and found that only 1/99 recipient cells possessed pM164. Thus, much like in bacteria, *traG* is essential for conjugation in *S. islandicus* (Waters et al., 1992).

Using *traG*⁺ donor we were able to obtain M.16.2 transconjugants possessing pM164. Transconjugants were verified for strain identity and the appropriate pM164 insertion event using PCR amplification of the *att* site (Supplementary Figure 1). 99% of colonies surveyed possessed pM164 in the insertion site, showing strong evidence of insertion site specificity and high efficiency of transfer (She et al., 2004). Highly efficient transfer is characteristic of conjugative plasmids in *Sulfolobus*, as has been observed previously by Prangishvili et al. (1998). Of the colonies verified to possess pM164, three were purified and sent for sequencing to further confirm the strain and the presence of pM164 in the appropriate locus.

We mated M.16.2p, a pM164 transconjugant of the parental strain M.16.2, with the naïve strain to assess the recombination frequencies of the two selectable markers (Figure 5). Through sequencing we confirmed that M.16.2p does not possess any other MGE other than pM164 and further confirmed that pM164 is integrated into the correct position. We show that high frequency recombination of *argD* follows insertion of pM164 into the appropriate tRNA-adjacent attachment site. Thus, we show direct evidence of an integrated plasmid influencing core genome recombination dynamics in Archaea. These results provide evidence that indeed pM164 and not the strain is the deciding factor in localized high frequency recombination of *argD*.

3.4. Transferred recombinant tracts suggest transfer from plasmid is unidirectional in a background of unselected recombinant tracts

To gain further understanding into the mechanism of pM164 transfer, we sequenced 10 RJW004 recombinants from three independent M.16.2p mating assays for a total of 30 isolates. Each

isolate represents independent recombination events within each experiment. We assess chromosomal regions that are transferred from M.16.2p to our tester strain, RJW004 (Figure 6A). The M.16.2p donor locus that stretches between *argD* and the pM164 integration site were differentiated from the recipient (RJW004) locus through SNP analysis (Figure 6). Because M.16.2 is highly similar to M.16.4, SNP positions scattered throughout the chromosome can be utilized as markers for recombination. We find consistent recombination tracts leading from the selected marker *argD* to pM164 in an uninterrupted manner. We also find a slope in recombinant SNPs upstream of *argD* that is not observed on the other end of pM164. From these data we infer that pM164 transfers unidirectionally while integrated in the chromosome, in a manner similar to Hfr strains in *E. coli*. In addition, we also found that 17/30 isolates possess unselected recombination events throughout the chromosome.

We find that long unselected donor tracts do not correlate with genomic distance from pM164. Thus, we hypothesize that although pM164 has a localized effect made apparent by the high frequency of unselected donor SNPs surrounding the *argD/att* region, the ME/Ced system most likely accounts for most of the intercellular mobilization of DNA throughout the chromosome. The localized high frequency effect is most likely due to unidirectional integrated plasmid transfer from a plasmid origin toward *argD*, these tracts are then resolved through homologous recombination between the chromosomes, defining the recombination tract lengths. These data, along with the essentiality of plasmid transfer through a TraG-mediated mechanism, show that in *S. islandicus* integrated plasmids can consistently and reproducibly create high-frequency marker exchange in nearby chromosomal loci.

4. Discussion

This study has uncovered evidence of plasmid-mediated high-frequency transfer of DNA in the strain M.16.4 from Kamchatka, Russia. We show that high frequency transfer requires TraG (Figure 4) and is transferable through plasmid conjugation to M.16.2, demonstrating that this mechanism depends upon the plasmid pM164. Genome sequencing of recombinants showed that tracts within the *argD* genomic context mostly extend through to the pM164 *att* site which we interpret as integrated plasmid transfer in a manner similar to Hfr. This process increases recombination frequencies unidirectionally in this region directly adjacent to *argD*. This study is the first to directly demonstrate a role for integrated plasmids in the mobilization of chromosomal loci in *Sulfolobus*.

In this work the ME system is active under uninduced conditions, which allows for genetic selections of chromosomal markers through co-incubation and plating on selective media. This was first noted by Grogan (1996) and has been the basis for many studies including those involved in the recombination mechanics with *Sulfolobus* and the UV-up-regulated system that allows for lateral gene transfer (Grogan, 1996; Reilly and Grogan, 2002; Hansen et al., 2005; Ajon et al., 2011; Rockwood et al., 2013; Zhang et al., 2013b; van Wolferen et al., 2020, 2016). Here, we find an association of a 37 kb integrated conjugative plasmid and a highly recombining marker located 8 kb from the integration site. Interestingly, Cadillo-Quiroz et al. (2012) found that M.16.4 and M.16.40 (both contain pM164 variants) are strong DNA donors when compared to M.16.2, which does not possess an integrated plasmid. Furthermore M.16.27, from a co-existing, closely related species, contains a plasmid that does not

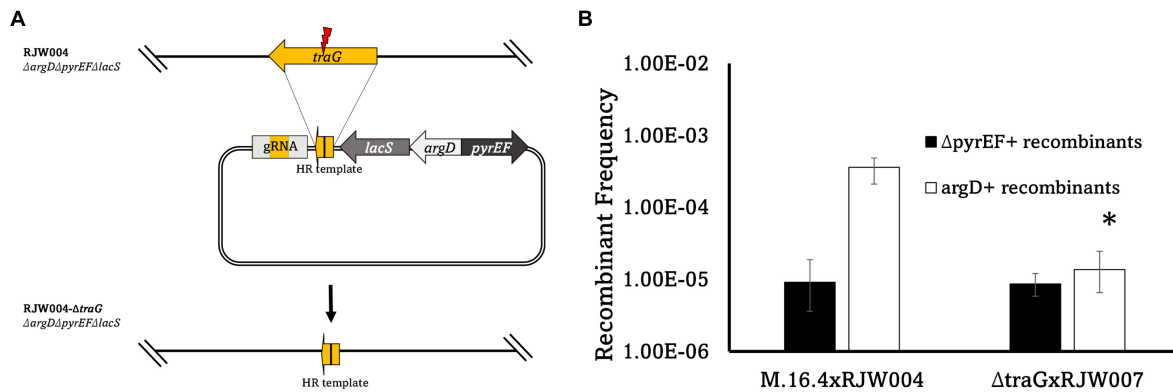


FIGURE 4

(A) Schematic diagram of *traG* gene deletion based on CRISPR-Cas system in *S. islandicus* M.16.4. The genome editing plasmid was designed by taking advantage of the native CRISPR-Cas system in *S. islandicus* to build a gRNA targeting *traG*. To avoid marker recombination, *argD* from *S. tokodaii* and *pyrEF/lacS* from *S. solfataricus* were utilized instead on native markers. (B) Recombinant frequencies of the *traG* deletion mutant compared to wild type. Error bars represent highest and lowest replicate for each assay, *represents a *p*-value of <0.05 when comparing to M.16.4 *argD* frequency.

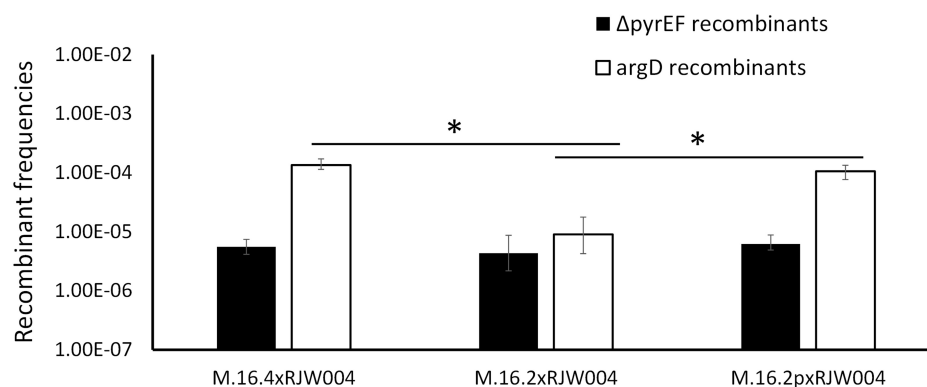


FIGURE 5

Recombination frequencies of genetic markers in M.16.2 transconjugant mating assays compared to parent strain. White bars represent the recombinant frequency of *argD*, while black bars represent the recombinant frequency of $\Delta pyrEF$. Error bars represent highest and lowest replicate for each assay, *represents a *p*-value of <0.05 when comparing to M.16.2 *argD* frequency.

confer this phenotype (Figure 2). Our findings, in conjunction with previous studies looking into defective plasmids which lose *traG*, have shown that the protein encoded by *traG* plays a major role in archaeal plasmid transfer and is essential for this mechanism, although further molecular studies on the transfer mechanism of *Sulfolobus* spp. plasmids are lacking (Stedman et al., 2000). Specifically, genetic and biochemical analyses are needed to decipher more of the transfer components and their respective functions. Both our recombination assay with the natural isolate M.16.2 (Figure 2), RJW004 $\Delta traG$ (Figure 4) and previous work in *S. acidocaldarius* show bidirectional recombination in the absence of an active plasmid, suggesting an additional host-encoded mechanism for DNA transport (Grogan, 1996; van Wolferen et al., 2016).

In *Sulfolobus* plasmids, *traG* is one of the only discernible transfer genes that can be identified through annotation as conjugation-specific proteins (the other being *trbE*). It is currently not known whether *traG* plays a similar role in *Sulfolobus* as it does in bacteria, as the molecular components of the *Sulfolobus* plasmid transfer apparatus have not been subjected to genetic or biochemical analysis in the way their bacterial counterparts have. The structure of TraG in *Sulfolobus*, however, possess similar motifs to bacterial T4SS

conjugation components, such as *E. coli* TrwB and other *Sulfolobus* pNOB8-like plasmids (Supplementary Figure 2; She et al., 1998).

Through donor SNP analysis of *argD*+ recombinant isolates we find that this high frequency phenomenon produces, in most cases, uninterrupted donor SNPs in the area between *argD*, which is under selection, and pM164, which is not (Figure 6). These results suggest that the high frequency recombination observed in *argD* is due to the co-transfer of chromosomal DNA through a TraG-mediated mechanism that is then recombined into the recipient. Because of the slope observed at only one side of pM164 and not the other, our current model is that high recombination of *argD* is driven by unidirectional transfer of pM164 in a manner similar to Hfr strains in *E. coli*. These recombination tracts, spanning at least 8 kb, are in contrast to previous recombination dynamics observed in *S. acidocaldarius* that show minimal recombination past 300 bp of distance between selective markers, showing evidence of differing mechanisms present in this study (Hansen et al., 2005).

We hypothesized that an Hfr-like mechanism through a plasmid may explain this phenomenon. Recombination frequencies for the *pyrEF* locus are in all cases $\sim 10^{-5}$ as is also shown in for the *argD* marker when pM164 is not present (Figures 2, 5). Interestingly,

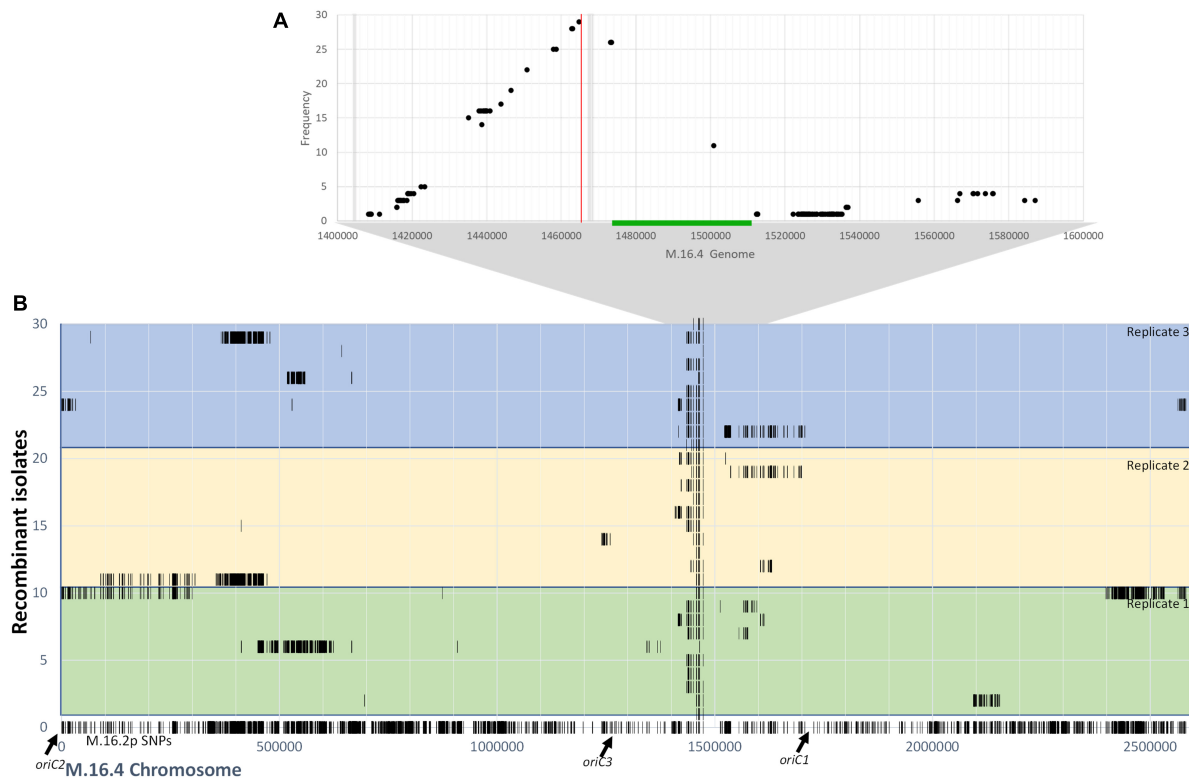


FIGURE 6

(A) Frequency of M.16.2p donor SNPs within the *argD*/pM164 region in RJW004 recombinants. The green bar represents pM164 and the red line represents the *argD* locus. Gray bar represents variable gene content between strains in M.16.2p. (B) Donor SNP distribution along the M.16.4 chromosome in RJW004 recombinants. The bottom (line 0) represents the SNP positions of M.16.2p (donor) mapped on to M.16.4 genome. The colored blocks represent each replicate, each containing 10 independent isolates.

this frequency is similar to the recombination frequency observed by Ajon et al. (2011) and Hansen et al. (2005) in *S. acidocaldarius* in induced condition. Here we find higher frequencies ($\sim 10^{-3}$) for *argD* in uninduced conditions. Similar phenomena, involving chromosomal high recombination near a mobile genetic element, have been observed in *Staphylococcus aureus* but over a much smaller scale and involving a conjugative transposon that circularizes with proximal parts of the host chromosome (Everitt et al., 2014). In this alternative model, plasmid DNA is excised imprecisely, taking flanking chromosomal regions with it, which then transfer through the plasmid-encoded transfer system and the genes in its chromosomal vicinity. Evidence of plasmid circularization has been noted in M.16.4 through plasmid end junction (Supplementary Figure 3). However, circularization would most likely encompass both distal regions of the inserted plasmid and not the unidirectional profile we find here (Figure 6). Thus, we find that two alternative but not mutually exclusive mechanisms involving plasmid transfer could be affecting chromosomal regions nearby such as *argD*.

Because recombination is occurring throughout the chromosome and the high sequence identity between isolates, it is difficult to observe the full effects that plasmid-mediated recombination has past the marker under selection. We found that SNPs were distributed mostly in two areas: (1) consistent tracts in the *argD*-pM164 region and (2) non-selected tracts that vary greatly in size and genomic location (Figure 6). These non-selected tracts although variable in length and continuous, in most cases, are not linked to the plasmid regions and are thus attributed to the broader ME/Ced system described by Grogan (1996) and van Wolferen et al. (2016).

The ME/Ced system has been demonstrated to be a UV inducible gene import system in *Sulfolobus* spp. that do not carry a pM164-like plasmid (Grogan, 1996; Ajon et al., 2011). Whether there is a compatible donor mechanism (plasmid or otherwise) is not known. Although it has not been tested explicitly, transfer of chromosomal DNA through the Ced system has been hypothesized to be responsible for unselected ME observed genetically (Hansen et al., 2005). The Ced system is present in *S. islandicus*, therefore, if the Ced mediated uptake of chromosomal DNA is correct, we suggest that the transfer of unselected tracts in our study may be happening through this alternate system. However, from the current set of experiments, we cannot exclude the possibility that the entire chromosome is transferred through an Hfr-like conjugation mechanism then recombined in patches outside of the selected region. The potential existence of multiple systems together implies that in nature portions of the chromosome, due to proximity to an integrated plasmid, will evolve differently and potentially more rapidly than others. Conjugative plasmids in *Sulfolobus* spp. are known to transfer between species, thus this mechanism lowers the barrier for DNA to be introduced in a population (Prangishvili et al., 1998). This is in contrast to the ME/Ced system which is known to be species-specific due to surface layer interactions (van Wolferen et al., 2020). Our findings show the effects of MGEs on M.16-type chromosome recombination dynamics, which were first observed in nature and then in the lab (Cadillo-Quiroz et al., 2012). We find that the transfer of pM164, an integrated conjugative plasmid, causes high frequency recombination of nearby chromosomal regions. How chromosomal DNA travels between cells is still largely unknown. In order to truly

understand the effects of MGEs on chromosome dynamics more studies are needed to further elucidate the mechanism by which DNA is transferred in *Sulfolobus* spp. more generally. Particularly, more genetic and biochemical approaches are needed to both understand archaeal biology in reference to MGEs and their cell biology as a whole. This study provides evidence that a relationship between mechanisms exists, and that this relationship has evolutionary implications within populations of *S. islandicus*.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA907032>.

Author contributions

RS-N and RW designed the research and drafted the manuscript. RS-N conducted experiments and performed data analysis and completed mutant construction and evaluation. CZ contributed to the experiment design and provided published strains. RS-N, CZ, and RW contributed to revising the manuscript. All authors made fundamental contributions to the manuscript and approved the final manuscript.

Funding

This work was funded mainly by a grant from the Division of Environmental Biology (DEB) (1355171 to RW), U.S. National Science Foundation. This work was also funded in part by the

National Aeronautics and Space Administration (NASA) through the NASA Astrobiology Institute under cooperative agreement no. NNA13AA91A, issued through the Science Mission Directorate.

Acknowledgments

The authors thank Alan Collins for support in Bioinformatic analyses. The authors also thank Jaya Chandrashekar for sequencing library prep and sample submission.

Conflict of interest

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.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2023.1114574/full#supplementary-material>

References

- Ajon, M., Fröls, S., van Wolferen, M., Stoecker, K., Teichmann, D., Driessen, A. J. M., et al. (2011). UV-inducible DNA exchange in hyperthermophilic archaea mediated by type IV pili. *Mol. Microbiol.* 82, 807–817. doi: 10.1111/j.1365-2958.2011.07861.x
- Arends, K., Celik, E. K., Probst, I., Goessweiner-Mohr, N., Fercher, C., Grumet, L., et al. (2013). TraG encoded by the pip501 type IV Secretion system is a Two-Domain Peptidoglycan-Degrading enzyme essential for conjugative transfer. *J. Bacteriol.* 195, 4436–4444. doi: 10.1128/JB.02263-12
- Bernstein, H., Byerly, H., Hopf, F., and Michod, R. (1985). Genetic damage, mutation, and the evolution of sex. *Science* 229, 1277–1281. doi: 10.1126/science.3898363
- Cadillo-Quiroz, H., Didelot, X., Held, N. L., Herrera, A., Darling, A., Reno, M. L., et al. (2012). Patterns of gene flow define species of thermophilic archaea. *PLoS Biol.* 10:e1001265. doi: 10.1371/journal.pbio.1001265
- Chen, L., Kim, B., Marie, S., Peter, R., Qunxin, S., Elfars, T., et al. (2005). The Genome of *Sulfolobus acidocaldarius*, a Model Organism of the Crenarchaeota. *J. Bacteriol.* 187, 4992–4999. doi: 10.1128/JB.187.14.4992
- Deatherage, D. E., and Barrick, J. E. (2014). Identification of mutations in laboratory-evolved microbes from next-generation sequencing data using breseq. *Methods Mol. Biol.* 1151, 165–188. doi: 10.1007/978-1-4939-0554-6_12
- Everitt, R. G., Didelot, X., Batty, E. M., Miller, R. R., Knox, K., Young, B. C., et al. (2014). Mobile elements drive recombination hotspots in the core genome of *Staphylococcus aureus*. *Nat. Commun.* 5:3956. doi: 10.1038/ncomms4956
- Feng, X., Sun, M., Han, W., Liang, Y. X., and She, Q. (2018). A transcriptional factor B paralogs functions as an activator to DNA damage-responsive expression in archaea. *Nucleic Acids Res.* 46, 7085–7096. doi: 10.1093/nar/gky236
- Firth, N., and Skurray, R. (1992). Characterization of the F plasmid bifunctional conjugation gene, traG. *Mol. Gen. Genet.* 232, 145–153. doi: 10.1007/BF00299147
- Gilchrist, C. L. M., and Chooi, Y. H. (2021). Clinker & clustermap.js: Automatic generation of gene cluster comparison figures. *Bioinformatics* 37, 2473–2475. doi: 10.1093/bioinformatics/btab007
- Greve, B., Jensen, S., Brügger, K., Zillig, W., and Garrett, R. (2004). Genomic comparison of archaeal conjugative plasmids from *Sulfolobus*. *Archaea* 1, 231–239. doi: 10.1155/2004/151926
- Grogan, D. W. (1996). Exchange of genetic markers at extremely high temperatures in the archaeon *Sulfolobus acidocaldarius*. *J. Bacteriol.* 178, 3207–3211.
- Gross, J., and Bhattacharya, D. (2010). Uniting sex and eukaryote origins in an emerging oxygenic world. *Biol. Direct* 5:53. doi: 10.1186/1745-6150-5-53
- Hansen, J. E., Dill, A. C., and Grogan, D. W. (2005). Conjugational genetic exchange in the hyperthermophilic archaeon *Sulfolobus acidocaldarius*: Intragenic recombination with minimal dependence on marker separation. *J. Bacteriol.* 187, 805–809. doi: 10.1128/JB.187.2.805-809.2005
- Kohler, P. L., Chan, Y. A., Hackett, K. T., Turner, N., Holly, L. H., Cloud-Hansen, C., et al. (2013). Mating pair formation homologue traG is a variable membrane protein essential for contact-independent type IV secretion of chromosomal DNA by *Neisseria gonorrhoeae*. *J. Bacteriol.* 195, 1666–1679. doi: 10.1128/JB.02098-12

- Krause, D. J., and Whitaker, R. J. (2015). Inferring speciation processes from patterns of natural variation in microbial genomes. *Syst. Biol.* 64, 926–935.
- Kurtz, S., Phillippy, A., Delcher, A., 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
- Li, P., Hwang, I., Miyagi, H., True, H., and Farrand, S. (1999). Essential components of the Ti plasmid trb system, a type IV macromolecular transporter. *J. Bacteriol.* 181, 5033–5041. doi: 10.1128/JB.181.16.5033-5041.1999
- Li, Y., Pan, S., Zhang, Y., Ren, M., Feng, M., Peng, N., et al. (2015). Harnessing type I and type III CRISPR-Cas systems for genome editing. *Nucleic Acids Res.* 44:e34. doi: 10.1093/nar/gkv1044
- Lipps, G. (2006). Plasmids and viruses of the thermoacidophilic crenarchaeote *Sulfolobus*. *Extremophiles* 10, 17–28. doi: 10.1007/s00792-005-0492-x
- Lloyd, R. G., and Buckman, C. (1995). Conjugational recombination in *Escherichia coli*: Genetic analysis of recombinant formation in Hfr x F⁻ crosses. *Genetics* 139, 1123–1148. doi: 10.1093/genetics/139.3.1123
- Ozer, E. A., Allen, J. P., and Hauser, A. R. (2014). Characterization of the core and accessory genomes of *Pseudomonas aeruginosa* using bioinformatic tools Spine and AGEnt. *BMC Genomics* 15:737. doi: 10.1186/1471-2164-15-737
- Prangishvili, D., Albers, S. V., Holz, I., Arnold, H. P., Stedman, K., Klein, T., et al. (1998). Conjugation in archaea: Frequent occurrence of conjugative plasmids in *Sulfolobus*. *Plasmid* 40, 190–202. doi: 10.1006/plas.1998.1363
- Reilly, M. S., and Grogan, D. W. (2002). Biological effects of DNA damage in the hyperthermophilic archaeon *Sulfolobus acidocaldarius*. *FEMS Microbiol. Lett.* 208, 29–34. doi: 10.1016/S0378-1097(01)00575-4
- Reno, M. L., Held, N. L., Fields, C. J., Burke, P. V., and Whitaker, R. J. (2009). Biogeography of the *Sulfolobus islandicus* pan-genome. *Proc. Natl. Acad. Sci. U.S.A.* 106, 8605–8610. doi: 10.1073/pnas.0808945106
- Rockwood, J., Mao, D., and Grogan, D. W. (2013). Homologous recombination in the archaeon *Sulfolobus acidocaldarius*: Effects of DNA substrates and mechanistic implications. *Microbiology* 159, 1888–1899. doi: 10.1099/mic.0.067942-0
- Schleper, C., Holz, I., Janekovic, D., Murphy, J., and Zillig, W. (1995). A multicopy plasmid of the extremely thermophilic archaeon *Sulfolobus* effects its transfer to recipients by mating. *J. Bacteriol.* 177, 4417–4426. doi: 10.1128/jb.177.15.4417-4426.1995
- Schmidt, K. J., Beck, K. E., and Grogan, D. W. (1999). UV stimulation of chromosomal marker exchange in *Sulfolobus acidocaldarius*: Implications for DNA repair, conjugation and homologous recombination at extremely high temperatures. *Genetics* 152, 1407–1415. doi: 10.1093/genetics/152.4.1407
- Schult, F., Le, T. N., Albersmeier, A., Rauch, B., Blumenkamp, P., Van Der Does, C., et al. (2018). Effect of UV irradiation on *Sulfolobus acidocaldarius* and involvement of the general transcription factor TFB3 in the early UV response. *Nucleic Acids Res.* 46, 7179–7192. doi: 10.1093/nar/gky527
- She, Q., Peng, X., Zillig, W., and Garrett, R. A. (2001). Gene capture in archaeal chromosomes. *Nature* 409, 478–478. doi: 10.1038/35054138
- She, Q., Phan, H., Garrett, R. A., Albers, S. V., Stedman, K. M., and Zillig, W. (1998). Genetic profile of pNOB8 from *Sulfolobus*: The first conjugative plasmid from an archaeon. *Extremophiles* 2, 417–425. doi: 10.1007/s007920050087
- She, Q., Shen, B., and Chen, L. (2004). Archaeal integrases and mechanisms of gene capture. *Biochem. Soc. Trans.* 32, 222–226. doi: 10.1042/BST0320222
- Stedman, K. M., She, Q., Phan, H., Holz, I., Singh, H., Prangishvili, D., et al. (2000). pING family of conjugative plasmids from the extremely thermophilic archaeon *Sulfolobus islandicus*: Insights into recombination and conjugation in Crenarchaeota. *J. Bacteriol.* 182, 7014–7020. doi: 10.1128/JB.182.24.7014-7020.2000
- Sun, M., Feng, X., Liu, Z., Han, W., Liang, Y. X., and She, Q. (2018). An Orc1/Cdc6 ortholog functions as a key regulator in the DNA damage response in Archaea. *Nucleic Acids Res.* 46, 6697–6711. doi: 10.1093/nar/gky487
- van Wolferen, M., Ma, X., and Albers, S. V. (2015). DNA processing proteins involved in the UV-induced stress response of sulfolobales. *J. Bacteriol.* 197, 2941–2951. doi: 10.1128/JB.00344-15
- van Wolferen, M., Shajahan, A., Heinrich, K., Brenzinger, S., Black, I. M., Wagner, A., et al. (2020). Species-specific recognition of sulfolobales mediated by uv-inducible pili and s-layer glycosylation patterns. *MBio* 11, e3014–e3019. doi: 10.1128/mBio.03014-19
- van Wolferen, M., Wagner, A., van der Does, C., and Albers, S.-V. (2016). The archaeal Ced system imports DNA. *Proc. Natl. Acad. Sci.* 113, 2496–2501. doi: 10.1073/pnas.1513740113
- Waters, V. L., Strack, B., Pansegrau, W., Lanka, E., and Guiney, D. G. (1992). Mutational analysis of essential IncPα plasmid transfer genes traF and traG and involvement of traF in phage sensitivity. *J. Bacteriol.* 174, 6666–6673. doi: 10.1128/jb.174.20.6666-6673.1992
- Zhang, C., and Whitaker, R. J. (2018). Microhomology-mediated high-throughput gene inactivation strategy for the hyperthermophilic crenarchaeon *sulfolobus islandicus*. *Appl. Env. Microbiol.* 84, 1–19. doi: 10.1128/AEM.02167-17
- Zhang, C., Cooper, T. E., Krause, D. J., and Whitaker, R. J. (2013a). Augmenting the genetic toolbox for *sulfolobus islandicus* with a stringent positive selectable marker for agmatine prototrophy. *Appl. Environ. Microbiol.* 79, 5539–5549. doi: 10.1128/AEM.01608-13
- Zhang, C., Krause, D. J., and Whitaker, R. J. (2013b). *Sulfolobus islandicus*: A model system for evolutionary genomics. *Biochem. Soc. Trans.* 41, 458–462.



OPEN ACCESS

EDITED BY

Marleen van Wolferen,
University of Freiburg,
Germany

REVIEWED BY

Ivan A. Berg,
University of Münster,
Germany
Jin-ho Lee,
Kyungsung University,
Republic of Korea

*CORRESPONDENCE

Haruyuki Atomi
✉ atomi@sbchem.kyoto-u.ac.jp

SPECIALTY SECTION

This article was submitted to
Biology of Archaea,
a section of the journal
Frontiers in Microbiology

RECEIVED 17 December 2022

ACCEPTED 31 January 2023

PUBLISHED 20 February 2023

CITATION

Su Y, Michimori Y and Atomi H (2023)
Biochemical and genetic examination of two
aminotransferases from the hyperthermophilic
archaeon *Thermococcus kodakarensis*.
Front. Microbiol. 14:1126218.
doi: 10.3389/fmicb.2023.1126218

COPYRIGHT

© 2023 Su, Michimori and Atomi. 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.

Biochemical and genetic examination of two aminotransferases from the hyperthermophilic archaeon *Thermococcus kodakarensis*

Yu Su¹, Yuta Michimori¹ and Haruyuki Atomi^{1,2*}

¹Department of Synthetic Chemistry and Biological Chemistry, Graduate School of Engineering, Kyoto University, Kyoto, Japan, ²Integrated Research Center for Carbon Negative Science, Kyoto University, Kyoto, Japan

The hyperthermophilic archaeon *Thermococcus kodakarensis* utilizes amino acids as a carbon and energy source. Multiple aminotransferases, along with glutamate dehydrogenase, are presumed to be involved in the catabolic conversion of amino acids. *T. kodakarensis* harbors seven Class I aminotransferase homologs on its genome. Here we examined the biochemical properties and physiological roles of two Class I aminotransferases. The TK0548 protein was produced in *Escherichia coli* and the TK2268 protein in *T. kodakarensis*. Purified TK0548 protein preferred Phe, Trp, Tyr, and His, and to a lower extent, Leu, Met and Glu. The TK2268 protein preferred Glu and Asp, with lower activities toward Cys, Leu, Ala, Met and Tyr. Both proteins recognized 2-oxoglutarate as the amino acceptor. The TK0548 protein exhibited the highest k_{cat}/K_m value toward Phe, followed by Trp, Tyr, and His. The TK2268 protein exhibited highest k_{cat}/K_m values for Glu and Asp. The TK0548 and TK2268 genes were individually disrupted, and both disruption strains displayed a retardation in growth on a minimal amino acid medium, suggesting their involvement in amino acid metabolism. Activities in the cell-free extracts of the disruption strains and the host strain were examined. The results suggested that the TK0548 protein contributes to the conversion of Trp, Tyr and His, and the TK2268 protein to that of Asp and His. Although other aminotransferases seem to contribute to the transamination of Phe, Trp, Tyr, Asp, and Glu, our results suggest that the TK0548 protein is responsible for the majority of aminotransferase activity toward His in *T. kodakarensis*. The genetic examination carried out in this study provides insight into the contributions of the two aminotransferases toward specific amino acids *in vivo*, an aspect which had not been thoroughly considered thus far.

KEYWORDS

Archaea, aminotransferase, metabolism, enzyme, genetics

Introduction

Aminotransferases, or transaminases, catalyze the reversible transfer of an amino group from a donor to a keto group of an acceptor compound (Toyokawa et al., 2021; Koper et al., 2022). The reaction is dependent on pyridoxal 5'-phosphate (PLP). Many aminotransferases recognize the α -amino group of α -amino acids and the keto group of 2-oxoacids, but some

recognize the amino group on side chains (Koszelewski et al., 2010; Malik et al., 2012; Zheng et al., 2018) or substrates other than α -amino acids (Gomm and O'Reilly, 2018; Kelly et al., 2020). The enzymes play a pivotal role in the carbon and nitrogen metabolism in a wide range of organisms. Aminotransferases are a focus of attention due to their involvement in a variety of diseases (Kunutsor et al., 2013; Montioli et al., 2021) and are also attractive enzymes for their use in biocatalysis (Malik et al., 2012; Gomm and O'Reilly, 2018; Kelly et al., 2020).

Aminotransferases can be divided into four classes (Class I to IV) based on their primary structure (Mehta et al., 1993). On the other hand, PLP-dependent enzymes can be classified by their folds (Fold type I to VII; Janssonius, 1998; Schneider et al., 2000; Koper et al., 2022). Aminotransferase Class I, II, and IV share a common ancestor with a Fold type I whereas Class III aminotransferases are members of Fold type IV PLP enzymes (Mehta et al., 1993). Class I enzymes constitute the majority of aminotransferases, including aromatic aminotransferases, aspartate aminotransferases, and alanine aminotransferases. Class II includes enzymes that can recognize amino groups other than the α -amino group of amino acids and utilize substrates such as γ -aminobutyric acid, or GABA, and ornithine. Class III enzymes with the distinct fold type IV utilize D-amino acids and branched-chain amino acids. Class IV enzymes utilize Ser, O-phosphoserine or Asp as the amino donor (Mehta et al., 1993; Mehta and Christen, 2000; Koper et al., 2022).

Thermococcales consists of three genera, *Pyrococcus*, *Thermococcus*, and *Palaeococcus*, and members readily utilize amino acids and peptides as a carbon and energy source for growth (Zillig et al., 1983; Fiala and Stetter, 1986; Takai et al., 2000). Different enzymes and pathways involved in amino acid catabolism and biosynthesis have been studied in these organisms. In particular, the enzymatic properties of a wide range of aminotransferases from *Pyrococcus furiosus* (Andreotti et al., 1995; Ward et al., 2000, 2002), *Pyrococcus horikoshii* (Matsui et al., 2000; Ura et al., 2001; Okada et al., 2012, 2014), *Thermococcus litoralis* (Andreotti et al., 1994; Sakuraba et al., 2004, 2008), *Thermococcus* sp. CKU-1 (Uchida et al., 2014) and *Thermococcus kodakarensis* (Kanai et al., 2015; Zheng et al., 2018) have been reported. Our group has been examining amino acid metabolism in *T. kodakarensis* (Shikata et al., 2007; Yokooji et al., 2013; Awano et al., 2014). In addition to the genome sequence (Fukui et al., 2005), a versatile genetic system (Sato et al., 2003, 2005; Matsumi et al., 2007; Santangelo et al., 2008, 2010) allows us to evaluate the physiological roles of individual genes in this organism.

A biochemical and genetic examination of enzymes/genes involved in Glu metabolism in *T. kodakarensis* confirmed the presence of glutamate dehydrogenase (TK1431), catalyzing an NADP-dependent interconversion between Glu and 2-oxoglutarate (Yokooji et al., 2013). The functions of 2-oxoglutarate:ferredoxin oxidoreductase (TK1123-TK1125, TK1131) and ADP-forming succinyl-CoA synthetase (TK1880, TK0943; Shikata et al., 2007) were also identified, constituting a route from Glu to succinate. The latter two reactions lead to the generation of reduced ferredoxin and ATP coupled to 2-oxoacid degradation, providing reducing equivalents and chemical energy (Yokooji et al., 2013). The presence of four 2-oxoacid:ferredoxin oxidoreductases and five ADP-forming acyl-CoA synthetases in *T. kodakarensis* suggest that other amino acids are catabolized through similar reactions. Based on the substrate specificities of the acyl-CoA synthetases (Mai and Adams, 1996; Glasemacher et al., 1997; Musfeldt et al., 1999; Shikata et al., 2007; Awano et al., 2014), it was suggested

that the following amino acids are subject to catabolism; Ala, Val, Leu, Ile, Met, Phe, Tyr, Trp, Glu/Gln, Cys, and His (Awano et al., 2014). However, dehydrogenases acting on amino acids other than Glu such as aspartate dehydrogenase and alanine dehydrogenase have been genetically confirmed to be absent in *T. kodakarensis* (Yokooji et al., 2013). This implies that 2-oxoacid generation from amino acids relies on aminotransferases. In order to better understand which amino acids are subject to catabolic degradation and which aminotransferases are involved, here we carried out a biochemical and genetic examination on two aminotransferases in *T. kodakarensis* encoded by TK0548 and TK2268.

Results

Multiple groups of aminotransferase homologs in Thermococcales

Figure 1 shows a phylogenetic tree of Class I aminotransferase homologs from four selected Thermococcales species, *T. kodakarensis*, *P. furiosus*, *P. horikoshii*, and *T. litoralis*, from which multiple aminotransferases have been experimentally studied. A clade including TK1990, a representative of a Class IV protein encoding cysteine desulfurase (Hidese et al., 2014), was used as the outgroup. There are eight groups (G1–G8) in which homologs occur in ten or more genomes of the 40 Thermococcales species examined. Among the eight groups, members of G1 (39 genomes among the 40 genomes examined harbor homologs; 39/40), G2 (40/40), G3 (40/40), G6 (40/40), and G8 (35/40) are present in most or all of the Thermococcales genomes used in our analysis. Members of G4 (12/40), G5 (10/40), and G7 (20/40) are less distributed. *T. litoralis* harbors a higher number of Class I aminotransferase homologs compared to the other three species. Homologs of OCC_03517 are not found on any of the other Thermococcales genomes, but homologs of OCC_10965 (3/40), OCC_02240 (10/40, G5), and OCC_11814 (5/40) can be identified in other species. A detailed phylogenetic tree based on sequences of all Class I aminotransferase homologs from Thermococcales species is shown in Supplementary Figure S1.

We considered the physiological roles of these groups of aminotransferases by first taking into account their gene location and previous biochemical studies on aminotransferases from members of Thermococcales. In terms of gene location, we observed that TK0250 (G7) is included within the His biosynthesis gene cluster (TK0242-TK0251) in *T. kodakarensis*. We also observed a complete co-occurrence between the TK0250 homologs and the His biosynthesis operon in Thermococcales members (Supplementary Table S1). A similar situation was observed for TK0260 (G4). The gene is situated in a gene cluster (TK0259-TK0261) involved in the biosynthesis of Phe/Tyr in *T. kodakarensis*, and co-occurrence is observed between TK0260 homologs and the gene cluster in Thermococcales species (Supplementary Table S2). These observations suggest that the physiological roles of members of G4 and G7 are related to Phe/Tyr and His biosynthesis, respectively. The members of G8 from Thermococcales have not been characterized, but display 30% identity to Thr decarboxylase (encoded by MM2060) from *Methanosarcina mazei* (Tavares et al., 2018, 2019) and are clustered with genes related to cobalamin salvage. By contrast, members of G1, G2, G3, G5, and G6 did not show a tendency to be included in a particular biosynthesis operon or gene cluster.

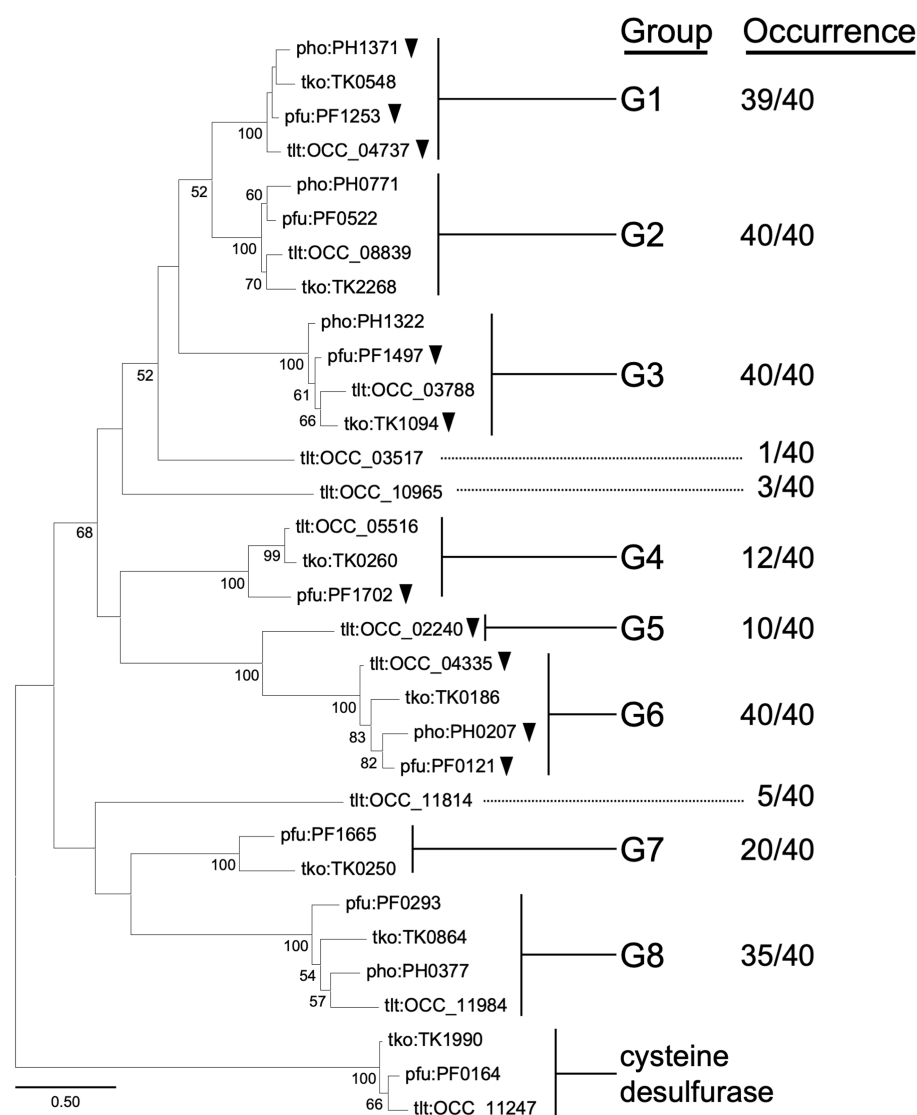


FIGURE 1

Phylogenetic analysis of Class I aminotransferases from selected species of Thermococcales. Amino acid sequences were collected from *T. kodakarensis*, *T. litoralis*, *P. furiosus*, and *P. horikoshii*. Homologs of cysteine desulfurase (TK1990) were added into the dataset as an outgroup. The sequences were aligned using MUSCLE algorithm (Edgar, 2004). The phylogenetic analysis was performed using the Maximum Likelihood method and JTT matrix-based model (Jones et al., 1992). The tree with the highest log likelihood (−9712.68) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Bootstrap values above 50 are shown. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the JTT model, then selecting the topology with superior log likelihood value. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. This analysis involved 32 amino acid sequences. All positions containing gaps and missing data were eliminated (complete deletion option). There were a total of 242 positions in the final dataset. Evolutionary analyses were conducted in MEGA11 (Tamura et al., 2021). Homologs that have been characterized genetically or biochemically are indicated with black arrowheads. Information on the groups or occurrence is indicated in the text.

Members of G1, G2, G3, and G6 are widely distributed among Thermococcales species, and a number of members have been biochemically characterized. In particular, members of G1 and G6 have been relatively well studied. In the case of G1, the structure of the PH1371 protein from *P. horikoshii* has been elucidated, and the amino donors most recognized were Tyr, Phe, Glu, His, and Trp (Matsui et al., 2000). The PF1253 protein from *P. furiosus*, referred to as AroAT II, displayed preference to Phe, Tyr and Trp, with highest k_{cat}/K_m toward Phe (Ward et al., 2002). The OCC_04737 protein from *T. litoralis*, referred to as ArAT II, also displayed similar properties, preferring Phe, Tyr, and Trp (Andreotti et al., 1994). In the case of G6,

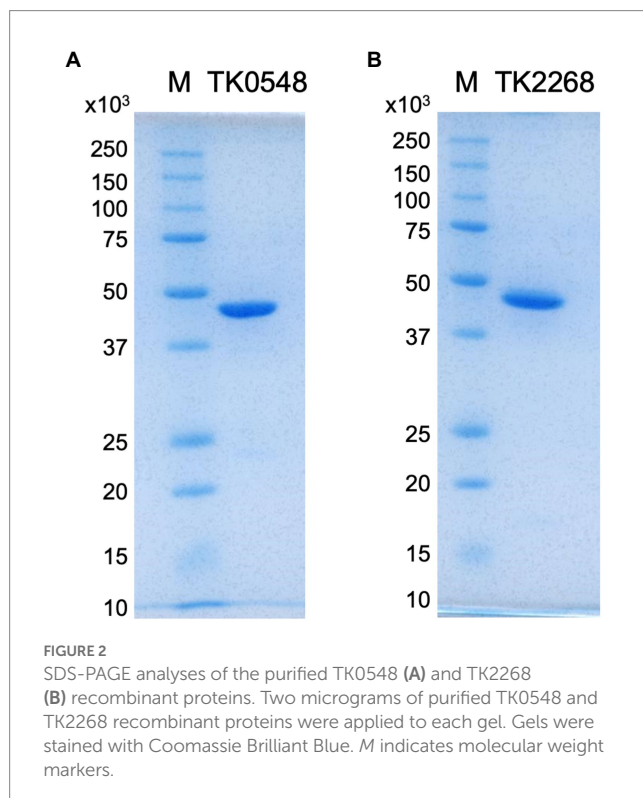
the PF0121 protein from *P. furiosus*, referred to as ArAT or AroAT I, displayed significant activity toward Phe, Trp, and Tyr (Andreotti et al., 1995). The OCC_04335 protein from *T. litoralis*, referred to as ArAT I, also recognized Phe, Tyr, and Trp (Andreotti et al., 1994). The PH0207 protein displays kynurenine aminotransferase activity which is involved in the degradation of Trp (Okada et al., 2012), and the crystal structure has been determined (Chon et al., 2005; Okada et al., 2014). Concerning other groups, the PF1497 protein from *P. furiosus* (G3), referred to as AlaAT, displays highest activity with Ala. The enzyme did not display activity toward Phe and Tyr (Ward et al., 2000). A genetic analysis on TK1094 from *T. kodakarensis* (G3)

suggested a role in the conversion between pyruvate and Ala (Kanai et al., 2015). The PF1702 protein (G4), referred to as AspAT, displays highest activity toward Asp and Glu, with low activities to a broad range of amino acids (Ward et al., 2002).

We carried out further analyses on TK2268 and TK0548. The TK2268 protein is a member of G2, which does not have any member that has been experimentally characterized. Determining its biochemical properties would contribute to our understanding of amino acid metabolism in *T. kodakarensis*. We also examined the TK0548 protein, the most closely related protein to the TK2268 protein (45.4% identical). None of the G1 and G2 genes have been examined genetically.

Expression of the TK0548 and TK2268 genes and purification of the recombinant proteins

In order to obtain recombinant proteins, the TK0548 and TK2268 genes were expressed in *Escherichia coli*. In the case of TK0548, soluble protein was obtained, and the recombinant TK0548 protein was purified through heat treatment, anion exchange chromatography and gel-filtration chromatography. In the case of TK2268, expression in *E. coli* resulted in the formation of inclusion bodies. The gene was thus expressed in the native host *T. kodakarensis* under the control of a strong constitutive promoter of the cell surface glycoprotein gene (*csg*, TK0895; Yokooji et al., 2009). Sequences to incorporate a His₆-tag on the C-terminus of the TK2268 protein were introduced. The soluble TK2268 protein obtained in the cell extract of *T. kodakarensis* cells was purified using a nickel affinity column and gel filtration chromatography. Both proteins were subjected to SDS-PAGE, confirming their apparent homogeneities (Figure 2).



Oligomeric state of the TK0548 and TK2268 proteins

The molecular masses of the purified, recombinant TK0548 and TK2268 proteins were examined with gel-filtration chromatography. The estimated molecular mass of the TK0548 protein was 85 kDa, and considering that the calculated molecular mass of the monomer was 43,608 Da, this suggested that the TK0548 protein was a dimer. In the case of the TK2268 protein, we consistently observed two peaks, corresponding to estimated molecular masses of 98 kDa and 327 kDa. As the calculated molecular mass of the monomer was 45,116 Da, the result suggested that the TK2268 protein forms a dimer unit, which may then further assemble to form an octamer.

The TK0548 and TK2268 proteins catalyze transamination in a PLP-dependent manner

The purified TK0548 and TK2268 proteins were examined for aminotransferase activity. We used Glu (10 mM) as the amino group donor and pyruvate (10 mM) as the amino acceptor. We observed aminotransferase activity in both proteins. When PLP was omitted from the reaction, we observed a partial decrease in activity in both cases (TK0548: 13% decrease, TK2268: 50% decrease) compared to those observed with the addition of PLP, suggesting that the TK0548 and TK2268 proteins are PLP-dependent aminotransferases. The only partial decrease in activity is most likely due to PLP bound to the proteins when they were produced in their respective host cells, *E. coli* (TK0548) and *T. kodakarensis* (TK2268). To further support the necessity of PLP for the reactions, we added hydroxylamine, a PLP inhibitor (Kito et al., 1978) to the enzymes in the absence of supplemental PLP. In this case, we observed a further 93% decrease in activity of the TK0548 protein and a 78% decrease in activity of the TK2268 protein with the addition of hydroxylamine (Supplementary Figure S2).

Transamination with varying amino donor and acceptor compounds

We first carried out an initial screening to identify the amino acids recognized by the TK0548 and TK2268 proteins. 2-Oxoglutarate or pyruvate was used as the amino acceptor, and the production of Glu or Ala, respectively, was examined after 15 min. The substrates were present in the reaction mixture at a concentration of 10 mM. As shown in Figure 3A, the TK0548 protein utilized Phe, Tyr, His, and Trp as an amino donor, as well as Met, Glu, and Leu to a lower extent. In the case of the TK2268 protein, Asp and Glu were utilized, along with Tyr, Leu, Ala, Met, and Cys (Figure 3B).

Focusing on the amino acids that were recognized by the proteins, we examined enzyme activity. The amino acids and amino acceptors were constant at 10 mM. As shown in Figure 4A, the TK0548 protein exhibited highest activity toward Tyr, followed by Phe, Trp, and His. Activity toward Leu, Met, and Glu were much lower. On the other hand, the TK2268 protein displayed highest activity toward Glu, followed by Asp. Lower levels of activity were observed using Cys, Leu, Ala, Met, and Tyr (Figure 4B).

Substrate specificity of the TK0548 and TK2268 proteins

Kinetic analyses were performed on the substrates that resulted in relatively high levels of activity. For the TK0548 protein, activities were first measured for varying concentrations of pyruvate or 2-oxoglutarate in the presence of 10 mM Phe. As shown in Table 1, the TK0548 protein exhibited a higher V_{\max} and lower K_m toward 2-oxoglutarate when compared to pyruvate. The k_{cat}/K_m value toward 2-oxoglutarate was over 70-fold higher than that for pyruvate. We next examined activity with varying concentrations of Phe, Tyr, Trp, His, and Met in the presence of 10 mM 2-oxoglutarate (Table 1). The highest k_{cat}/K_m

value was observed with Phe, with high values also observed for Tyr and Trp. A lower k_{cat}/K_m value was observed with His, and that for Met was extremely low. The results suggested that the TK0548 protein mainly utilizes the aromatic amino acids Phe, Tyr, and Trp as substrates, with His also a possible substrate.

Concerning the TK2268 protein, Leu (50 mM) was used as the amino donor to measure activity with varying concentrations of pyruvate and 2-oxoglutarate. The k_{cat}/K_m value toward 2-oxoglutarate was higher than that toward pyruvate (Table 2). This was mainly due to differences in the K_m value, as their V_{\max} values were comparable. As the product of the aminotransferase reaction with 2-oxoglutarate

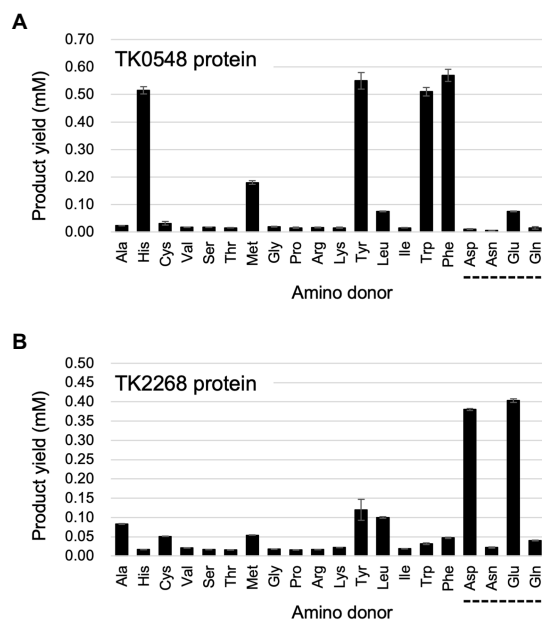


FIGURE 3

Examination of the amino donors recognized by the recombinant TK0548 (A) and TK2268 (B) proteins. Each protein was incubated for 15 min at 80°C with the indicated amino acids and 2-oxoglutarate and the generation of Glu was examined. The amino acids indicated with a dotted line below them were examined using pyruvate as the amino acceptor, and the generation of Ala was examined. Amino donors and acceptors were added at a concentration of 10 mM, with the exception of Tyr (6 mM). The results are the means of three independent assays and error bars indicate standard deviations.

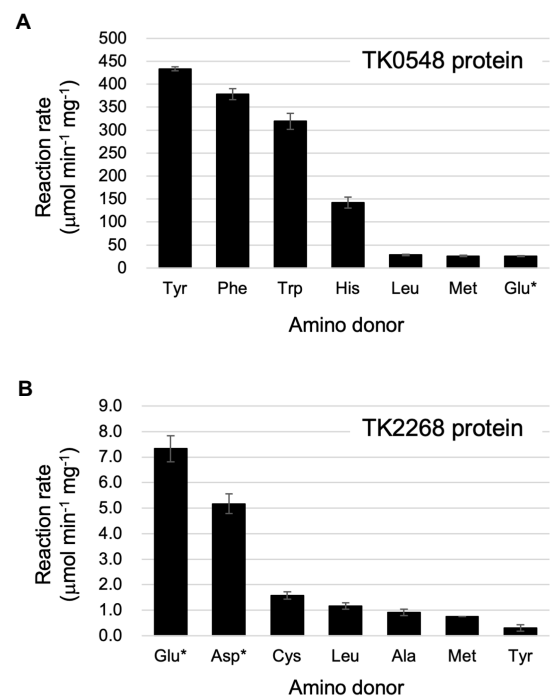


FIGURE 4

Reaction rates of the recombinant TK0548 (A) and TK2268 (B) proteins with various amino donors. Each protein was incubated for varying periods of time at 80°C with the indicated amino acids and 2-oxoglutarate or pyruvate (indicated with asterisks), and the reaction rates of Glu or Ala production, respectively, were calculated. Amino donors and acceptors were added at a concentration of 10 mM (Tyr: 6 mM). The results are the means of three independent assays and error bars indicate standard deviations.

TABLE 1 Kinetic parameters of the aminotransferase reaction catalyzed by the TK0548 protein.

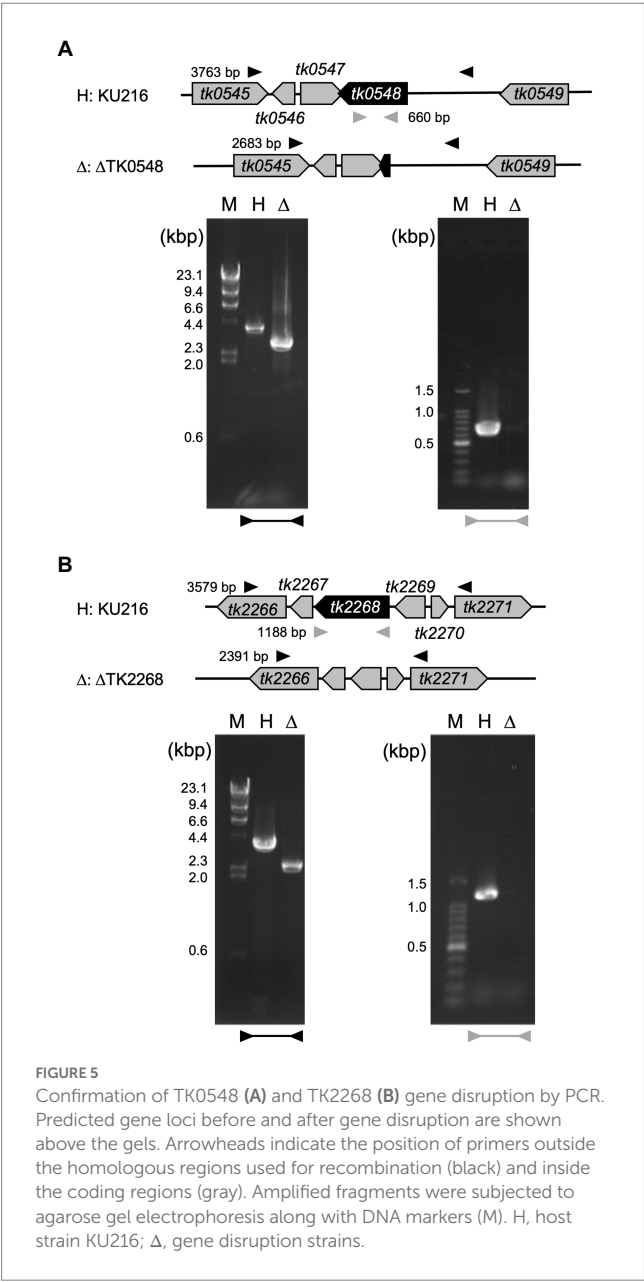
Substrate		V_{\max} ($\mu\text{mol min}^{-1} \text{mg}^{-1}$)	K_m (mM)	k_{cat} (s^{-1})	k_{cat}/K_m ($\text{mM}^{-1} \text{s}^{-1}$)
Amino donor	Amino acceptor				
<i>Phenylalanine</i>	2-Oxoglutarate ^a	483 ± 10	0.67 ± 0.06	351 ± 10	524
<i>Tyrosine</i>		419 ± 18	1.4 ± 0.2	305 ± 19	218
<i>Tryptophan</i>		359 ± 10	1.1 ± 0.1	261 ± 10	237
<i>Histidine</i>		179 ± 6	4.4 ± 0.4	130 ± 6	29.6
<i>Methionine</i>		169 ± 14	64 ± 9	123 ± 14	1.92
Phenylalanine ^a	2-Oxoglutarate	334 ± 10	1.5 ± 0.2	243 ± 10	162
	Pyruvate	137 ± 7	48 ± 5	100 ± 7	2.07

The kinetic parameters are those for the substrates indicated in italic. ^aConcentration fixed at 10 mM.

TABLE 2 Kinetic parameters of the aminotransferase reaction catalyzed by the TK2268 protein.

Substrate		V_{\max} ($\mu\text{mol min}^{-1} \text{mg}^{-1}$)	K_m (mM)	k_{cat} (s^{-1})	k_{cat}/K_m ($\text{mM}^{-1} \text{s}^{-1}$)
Amino donor	Amino acceptor				
<i>Aspartate</i>	Pyruvate ^a	5.42 ± 0.18	2.14 ± 0.22	4.07 ± 0.14	1.90
<i>Glutamate</i>		7.42 ± 0.31	2.17 ± 0.27	5.57 ± 0.23	2.57
<i>Leucine</i>		4.68 ± 0.22	36.3 ± 4.3	3.52 ± 0.23	0.10
<i>Leucine</i>	2-Oxoglutarate ^a	5.48 ± 0.22	39.6 ± 3.8	4.12 ± 0.23	0.10
<i>Tyrosine</i>		0.34 ± 0.02	1.45 ± 0.26	0.26 ± 0.02	0.18
<i>Leucine</i> ^b	2-Oxoglutarate	4.45 ± 0.14	0.19 ± 0.03	3.35 ± 0.15	17.6
	Pyruvate	3.76 ± 0.15	2.72 ± 0.35	2.83 ± 0.16	1.04

The kinetic parameters are those for the substrates indicated in *italic*. ^aConcentration fixed at 10 mM. ^bConcentration fixed at 50 mM.



would not be possible with Glu as the amino donor, 10 mM pyruvate was used as the amino acceptor to examine the activities with varying concentrations of Glu, Asp, and Leu. As a result, relatively high k_{cat}/K_m

values were observed for both Asp and Glu compared to Leu and Tyr. The results suggest that the TK2268 protein is an aminotransferase with specificity toward the acidic amino acids Glu and Asp.

Gene disruption and growth of the ΔTK0548 and ΔTK2268 strains

To understand the contribution of the TK0548 and TK2268 genes to amino acid catabolism in *T. kodakarensis*, we disrupted each gene using *T. kodakarensis* KU216 (ΔpyrF) as a host strain. Five transformants were chosen for each gene disruption and their loci were examined by PCR. Examples of transformants whose TK0548 gene (Figure 5A) or TK2268 gene (Figure 5B) was disrupted are shown. The respective loci were sequenced, confirming that gene disruption had occurred as intended.

The ΔTK0548 and ΔTK2268 disruption strains were first grown in the nutrient-rich ASW-YT-m1-S⁰ medium containing yeast extract and tryptone (Figure 6A). The growth of both disruption strains did not display significant differences to that of the host strain KU216. Therefore, in order to increase the dependency of growth on amino acid catabolism, the strains were grown in ASW-AA-m1-S⁰(+Ura) medium (Figure 6B). This medium is a synthetic medium with amino acids as the only major carbon and energy source. In this case, we observed a retardation in growth in both gene disruption strains. Cell yields of the cultures eventually reached similar levels, but the gene disruption strains took a 4-h longer period of time to reach maximum cell density. The results suggest that both proteins are involved in the utilization of amino acids for growth of *T. kodakarensis*.

Aminotransferase activity in *Thermococcus kodakarensis* cell extracts

The substrate specificities of multiple aminotransferases from members of Thermococcales have been determined *in vitro*. However, the contribution of each protein in the conversion of a particular amino acid *in vivo* cannot be determined by biochemical properties alone, as multiple aminotransferases, in some cases with overlapping substrate specificities or different expression levels, are present in the cell. We thus measured and compared aminotransferase activity in the cell extracts of *T. kodakarensis* host strain and gene disruption strains. As *in vitro* studies indicated that the TK0548 protein preferred Trp, Phe, Tyr, and His, while the TK2268 protein recognized Asp and Glu, activities in the cell extracts toward Trp, Phe, Tyr, His, Asp, and Glu were measured.

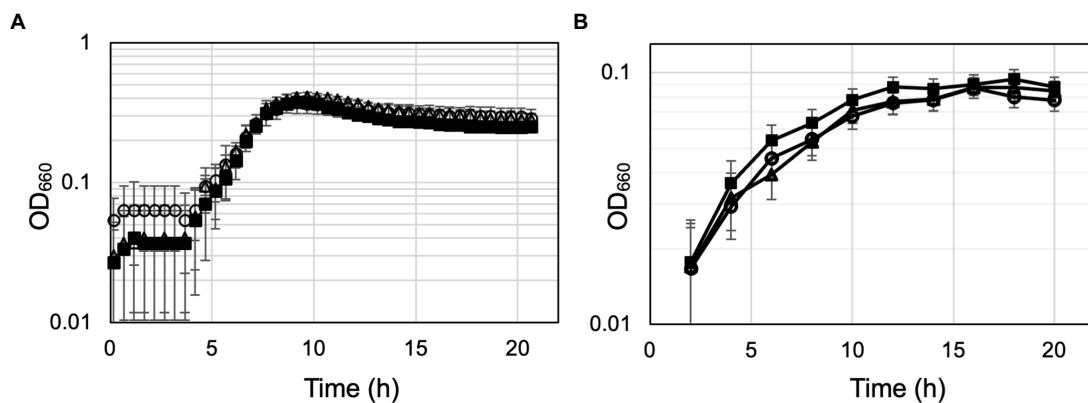


FIGURE 6

Growth properties of the host strain KU216 and the gene disruption strains. Growth of the host KU216 strain (closed squares) and the Δ TK0548 (open triangles) and Δ TK2248 (open circles) gene disruption strains were examined in a nutrient-rich ASW-YT-m1-S⁰ medium containing yeast extract and tryptone (A), and in a synthetic amino acid medium ASW-AA-m1-S⁰(+Ura) (B). Growth was measured at 85°C. Error bars indicate the standard deviations of three independent culture experiments. The vertical axis is represented in logarithmic scale.

We also used Ile as an amino donor. *In vitro* studies indicated that the recognition of both TK0548 and TK2268 proteins toward Ile was minimal (Figure 3), so the effects of TK0548 and TK2268 disruption on intracellular aminotransferase activity toward Ile would be expected to be low. As shown in Figure 7, aminotransferase activity for all seven amino acids was clearly observed in KU216 cell extracts. We observed that the levels of aminotransferase activity toward different amino acids greatly differed. Activities toward Phe or Glu were particularly high, whereas that toward Asp was notably low, two orders of magnitude lower than those observed for Phe or Glu. When we examined the effects of gene disruption, the disruption of TK0548 and TK2268 had no effect on intracellular aminotransferase activity toward Ile, consistent with our *in vitro* results that neither enzyme recognized Ile (Figure 3). We next examined differences observed upon gene disruption, focusing on those with *p* values below 0.01. We observed that the disruption of TK2268 resulted in a 35% decrease in aminotransferase activity toward Asp, one of the substrates preferred by the protein in *in vitro* experiments. Disruption of TK2268 did not affect the Glu aminotransferase activity in cell extracts, suggesting that the contribution of the protein to this activity is relatively small. This is not surprising though, as *in vitro* analysis indicated that the activity of the TK2268 protein toward Glu and Asp are comparable (Table 2). As the total Glu aminotransferase activity in *T. kodakarensis* cell extracts is over 100-fold higher than that toward Asp, a decrease in Glu aminotransferase activity comparable to the levels toward Asp would correspond to only a very small fraction of the total activity toward Glu. We also observed a decrease in His aminotransferase activity, but this could not be explained by the results of *in vitro* analysis. When TK0548 was disrupted, decreases in activity toward Trp (32%), Tyr (64%), and His (89%) were observed. The TK0548 protein seems to be a major contributor for Tyr and His aminotransferase activity in *T. kodakarensis*.

Discussion

Biochemical studies have been carried out on wealth of aminotransferases from Thermococcales species. Proteins from *T. kodakarensis*, *P. furiosus*, *P. horikoshii*, and *T. litoralis* that are

classified in the Class I to Class IV aminotransferases are listed in Table 3. Some proteins have been shown to catalyze reactions other than transamination, and include racemases, decarboxylases, desulfurases, hydroxymethyltransferases and others. Concerning the aminotransferases that have been biochemically examined, the amino donor and acceptor molecules are indicated. The TK0548 protein was shown here to prefer the aromatic amino acids Phe, Tyr and Trp, which is consistent with other members of G1. To a lower degree, the protein also recognized His. The PH1371 protein also displays activity toward His (Matsui et al., 2000). The activities of the PF1253 (Ward et al., 2002) and OCC_04737 (Andreotti et al., 1994) proteins toward His have not been examined. In the case of the TK2268 protein, it is the first characterized representative of G2, and displays specificity toward the acidic amino acids Asp and Glu. Although the range of substrates examined in separate studies differ, enzymes that belong to the same group have been shown to display similar substrate specificities, such as the four enzymes from G1 and the three enzymes from G6. Our results with the TK2268 protein thus raise the possibilities that members of G2 recognize acidic amino acids as amino donors.

In order to understand the contribution of each enzyme to the transamination of a specific amino acid *in vivo*, we examined and compared specific aminotransferase activities among cells of the host strain KU216 and Δ TK0548 and Δ TK2268 disruption strains. This takes into account not only the substrate specificity and activity levels of each enzyme, but also their expression levels in the cell. The results suggested that TK0548 contributed to the aminotransferase activity toward Phe and Trp, and to a higher degree Tyr. The considerable levels of activity still observed in the disruption strains most likely reflects the activity of the TK0186 protein, a member of G6 (Table 3). Interestingly, the TK0548 protein seems to be the predominant His aminotransferase in *T. kodakarensis*, accounting for approximately 90% of the activity in cell extracts.

Amino acid catabolism in members of the Thermococcales proceeds via amino acid, 2-oxoacid, acyl-CoA and acid (Figure 8). Concerning the conversion from amino acids to 2-oxoacids, our group has previously shown that Glu is the only amino acid that is converted by a dehydrogenase, a glutamate dehydrogenase encoded by TK1431. All

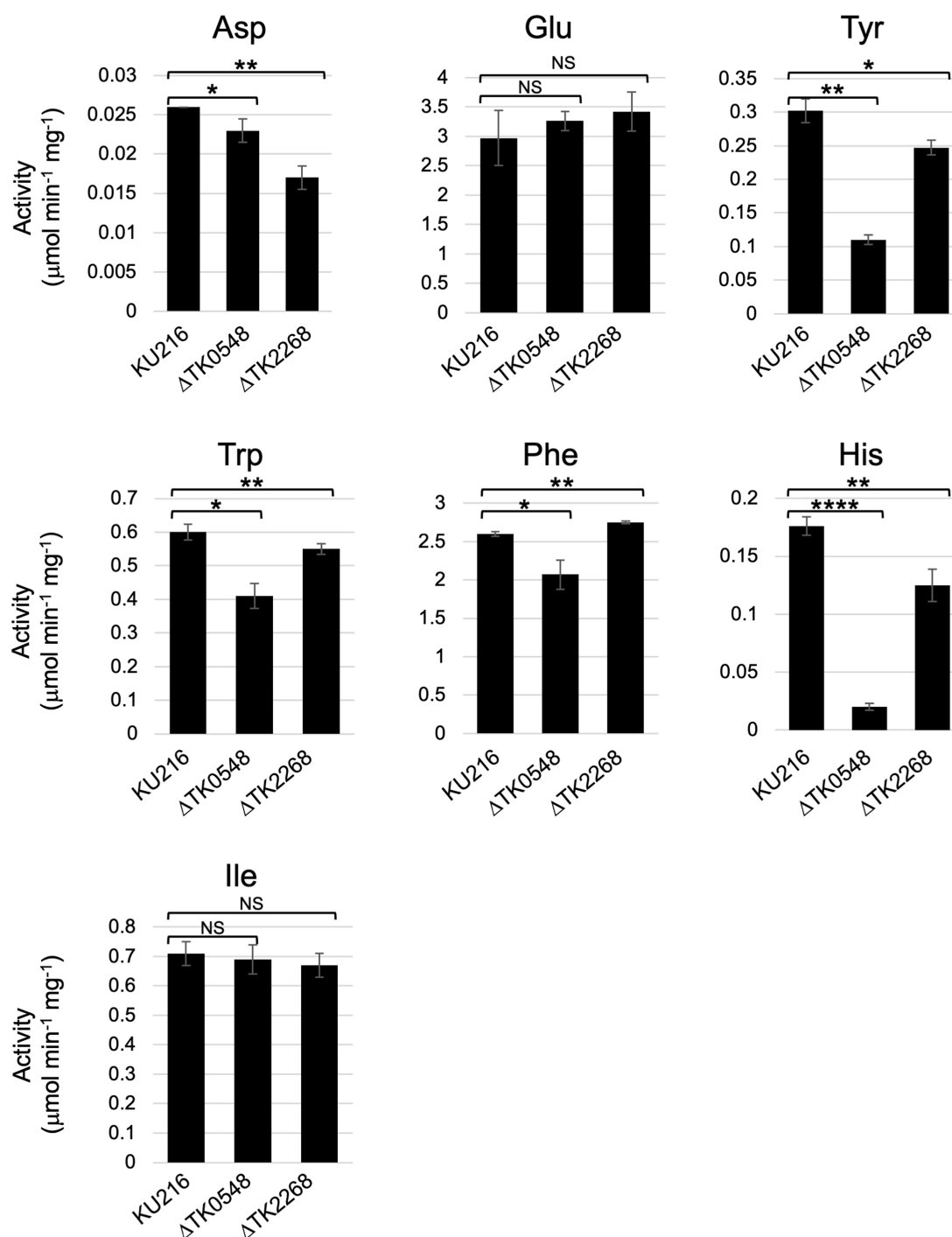


FIGURE 7

Aminotransferase activity in *T. kodakarensis* cell extracts. Cell extracts were incubated with different amino acids and 2-oxoglutarate, or pyruvate (for Glu and Asp), at 80°C for various periods of time. Activity was calculated based on the generation of Glu or Ala. Amino donors and acceptors were added at a concentration of 10mM (Tyr: 6mM). 0.01<*p*<0.05*; 0.001<*p*<0.01**; 0.0001<*p*<0.001***; *p*<0.0001****; not significant: NS.

other amino acids that are converted to 2-oxoacids would have to rely on aminotransferases (Yokooji et al., 2013). The 2-oxoacids are converted to acyl-CoAs via 2-oxoacid:ferredoxin oxidoreductases. There are seven potential sets of genes encoding 2-oxoacid:ferredoxin oxidoreductases in members of the Thermococcales, and four protein complexes from *P. furiosus* have been biochemically examined; pyruvate:ferredoxin oxidoreductase (POR), 2-ketoisovalerate:ferredoxin oxidoreductase

(VOR), indolepyruvate:ferredoxin oxidoreductase (IOR), and 2-ketoglutarate:ferredoxin oxidoreductase (KGOR) (Blamey and Adams, 1993; Mai and Adams, 1994; Heider et al., 1996; Mai and Adams, 1996). The KGOR homolog in *T. kodakarensis* has been genetically examined, confirming its role in Glu metabolism, converting 2-oxoglutarate to succinyl-CoA (Yokooji et al., 2013). Finally, the acyl-CoAs are hydrolyzed by NDP-forming acyl-CoA synthetases. There are five NDP-forming

TABLE 3 An overview of aminotransferase homologs in Thermococcales.

Species	Class	Gene number	Group	Function (Annotation)*	Amino donor**	Amino acceptor**	Reference
<i>Pyrococcus furiosus</i>	I	PF1253	G1	Aromatic aminotransferase	Phe, Tyr, Trp	2-OG	Ward et al. (2002)
		PF0522	G2	(Aspartate aminotransferase)			
		PF1497	G3	Alanine aminotransferase	Ala, Glu, (Asp, Ile, Leu)	2-OG, Pyr	Ward et al. (2000)
		PF1702	G4	Aspartate aminotransferase	Asp, (Glu)	2-OG, (Pyr, phenylpyruvate)	Ward et al. (2002)
		PF0121	G6	Aromatic aminotransferase	Phe, Tyr, Trp	2-OG	Andreotti et al. (1995)
		PF1665	G7	(Histidinol-phosphate aminotransferase)			
		PF0293	G8	(Histidinol-phosphate aminotransferase)			
	II	PF1685		(Acetylornithine aminotransferase)			
		PF1421		(4-aminobutyrate aminotransferase)			
		PF1232		(4-Aminobutyrate aminotransferase)			
		PF0513		(Putative glutamate aminotransferase)			
	III	ND		ND			
	IV	PF0164		(Cysteine desulfurase)			
		PF1066		(Putative aminotransferase)			
		PF1472		(Aspartate/serine transaminase)			
		PF1778		(Serine hydroxymethyltransferase)			
		PF1999		(Glycine dehydrogenase subunit 1)			
		PF2000		(Glycine dehydrogenase subunit 2)			
<i>Pyrococcus horikoshii</i>	I	PH1371	G1	Aromatic aminotransferase	Phe, Tyr, Trp, (His)	Phenylpyruvate, 2-OG	Matsui et al. (2000)
		PH0771	G2	(Aspartate aminotransferase)			
		PH1322	G3	(Alanine-synthesizing aminotransferase)			
		PH0207	G6	Kynurenine aminotransferase	Kynurenine	2-OG, OAA, 2-oxobutyrate, 2-oxo-4-methylthiobutyrate	Okada et al. (2012)
		PH0377	G8	(Histidinol-phosphate aminotransferase)			
	II	PH1716		(Long hypothetical acetylornithine aminotransferase)			
		PH1501		(Long hypothetical aminotransferase)			
		PH1423		Ornithine δ -aminotransferase	L-Orn, L-Lys, D-Orn, D-Lys, (5-aminovalerate, 6-aminohexanoate, GABA)	2-OG, (Pyr)	Kawakami et al. (2022)
		PH0782		Alanine/serine racemase	Ala, Ser***		Kawakami et al. (2018)
		PH0138		Amino acid racemase	Phe, Leu, Met, (Tyr, Ile, Val, Trp, Ala)***		Kawakami et al. (2015, 2017)
	III	ND		ND			
	IV	PH0626		(Long hypothetical protein)			
		PH1308		(Long hypothetical serine aminotransferase)			
		PH1654		(Long hypothetical serine hydroxymethyltransferase)			
		PH1994		(Glycine dehydrogenase subunit 2)			
		PH1995		(Glycine dehydrogenase subunit 1)			
<i>Thermococcus litoralis</i>	I	OCC_04737	G1	Aromatic aminotransferase	Phe, Tyr, Trp	2-OG	Andreotti et al. (1994)
		OCC_08839	G2	(Aminotransferase)			
		OCC_03788	G3	(Alanine aminotransferase)			
		OCC_03517		(Hypothetical protein)			
		OCC_10965		Aspartate aminotransferase			
		OCC_05516	G4	(Aspartate aminotransferase)			

(Continued)

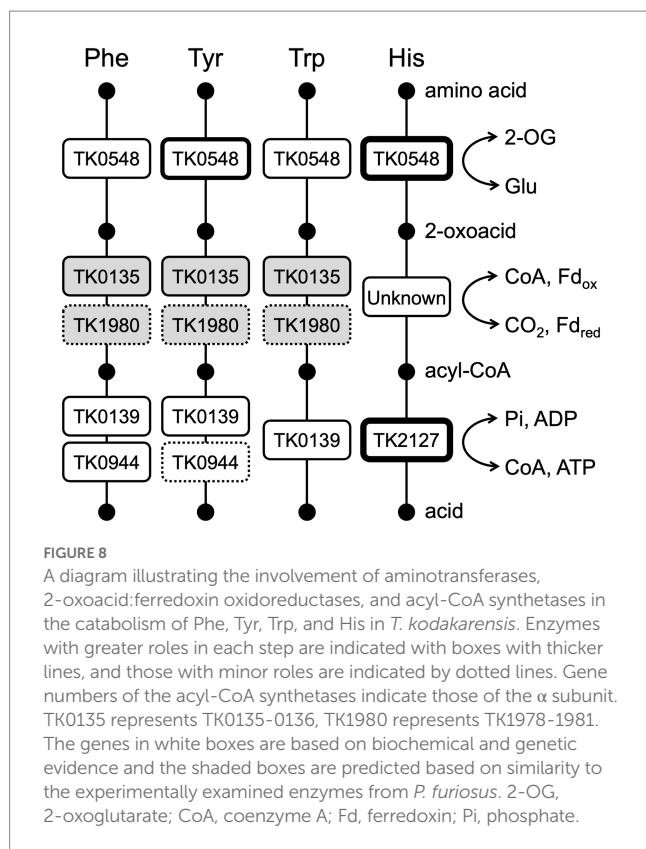
TABLE 3 (Continued)

Species	Class	Gene number	Group	Function (Annotation)*	Amino donor**	Amino acceptor**	Reference
		OCC_02240	G5	Alanine glyoxylate aminotransferase	Ala	Glyoxylate	Sakuraba et al. (2004)
		OCC_04335	G6	Aromatic aminotransferase	Phe, Tyr, Trp	2-OG	Andreotti et al. (1994)
		OCC_11814		(Histidinol-phosphate aminotransferase)			
		OCC_11984		(Histidinol-phosphate aminotransferase)			
	II	OCC_00582		(4-Aminobutyrate aminotransferase)			
		OCC_08410		(4-Aminobutyrate aminotransferase)			
		OCC_10945		Moderate-substrate specificity amino acid racemase	Met, Leu, (Phe, Ala, Ser)***		Kawakami et al. (2021)
	III	ND		ND			
	IV	OCC_11247		(Cysteine desulfurase)			
		OCC_00972		(Aminotransferase)			
		OCC_11879		(Aspartate aminotransferase)			
		OCC_07124		(Glycine dehydrogenase subunit 1)			
		OCC_07119		(Glycine dehydrogenase subunit 2)			
<i>Thermococcus kodakarensis</i>	I	TK0548	G1	Aromatic aminotransferase	Tyr, Phe, Trp, His	2-OG, (Pyr)	This study
		TK2268	G2	Aspartate aminotransferase	Glu, Asp	2-OG, (Pyr)	This study
		TK1094	G3	(Alanine aminotransferase)			Kanai et al. (2015)
		TK0260	G4	Aspartate aminotransferase			
		TK0186	G6	(2-Aminoacidipate transaminase)			
		TK0250	G7	(Histidinol-phosphate aminotransferase)			
		TK0864	G8	(Threonine- O-3-phosphate decarboxylase)			
	II	TK0275		LysW-γ-L-lysine aminotransferase			Yoshida et al. (2016)
		TK1211		Leu/Met racemase	Leu, Met***		Zheng et al. (2021)
		TK2101		Ornithine ω-aminotransferase	L-Orn, L-Lys, (D-Orn, D-Lys)	2-OG, 2-oxoadipate, (OAA, Pyr)	Zheng et al. (2018)
	III	ND		ND			
	IV	TK1990		Cysteine desulfurase	Cys***		Hidese et al. (2014)
		TK1303		(Hypothetical protein, conserved)			
		TK1548		(Probable serine-glyoxylate aminotransferase)			
		TK1379		(Glycine cleavage system protein P, subunit 2)			
		TK1380		(Glycine cleavage system protein P, subunit 1)			

*Functions in parentheses are based on annotation. **Compounds in parentheses are less recognized substrates. ***Substrates for the indicated enzymes other than aminotransferases. 2-OG, 2-oxoglutarate; OAA, oxaloacetate; Orn, ornithine; Pyr, pyruvate; GABA, γ-aminobutyric acid; ND, not detected.

acyl-CoA synthetases in Thermococcales; ACS I, ACS II, ACS III, succinyl-CoA synthetase (SCS), and 2-(imidazol-4-yl)acetyl-CoA synthetase (ICS). ACS I and ACS II from *P. furiosus* and ACS II, ACS III, SCS and ICS from *T. kodakarensis* have been biochemically examined ([Mai and Adams, 1996](#); [Glasemacher et al., 1997](#); [Musfeldt et al., 1999](#); [Shikata et al., 2007](#); [Awano et al., 2014](#)). The involvement of SCS in Glu metabolism has been genetically confirmed ([Yokooji et al., 2013](#)). As the NDP-forming acyl-CoA synthetases catalyze the reaction in which substrate-level phosphorylation occurs, it is possible to predict the amino acids that are subject to this catabolism based on the substrate specificities of the acyl-CoA synthetases. The substrates recognized by the five acyl-CoA synthetases combined suggest that the amino acids that are subject to this mode of catabolism are Ala, Val, Ile, Leu, Met, Phe, Tyr,

Trp, Glu/Gln, Cys, and His ([Awano et al., 2014](#)). Aminotransferases that recognize Ala, Val, Ile, Leu, Met, Phe, Tyr, Trp, and Glu/Gln have been reported, as well as 2-oxoacid:ferredoxin oxidoreductases that act on the corresponding 2-oxoacids after transamination. However, aminotransferases and 2-oxoacid:ferredoxin oxidoreductases that are related to Cys and His degradation are still not known. In particular, ICS displays a remarkably high specificity toward 2-(imidazol-4-yl)-acetate, suggesting that it is involved solely in His metabolism ([Awano et al., 2014](#)). The results of this study strongly suggest that the TK0548 protein and ICS are metabolically linked in His catabolism ([Figure 8](#)). To a certain extent, the TK0548 protein also contributes in the catabolism of aromatic amino acids, particularly Tyr. Finally, the specificities of the five acyl-CoA synthetases combined raise the possibilities that Asp, Asn, Gly, Lys, Arg,



Pro, Ser, and Thr may not be directed to the catabolic degradation involving 2-oxoacid:ferredoxin oxidoreductases and NDP-forming acyl-CoA synthetases, and this may be related to the notably low Asp aminotransferase activity observed in *T. kodakarensis* cell extracts and the TK2268 protein. As we have previously shown that *T. kodakarensis* does not harbor an Asp dehydrogenase, our present knowledge suggests that Asp is not catabolized through the pathway involving 2-oxoacid:ferredoxin oxidoreductases and NDP-forming acyl-CoA synthetases. We also could not find an aspartate ammonia lyase homolog on the genome, which would lead to the generation of fumarate. In addition, *T. kodakarensis* possesses a fumarase homolog, but lacks a number of homologs of the citric acid cycle including succinate dehydrogenase, malate dehydrogenase and citrate synthase. Other than the TK2268 protein, enzymes potentially related to oxaloacetate metabolism are a putative fumarase, malic enzyme (Fukuda et al., 2005), and phosphoenolpyruvate carboxykinase (Fukuda et al., 2004). *T. kodakarensis* may display a C4-compound metabolism distinct to those found in bacteria. Further understanding of oxaloacetate metabolism may provide valuable clues to elucidate the metabolism of Asp and the physiological function of TK2268. We would like to note, however, that we cannot rule out the possibility that the TK2268 protein recognizes a substrate completely different from those considered in this study.

Materials and methods

Strains and culture conditions

Thermococcus kodakarensis was isolated from Kodakara Island, Kagoshima, Japan (Morikawa et al., 1994; Atomi et al., 2004).

T. kodakarensis KU216 (Sato et al., 2003, 2005) and derivative strains were cultivated under strictly anaerobic conditions at 85°C in nutrient-rich medium (ASW-YT-m1-S⁰ or ASW-YT-m1-pyruvate) or synthetic medium (ASW-AA-m1-S⁰). ASW-YT-m1-S⁰, ASW-YT-m1-pyruvate, and ASW-AA-m1-S⁰ are modified versions of ASW-YT-S⁰, ASW-YT-pyruvate, and ASW-AA-S⁰ media, respectively. ASW-YT-S⁰ was composed of 0.8 × artificial seawater (ASW) (Robb and Place, 1995), 5 g L⁻¹ yeast extract, 5 g L⁻¹ tryptone, and 2 g L⁻¹ elemental sulfur. In ASW-YT-m1-S⁰, 20 μM KI, 20 μM H₃BO₃, 10 μM NiCl₂, and 10 μM Na₂WO₄ were supplemented. In ASW-YT-m1-pyruvate medium, elemental sulfur was replaced with 5 g L⁻¹ sodium pyruvate. ASW-AA-S⁰ was composed of 0.8 × ASW, a mixture of 20 amino acids, modified Wolfe's trace minerals and a mixture of vitamins (Sato et al., 2003). In ASW-AA-m1-S⁰, 20 μM KI, 20 μM H₃BO₃, 10 μM NiCl₂, and 10 μM Na₂WO₄ were supplemented, and the concentrations of L-arginine hydrochloride and L-valine were increased (from 125 mg L⁻¹ to 250 mg L⁻¹ and from 50 mg L⁻¹ to 200 mg L⁻¹, respectively). When cells without a *pyrF* gene were grown, 10 μg mL⁻¹ uracil was added to make ASW-AA-m1-S⁰(+Ura). To remove oxygen in the medium, 5% (w/v) Na₂S solution was added until the medium became colorless. Resazurine (0.5 mg L⁻¹) was also added to all media as an oxygen indicator. For solid medium used to isolate transformants, 10 g L⁻¹ gelrite, 7.5 g L⁻¹ 5-fluoroorotic acid (5-FOA), 10 μg mL⁻¹ uracil, 4.5 mL of 1 M NaOH and 0.2% (v/v) polysulfide solution (10 g Na₂S 9H₂O and 3 g sulfur flowers in 15 mL H₂O) rather than elemental sulfur was supplemented to ASW-AA-m1 medium. *Escherichia coli* DH5α (Takara Bio, Kusatsu, Japan) and BL21-Codonplus(DE3)-RIL strains (Agilent Technologies, Santa Clara, CA) were cultivated at 37°C in Lysogeny broth (LB) medium supplemented with ampicillin (100 mg L⁻¹). *E. coli* DH5α was used for recombinant plasmid construction and *E. coli* BL21-Codonplus (DE3)-RIL was used for heterologous gene expression. Chemicals were purchased from Wako Pure Chemicals (Osaka, Japan) or Nacalai Tesque (Kyoto, Japan) unless mentioned otherwise.

Expression of the TK0548 and TK2268 genes

In this study, pET21a(+) was used as an expression vector for TK0548 which was amplified from genomic DNA of *T. kodakarensis* KU216 using the primer set TK0548F/TK0548R (Table 4). The restriction enzyme sites *Nde*I and *Bam*HI were incorporated into the 5'- and 3'-termini of the fragments, respectively, during PCR. The amplified product and pET21a(+) were digested with *Nde*I and *Bam*HI, ligated using Ligation high Ver. 2, and introduced into *E. coli* DH5α cells. Positive colonies were selected by PCR analysis and confirmed by DNA sequencing. Plasmids were introduced into *E. coli* BL21-Codonplus (DE3)-RIL for gene expression.

The TK2268 gene was amplified from the genomic DNA of *T. kodakarensis* KU216 using the primer set TK2268F/R (Table 4). *Nde*I and *Sal*I sites as well as a sequence to introduce a C-terminal His₆-tag were incorporated during amplification. The amplified product was digested with *Nde*I and *Sal*I and inserted into a *T. kodakarensis*-*E. coli* shuttle plasmid previously used for heterologous expression of TK2101 and TK1211 (Zheng et al., 2018, 2021). After confirming the absence of unintended mutations by DNA sequencing, the plasmid was introduced into *T. kodakarensis* KPD2 (Δ *pyrF*,

TABLE 4 Primers used in the study.

Primer	Sequence (5' to 3')	Use
TK0548F	AAAAACATATGGCGCTGAGCGACAGGCTTGACC	Construction of expression plasmids
TK0548R	AAAGGATCCTTAAACGAGCTTTTCTCCTTCAGG	
TK2268F	AAAAAAACATATGAGGTATAAGAAGAGAAAGTAC	
TK2268R	AAAGTCGACTCAGTGGTGGTGGTGGTGGTGCAGCTTCGAGAGGGCCT	
TK0548seqF1	TCAGCGAGCTTATGCTCAAG	Sequencing
TK0548seqR1	GCATTGGAAGGCCCGCTTCGG	
TK2268seqF1	CCCCAACGCAGGCATCC	
TK2268seqF2	CCAAAATAGCCGAGGCAGGG	
TK2268seqF3	TACATCCTCAGCGACGAGCCCC	
TK2268seqF4	CGGCAACGT'TACGTCCTTCATCC	
TK2268seqF5	TCGATATGACCAGCGAAGAC	
TK2268seqR1	CCCTGCCTCGGCTATTTTGG	
dTK0548seqF1	GGCAGCCCTGAGTGAGGGCGTTG	
dTK0548seqF2	GGCGGTGTTGATGCGTTCTTC	
dTK0548seqF3	ATTGAAATTCGCGAGAGCCATTG	
dTK0548seqF4	CCGCGTCAACTTCGAGGCTCTC	
dTK0548seqF5	GTGATACAATATGGAGAAGG	
dTK0548seqR1	CATAATCCTCTCCACGCTCCTCCGAAG	
dTK0548seqR2	CCGGGGACATAACGAACTTC	
dTK0548seqR3	CCTTCTCCATATTGTATCAC	
dTK0548seqR4	CCTCACCACTCCCCAAAGTC	
TK0548seqF3	ATGGTAATTAGCGATGAAGTTTACG	
TK0548seqF4	CCTGTCCGGTCACCTTCGC	
TK0548seqF5	TCAGCGAGCTTATGCTCAAG	
TK0548seqR2	GAACGCCTGGTTGGCTCCGGTTAGG	
TK0548seqF6	CCGAACGCGGGCCTTCCAATGC	
TK0548seqF2	AAAGCCCGTTGAGGTTCC	
dTK2268seqF1	GAATGACCTCACTCCTGTTTC	
dTK2268seqF2	CTGAACGATGAGACAGGTGATTG	
dTK2268seqF3	CCCACCAGATAAGGGAAGCAATAAAAG	
dTK2268seqF4	TTGTGCGAGAAAGACCCTTCGCCAAG	
dTK2268seqR1	ATCCCCCTCACATTTTCC	
dTK2268seqR2	AGTGCTATCGCGACGCTCCTC	
dTK2268seqR3	TTCCAGCCAACCGCTTCGAC	
dTK2268seqR4	CTCCCGCAGGGGCTAATAAACG	
dTK0548F1	GAGAATTGAAACAAAAAGGTGGTGGTG	Construction of disruption plasmids
dTK0548R1	GAGATCAGACTGCGGGAGCGCTCTGCTTG	
dTK0548F2	ATCAGGATAAGCTACGCCACGGCTAC	
dTK0548R2	TACCATCACCGCTCTCCCTTC	
dTK0548inF	CGCTGGGATGCAGGATGTTATC	
dTK0548inR	GTAGCCTTCTCCGGCTTTC	
dTK0548outF	ATTGAAACTATAGGGAAATATAGGG	
dTK0548outR	TAATCCCGAAGAAGGGCTACTC	
dTK2268F1	ATCGTCGTCATCGAACTGAG	
dTK2268R1	TGGAGACCTGGTCTTTGTCC	
dTK2268F2	ATGGAAAACCTGCTTGCCGTCTTCGTTTC	
dTK2268R2	CGATGTTCCCCCGGGCAGTTCGGGAATG	
dTK2268inF	ATGAGGTATAAGAAGAGAAAGTACTTCATGGCCGGCAGGATAA	
dTK2268inR	GGAACCACTCACAGCTTCGAGAGGGCCTCTTTCAATC	
dTK2268outF	GGATCGCCGCCAGAACCTTTC	
dTK2268outR	CATGGCATAGTTCCTGCAAGCATTG	

$\Delta pdaD$, $\Delta chiA$) for gene expression. *pdaD* corresponds to TK0149, and disruption of this gene results in agmatine auxotrophy (Fukuda et al., 2008). For transformation, *T. kodakarensis* KPD2 was grown in ASW-YT-m1-S⁰ medium supplemented with agmatine (1.0 mM) at 85°C for 12 h. Cells were harvested by centrifugation (12,000 × g, 5 min, 4°C), and resuspended in 200 µL 0.8 × ASW-m1, followed by incubation on ice for 30 min. After mixing with 3.0 µg of the expression plasmid, the mixture was further incubated on ice for 1 h. Cells were inoculated into 20 mL ASW-YT-m1-S⁰ medium. After incubation at 85°C for 24 h, a 200-µL aliquot was further inoculated into 20 mL ASW-YT-m1-S⁰ medium. After incubation at 85°C 24 h, cells were spread onto solid ASW-YT-m1-S⁰ medium. After incubation at 85°C for 24 h, transformants displaying agmatine prototrophy were isolated and cultivated in ASW-YT-m1-S⁰. The presence of recombinant plasmids and absence of unintended mutation were confirmed by PCR and DNA sequencing, respectively.

Purification of the recombinant proteins

Transformants with TK0548 expression plasmid were cultivated in LB medium (100 mg L⁻¹ ampicillin and 30 mg L⁻¹ chloramphenicol) at 37°C until the OD₆₀₀ reached 0.6. Heterologous gene expression was induced by adding isopropyl-1-thio-β-D-galactopyranoside (IPTG) to a final concentration 0.1 mM followed by cultivation at 18°C for 20 h. Cells were harvested via centrifugation (12,000 × g, 15 min, 4°C), and resuspended in 50 mM HEPES buffer (containing 150 mM NaCl, pH 7.5). Sonication was used to lyse the cells and the insoluble cell debris was separated by centrifugation at 12,000 × g for 15 min at 4°C. The soluble cell extract was heat treated at 85°C for 15 min, and the thermolabile proteins derived from the host were removed by centrifugation (12,000 × g, 15 min, 4°C). The supernatant was loaded onto an anion exchange column (Resource Q) equilibrated with 50 mM HEPES buffer (pH 7.5). Protein was eluted with a linear gradient of 0 to 1.0 M NaCl. Fractions that contain TK0548 protein were collected and concentrated with an Amicon Ultra centrifugal filter unit (MWCO 10000). The resulting protein solution was applied to a Superdex 200 10/300 G1 gel-filtration column (GE Healthcare) with a mobile phase of 50 mM HEPES buffer (containing 150 mM NaCl, pH 7.5) at a flow rate of 0.7 mL min⁻¹.

The TK2268 gene expression strain was cultivated in ASW-YT-m1-pyruvate medium at 85°C for 20 h, and cells were collected by centrifugation (6,000 × g, 15 min, 4°C). After washing with 0.8 × ASW-m1, cells were resuspended in binding buffer (50 mM HEPES buffer, 20 mM imidazole, 500 mM KCl, 10% (v/v) glycerol, pH 7.5), then disrupted by sonication. Insoluble cell debris was removed by centrifugation (12,000 × g, 15 min, 4°C). The soluble cell extract was applied to a His GraviTrap column (GE Healthcare) which had been equilibrated with binding buffer. The TK2268 protein with a His₆-tag at its C terminus was eluted by elution buffer (50 mM HEPES buffer, 500 mM imidazole, 500 mM KCl, 10% (v/v) glycerol, pH 7.5). After concentrating the eluate, it was applied to a Superdex 200 10/300 G1 gel-filtration column (GE Healthcare). The TK2268 protein was eluted with a mobile phase of 50 mM HEPES buffer (pH 7.5) containing 500 mM KCl and 10% (v/v) glycerol, at a flow rate of 0.7 mL min⁻¹.

For examining the molecular mass of proteins, Blue 2000 was used to examine the void volume of the column, and ribonuclease A (13.7 kDa), carbonic anhydrase (29 kDa), ovalbumin (44 kDa),

conalbumin (75 kDa), aldolase (158 kDa), and ferritin (440 kDa; GE Healthcare) were used as standards. The mobile phase was 50 mM HEPES buffer (pH 7.5) containing 500 mM KCl, 10% (v/v) glycerol, and the flow rate was 0.7 mL min⁻¹. Protein concentrations were determined with a Protein Assay kit (Bio-Rad, Hercules, CA) using bovine serum albumin as standard.

Construction of gene disruption strains Δ TK0548 and Δ TK2268

Gene disruption strains were constructed using *T. kodakarensis* KU216 ($\Delta pyrF$), which shows uracil auxotrophy, as a host strain. For disrupting the TK0548 gene, the region from the start codon to base number 1080 of TK0548 gene was deleted instead of the stop codon to avoid disturbing expression of the overlapping downstream gene. In the case of TK2268, the entire coding region, along with 9 bases of its 3'-flanking region, was deleted. The TK0548 and TK2268 genes along with their 5'- and 3'-flanking regions (~1.0 kbp) were amplified from the genome of *T. kodakarensis* KU216 using the primer sets dTK0548F1/R1 and dTK2268F1/R1 (Table 4). The amplified products were inserted in the *HincII* site of the plasmid pUD3 which contains the *pyrF* gene of *T. kodakarensis* inserted in the *ApaI* site of pUC118 (Yokooji et al., 2009). Inverse PCR was performed with the primer sets dTK0548F2/0548R2 and dTK2268F2/2268R2 (Table 4) to remove sequences from the recombinant plasmid. The sequences of relevant regions were confirmed by DNA sequencing.

Thermococcus kodakarensis KU216 was cultivated in ASW-YT-m1-S⁰ medium for 12 h. Cells were harvested and resuspended in 200 µL of 0.8 × ASW-m1, then kept on ice for 30 min. After addition of 3.0 µg of the disruption plasmid and further incubation on ice for 1 h, cells were cultivated in ASW-AA-m1-S⁰ medium without uracil for 48 h at 85°C. A 200 µL aliquot was inoculated into fresh ASW-AA-m1-S⁰ medium and further cultivated under the same conditions to enrich transformants displaying uracil prototrophy. The culture was spread onto ASW-YT-m1 solid medium supplemented with 7.5 g L⁻¹ 5-FOA and 60 mM NaOH. Only cells that have undergone a pop-out recombination can grow in the presence of 5-FOA. After cultivation at 85°C for 48 h, colonies were selected, and their genotypes were analyzed by PCR using primer sets dTK0548outF/0548outR and dTK2268outF/2268outR (Table 4). Transformants that led to amplification of DNA products with the expected size were chosen and cultivated in ASW-YT-m1-S⁰ medium. Gene disruption was also confirmed by DNA sequencing.

Enzyme activity measurements

Initial examination of the aminotransferase activity of TK0548 and TK2268 proteins were carried out with Glu and pyruvate as amino donor and amino acceptor, respectively. Aminotransferase activity was measured at 80°C unless mentioned otherwise. The standard reaction mixture of TK0548 protein contained 20 µM PLP, 10 mM Glu, 10 mM pyruvate, 6.24 mM NaCl, and 0.4 µg mL⁻¹ recombinant protein in 50 mM HEPES buffer (pH 7.4). The standard reaction mixture of TK2268 protein contained 20 µM PLP, 10 mM Glu, 10 mM pyruvate, 8.8 mM KCl, 0.18% (v/v) glycerol and 4 µg mL⁻¹ recombinant protein in 50 mM HEPES buffer (pH 7.4). When the PLP-dependency of

activity was measured, PLP was omitted and activity with or without 10 mM hydroxylamine was measured. When PLP was omitted without addition of hydroxylamine, the amino donor and acceptor were 10 mM Leu and 10 mM 2-oxoglutarate, respectively. In the presence of hydroxylamine, the TK0548 protein reaction was measured with Phe and 2-oxoglutarate, while the TK2268 protein reaction was measured with Asp and pyruvate. The reaction mixture without amino donor and amino acceptor was pre-incubated at 80°C for 2 min, and the amino donor and acceptor were added to start the reaction. After further incubation at 80°C for 5 or 15 min, the reaction was stopped through cooling the reaction mixture on ice for 10 min. Proteins were removed with an Amicon Ultra-0.5 centrifugal filter unit with an Ultracel-10 membrane (Millipore). The formation of Glu and Ala was detected and quantified by HPLC after derivatization. The derivatization mixture (100 μ L) contained 10 μ L of reaction mixture, 70 μ L of solution B [borate sodium hydroxide buffer (0.4 M, pH 10.4)] and 20 μ L of solution A (8 mg *o*-phthalaldehyde and 10 mg *N*-acetylcysteine were dissolved in 1 mL methanol). After derivatization for 5 min at room temperature, an aliquot (10 μ L) of the solution was applied to a COSMOSIL 5C18-PAQ packed column (4.6 ID \times 250 mm) using a Nexera X2 liquid chromatography system with a fluorescence detector RF-20A XS (Shimadzu, Kyoto, Japan). Compounds were eluted with a solution of 20 mM sodium acetate (pH 5.6) and methanol at a flow rate of 0.7 mL min⁻¹. The excitation and emission wavelength were 350 and 450 nm, respectively.

To screen the amino acids recognized by TK0548 and TK2268 proteins, 20 amino acids were used as amino donor, and 2-oxoglutarate or pyruvate was used as amino acceptor. To analyze the aminotransferase activity of TK0548 protein, Tyr, Phe, Trp, His, Leu, Met, and Glu were chosen as amino donor, and 2-oxoglutarate or pyruvate was used as amino acceptor. For the TK2268 protein, Glu, Asp, Cys, Leu, Ala, Met, and Tyr were chosen as amino donor, and 2-oxoglutarate or pyruvate was used as amino acceptor. The standard reaction mixture was incubated at 80°C for 3, 5, and 7 min (reaction mixture with His and Glu were incubated at 80°C for 1, 2, 3 min) to confirm that product formation was linear with time.

For kinetic analysis of the TK0548 protein reaction, reaction rates with various concentrations of Phe, Trp, Tyr, Met, and His were examined with 10 mM 2-oxoglutarate. Reaction rates with various concentrations of 2-oxoglutarate and pyruvate were examined with 10 mM Phe. For analysis of the TK2268 protein, reaction rates with various concentrations of Asp, Glu, Tyr, and Leu were examined with 10 mM pyruvate or 10 mM 2-oxoglutarate. Reaction rates with various concentrations of 2-oxoglutarate and pyruvate were examined with 10 mM Leu. Kinetic parameters were obtained by fitting the data to the Michaelis–Menten equation using IGORPRO, version 6.03 (Wave-Metrics, Lake, Oswego, OR).

Growth measurements

Growth properties of the host strain KU216 and the Δ TK0548 and Δ TK2268 gene disruption strains were examined in ASW-YT-m1-S⁰ and ASW-AA-m1-S⁰(+Ura) media. Cells were precultured in the nutrient-rich medium ASW-YT-m1-S⁰ for 15 h until the stationary phase and inoculated into ASW-YT-m1-S⁰ medium or synthetic medium ASW-AA-m1-S⁰(+Ura). The OD₆₆₀ of the culture was monitored.

Activity measurements in cell-free extracts

Thermococcus kodakarensis KU216, Δ TK0548 and Δ TK2268 disruption strains were cultivated at ASW-YT-m1-pyruvate medium for 20 h and cells were collected by centrifugation (6,000 \times g, 15 min, 4°C). Cells were disrupted by sonication and insoluble cell debris was removed (12,000 \times g, 30 min, 4°C). After exchanging the buffer with 50 mM HEPES (containing 150 mM NaCl, pH7.4) using Amicon Ultra centrifugal filter unit (MWCO 10000), the aminotransferase activity in the cell-free extract was measured. For the aminotransferase activity toward His, Tyr, and Asp, the reaction mixture contained 20 μ M PLP, 10 mM amino donor (the final concentration of Tyr was 6 mM), 10 mM amino acceptor (2-oxoglutarate or pyruvate), 5.76 mM NaCl and 0.384 mg mL⁻¹ cell-free extracts in 50 mM HEPES buffer (pH7.4). For Trp, the reaction mixture contained 20 μ M PLP, 10 mM amino donor, 10 mM amino acceptor (2-oxoglutarate), 2.88 mM NaCl and 0.182 mg mL⁻¹ cell-free extracts in 50 mM HEPES buffer (pH7.4). In the case of Glu, Ile, and Phe, the reaction mixture contained 20 μ M PLP, 10 mM amino donor, 10 mM amino acceptor (2-oxoglutarate or pyruvate), 0.576 mM NaCl and 0.0384 mg mL⁻¹ cell-free extracts in 50 mM HEPES buffer (pH7.4). The reaction mixture without amino acceptor was pre-incubated at 80°C for 2 min, and the amino acceptor was added to start the reaction. After further incubation at 80°C for 1, 2, and 3 min (When Asp and Ile were used as amino donor, the reaction mixtures were incubated at 80°C for 4, 6, and 8 min and 2, 4, and 6 min, respectively). The reaction was stopped through cooling the reaction mixture on ice for 10 min. Proteins were removed with an Amicon Ultra-0.5 centrifugal filter unit with an Ultracel-10 membrane (Millipore). The formation of Glu and Ala was determined by HPLC after derivatization.

Data availability statement

The original contributions presented in the study are included in the article/[Supplementary material](#), further inquiries can be directed to the corresponding author.

Author contributions

HA designed the experiments. YS carried out the biochemical and genetic experiments. YM carried out the bioinformatic analyses. All authors contributed in data analyses, writing the manuscript, and approved the submitted version.

Funding

This study was partially supported by JSPS KAKENHI grant numbers JP19H05679 (Post-Koch Ecology) and JP19H05684 to HA. This work was partially supported by JST SPRING, Grant Number JPMJSP2110 and JST, the establishment of university fellowships toward the creation of science technology innovation, Grant Number JPMJFS2123.

Conflict of interest

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.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated

organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2023.1126218/full#supplementary-material>

References

- Andreotti, G., Cubellis, M. V., Nitti, G., Sannia, G., Mai, X., Adams, M. W., et al. (1995). An extremely thermostable aromatic aminotransferase from the hyperthermophilic archaeon *Pyrococcus furiosus*. *Biochim. Biophys. Acta* 1247, 90–96. doi: 10.1016/0167-4838(94)00211-X
- Andreotti, G., Cubellis, M. V., Nitti, G., Sannia, G., Mai, X., Marino, G., et al. (1994). Characterization of aromatic aminotransferases from the hyperthermophilic archaeon *Thermococcus litoralis*. *Eur. J. Biochem.* 220, 543–549. doi: 10.1111/j.1432-1033.1994.tb18654.x
- Atomi, H., Fukui, T., Kanai, T., Morikawa, M., and Imanaka, T. (2004). Description of *Thermococcus kodakaraensis* sp. nov., a well studied hyperthermophilic archaeon previously reported as *Pyrococcus* sp. KOD1. *Archaea* 1, 263–267. doi: 10.1155/2004/204953
- Awano, T., Wilming, A., Tomita, H., Yokooji, Y., Fukui, T., Imanaka, T., et al. (2014). Characterization of two members among the five ADP-forming acyl coenzyme A (acyl-CoA) synthetases reveals the presence of a 2-(imidazol-4-yl)acetyl-CoA synthetase in *Thermococcus kodakaraensis*. *J. Bacteriol.* 196, 140–147. doi: 10.1128/JB.00877-13
- Blamey, J. M., and Adams, M. W. (1993). Purification and characterization of pyruvate ferredoxin oxidoreductase from the hyperthermophilic archaeon *Pyrococcus furiosus*. *Biochim. Biophys. Acta* 1161, 19–27. doi: 10.1016/0167-4838(93)90190-3
- Chon, H., Matsumura, H., Koga, Y., Takano, K., and Kanaya, S. (2005). Crystal structure of a human kynurenine aminotransferase II homologue from *Pyrococcus horikoshii* OT3 at 2.20 Å resolution. *Proteins* 61, 685–688. doi: 10.1002/prot.20614
- 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
- Fiala, G., and Stetter, K. O. (1986). *Pyrococcus furiosus* sp. nov. represents a novel genus of marine heterotrophic archaeobacteria growing optimally at 100°C. *Arch. Microbiol.* 145, 56–61. doi: 10.1007/BF00413027
- Fukuda, W., Fukui, T., Atomi, H., and Imanaka, T. (2004). First characterization of an archaeal GTP-dependent phosphoenolpyruvate carboxykinase from the hyperthermophilic archaeon *Thermococcus kodakaraensis* KOD1. *J. Bacteriol.* 186, 4620–4627. doi: 10.1128/JB.186.14.4620-4627.2004
- Fukuda, W., Ismail, Y. S., Fukui, T., Atomi, H., and Imanaka, T. (2005). Characterization of an archaeal malic enzyme from the hyperthermophilic archaeon *Thermococcus kodakaraensis* KOD1. *Archaea* 1, 293–301. doi: 10.1155/2005/250757
- Fukuda, W., Morimoto, N., Imanaka, T., and Fujiwara, S. (2008). Agmatine is essential for the cell growth of *Thermococcus kodakaraensis*. *FEMS Microbiol. Lett.* 287, 113–120. doi: 10.1111/j.1574-6968.2008.01303.x
- Fukui, T., Atomi, H., Kanai, T., Matsumi, R., Fujiwara, S., and Imanaka, T. (2005). Complete genome sequence of the hyperthermophilic archaeon *Thermococcus kodakaraensis* KOD1 and comparison with *Pyrococcus* genomes. *Genome Res.* 15, 352–363. doi: 10.1101/gr.3003105
- Glasemacher, J., Bock, A. K., Schmid, R., and Schönheit, P. (1997). Purification and properties of acetyl-CoA synthetase (ADP-forming), an archaeal enzyme of acetate formation and ATP synthesis, from the hyperthermophile *Pyrococcus furiosus*. *Eur. J. Biochem.* 244, 561–567. doi: 10.1111/j.1432-1033.1997.00561.x
- Gomm, A., and O'Reilly, E. (2018). Transaminases for chiral amine synthesis. *Curr. Opin. Chem. Biol.* 43, 106–112. doi: 10.1016/j.cbpa.2017.12.007
- Heider, J., Mai, X., and Adams, M. W. (1996). Characterization of 2-ketoisovalerate ferredoxin oxidoreductase, a new and reversible coenzyme A-dependent enzyme involved in peptide fermentation by hyperthermophilic archaea. *J. Bacteriol.* 178, 780–787. doi: 10.1128/jb.178.3.780-787.1996
- Hidese, R., Inoue, T., Imanaka, T., and Fujiwara, S. (2014). Cysteine desulphurase plays an important role in environmental adaptation of the hyperthermophilic archaeon *Thermococcus kodakaraensis*. *Mol. Microbiol.* 93, 331–345. doi: 10.1111/mmi.12662
- Janssonius, J. N. (1998). Structure, evolution and action of vitamin B6-dependent enzymes. *Curr. Opin. Struct. Biol.* 8, 759–769. doi: 10.1016/S0959-440X(98)80096-1
- Jones, D. T., Taylor, W. R., and Thornton, J. M. (1992). The rapid generation of mutation data matrices from protein sequences. *Comput. Appl. Biosci.* 8, 275–282. doi: 10.1093/bioinformatics/8.3.275
- Kanai, T., Simons, J. R., Tsukamoto, R., Nakajima, A., Omori, Y., Matsuoka, R., et al. (2015). Overproduction of the membrane-bound [NiFe]-hydrogenase in *Thermococcus kodakaraensis* and its effect on hydrogen production. *Front. Microbiol.* 6:847. doi: 10.3389/fmicb.2015.00847
- Kawakami, R., Kinoshita, C., Kawase, T., Sato, M., Hayashi, J., Sakuraba, H., et al. (2021). Characterization of a novel moderate-substrate specificity amino acid racemase from the hyperthermophilic archaeon *Thermococcus litoralis*. *Biosci. Biotechnol. Biochem.* 85, 1650–1657. doi: 10.1093/bbb/zbab078
- Kawakami, R., Ohmori, T., Sakuraba, H., and Ohshima, T. (2015). Identification of a novel amino acid racemase from a hyperthermophilic archaeon *Pyrococcus horikoshii* OT-3 induced by D-amino acids. *Amino. Acids.* 47, 1579–1587. doi: 10.1007/s00726-015-2001-6
- Kawakami, R., Ohshida, T., Sakuraba, H., and Ohshima, T. (2018). A novel PLP-dependent alanine/serine racemase from the hyperthermophilic archaeon *Pyrococcus horikoshii* OT-3. *Front. Microbiol.* 9:1481. doi: 10.3389/fmicb.2018.01481
- Kawakami, R., Ohshida, T., Hayashi, J., Yoneda, K., Furumoto, T., Ohshima, T., et al. (2022). Crystal structure of a novel type of ornithine d-aminotransferase from the hyperthermophilic archaeon *Pyrococcus horikoshii*. *Int. J. Biol. Macromol.* 208, 731–740. doi: 10.1016/j.ijbiomac.2022.03.114
- Kawakami, R., Sakuraba, H., Ohmori, T., and Ohshima, T. (2017). First characterization of an archaeal amino acid racemase with broad substrate specificity from the hyperthermophile *Pyrococcus horikoshii* OT-3. *J. Biosci. Bioeng.* 124, 23–27. doi: 10.1016/j.jbiosc.2017.02.004
- Kelly, S. A., Mix, S., Moody, T. S., and Gilmore, B. F. (2020). Transaminases for industrial biocatalysis: novel enzyme discovery. *Appl. Microbiol. Biotechnol.* 104, 4781–4794. doi: 10.1007/s00253-020-10585-0
- Kito, K., Sanada, Y., and Katunuma, N. (1978). Mode of inhibition of ornithine aminotransferase by L-canaline. *J. Biochem.* 83, 201–206. doi: 10.1093/oxfordjournals.jbchem.a131892
- Koper, K., Han, S. W., Pastor, D. C., Yoshikuni, Y., and Maeda, H. A. (2022). Evolutionary origin and functional diversification of aminotransferases. *J. Biol. Chem.* 298:102122. doi: 10.1016/j.jbc.2022.102122
- Koszelewski, D., Tauber, K., Faber, K., and Kroutil, W. (2010). ω-Transaminases for the synthesis of non-racemic alpha-chiral primary amines. *Trends Biotechnol.* 28, 324–332. doi: 10.1016/j.tibtech.2010.03.003
- Kunutsor, S. K., Apekey, T. A., and Walley, J. (2013). Liver aminotransferases and risk of incident type 2 diabetes: a systematic review and meta-analysis. *Am. J. Epidemiol.* 178, 159–171. doi: 10.1093/aje/kws469
- Mai, X., and Adams, M. W. (1994). Indolepyruvate ferredoxin oxidoreductase from the hyperthermophilic archaeon *Pyrococcus furiosus*. *J. Biol. Chem.* 269, 16726–16732. doi: 10.1016/S0021-9258(19)89451-6
- Mai, X., and Adams, M. W. (1996). Purification and characterization of two reversible and ADP-dependent acetyl coenzyme A synthetases from the hyperthermophilic archaeon *Pyrococcus furiosus*. *J. Bacteriol.* 178, 5897–5903. doi: 10.1128/jb.178.20.5897-5903.1996
- Malik, M. S., Park, E. S., and Shin, J. S. (2012). Features and technical applications of ω-transaminases. *Appl. Microbiol. Biotechnol.* 94, 1163–1171. doi: 10.1007/s00253-012-4103-3
- Matsui, I., Matsui, E., Sakai, Y., Kikuchi, H., Kawarabayashi, Y., Ura, H., et al. (2000). The molecular structure of hyperthermostable aromatic aminotransferase with novel substrate specificity from *Pyrococcus horikoshii*. *J. Biol. Chem.* 275, 4871–4879. doi: 10.1074/jbc.275.7.4871

- Matsumi, R., Manabe, K., Fukui, T., Atomi, H., and Imanaka, T. (2007). Disruption of a sugar transporter gene cluster in a hyperthermophilic archaeon using a host-marker system based on antibiotic resistance. *J. Bacteriol.* 189, 2683–2691. doi: 10.1128/JB.01692-06
- Mehta, P. K., and Christen, P. (2000). The molecular evolution of pyridoxal-5'-phosphate-dependent enzymes. *Adv. Enzymol. Relat. Areas Mol. Biol.* 74, 129–184. doi: 10.1002/9780470123201.ch4
- Mehta, P. K., Hale, T. I., and Christen, P. (1993). Aminotransferases: demonstration of homology and division into evolutionary subgroups. *Eur. J. Biochem.* 214, 549–561. doi: 10.1111/j.1432-1033.1993.tb17953.x
- Montioli, R., Bellezza, I., Desbats, M. A., Borri Voltattorni, C., Salvati, L., and Cellini, B. (2021). Deficit of human ornithine aminotransferase in gyrate atrophy: molecular, cellular, and clinical aspects. *Biochim. Biophys. Acta Proteins Proteom.* 1869:140555. doi: 10.1016/j.bbapap.2020.140555
- Morikawa, M., Izawa, Y., Rashid, N., Hoaki, T., and Imanaka, T. (1994). Purification and characterization of a thermostable thiol protease from a newly isolated hyperthermophilic *Pyrococcus* sp. *Appl. Environ. Microbiol.* 60, 4559–4566. doi: 10.1128/aem.60.12.4559-4566.1994
- Musfeldt, M., Selig, M., and Schönheit, P. (1999). Acetyl coenzyme A synthetase (ADP forming) from the hyperthermophilic archaeon *Pyrococcus furiosus*: identification, cloning, separate expression of the encoding genes, *acdAI* and *acdBI*, in *Escherichia coli*, and *in vitro* reconstitution of the active heterotetrameric enzyme from its recombinant subunits. *J. Bacteriol.* 181, 5885–5888. doi: 10.1128/JB.181.18.5885-5888.1999
- Okada, K., Angkawidjaja, C., Koga, Y., and Kanaya, S. (2014). Structural and mechanistic insights into the kynurenine aminotransferase-mediated excretion of kynurenine acid. *J. Struct. Biol.* 185, 257–266. doi: 10.1016/j.jsb.2014.01.009
- Okada, K., Angkawidjaja, C., Koga, Y., Takano, K., and Kanaya, S. (2012). Characteristic features of kynurenine aminotransferase allosterically regulated by (alpha)-ketoglutarate in cooperation with kynurenine. *PLoS One* 7:e40307. doi: 10.1371/journal.pone.0040307
- Robb, F., and Place, A. (1995). *Media for Thermophiles*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.
- Sakuraba, H., Kawakami, R., Takahashi, H., and Ohshima, T. (2004). Novel archaeal alanine:glyoxylate aminotransferase from *Thermococcus litoralis*. *J. Bacteriol.* 186, 5513–5518. doi: 10.1128/JB.186.16.5513-5518.2004
- Sakuraba, H., Yoneda, K., Takeuchi, K., Tsuge, H., Katunuma, N., and Ohshima, T. (2008). Structure of an archaeal alanine: glyoxylate aminotransferase. *Acta Crystallogr. D Biol. Crystallogr.* 64, 696–699. doi: 10.1107/S0907444908006732
- Santangelo, T. J., Cubonova, L., and Reeve, J. N. (2008). Shuttle vector expression in *Thermococcus kodakaraensis*: contributions of *cis* elements to protein synthesis in a hyperthermophilic archaeon. *Appl. Environ. Microbiol.* 74, 3099–3104. doi: 10.1128/AEM.00305-08
- Santangelo, T. J., Cubonova, L., and Reeve, J. N. (2010). *Thermococcus kodakaraensis* genetics: TK1827-encoded β -glycosidase, new positive-selection protocol, and targeted and repetitive deletion technology. *Appl. Environ. Microbiol.* 76, 1044–1052. doi: 10.1128/AEM.02497-09
- Sato, T., Fukui, T., Atomi, H., and Imanaka, T. (2003). Targeted gene disruption by homologous recombination in the hyperthermophilic archaeon *Thermococcus kodakaraensis* KOD1. *J. Bacteriol.* 185, 210–220. doi: 10.1128/JB.185.1.210-220.2003
- Sato, T., Fukui, T., Atomi, H., and Imanaka, T. (2005). Improved and versatile transformation system allowing multiple genetic manipulations of the hyperthermophilic archaeon *Thermococcus kodakaraensis*. *Appl. Environ. Microbiol.* 71, 3889–3899. doi: 10.1128/AEM.71.7.3889-3899.2005
- Schneider, G., Kack, H., and Lindqvist, Y. (2000). The manifold of vitamin B6 dependent enzymes. *Structure* 8, R1–R6. doi: 10.1016/S0969-2126(00)00085-X
- Shikata, K., Fukui, T., Atomi, H., and Imanaka, T. (2007). A novel ADP-forming succinyl-CoA synthetase in *Thermococcus kodakaraensis* structurally related to the archaeal nucleoside diphosphate-forming acetyl-CoA synthetases. *J. Biol. Chem.* 282, 26963–26970. doi: 10.1074/jbc.M702694200
- Takai, K., Sugai, A., Itoh, T., and Horikoshi, K. (2000). *Palaeococcus ferrophilus* gen. nov., sp. nov., a barophilic, hyperthermophilic archaeon from a deep-sea hydrothermal vent chimney. *Int. J. Syst. Evol. Microbiol.* 50 Pt 2, 489–500. doi: 10.1099/00207713-50-2-489
- Tamura, K., Stecher, G., and Kumar, S. (2021). MEGA11: molecular evolutionary genetics analysis version 11. *Mol. Biol. Evol.* 38, 3022–3027. doi: 10.1093/molbev/msab120
- Tavares, N. K., Stracey, N., Brunold, T. C., and Escalante-Semerena, J. C. (2019). The L-Thr kinase/L-Thr-phosphate decarboxylase (CobD) enzyme from *Methanosarcina mazei* Go1 contains metallocenters needed for optimal activity. *Biochemistry* 58, 3260–3279. doi: 10.1021/acs.biochem.9b00268
- Tavares, N. K., Zayas, C. L., and Escalante-Semerena, J. C. (2018). The *Methanosarcina mazei* MM2060 gene encodes a bifunctional kinase/decarboxylase enzyme involved in Cobamide biosynthesis. *Biochemistry* 57, 4478–4495. doi: 10.1021/acs.biochem.8b00546
- Toyokawa, Y., Koonthongkaew, J., and Takagi, H. (2021). An overview of branched-chain amino acid aminotransferases: functional differences between mitochondrial and cytosolic isozymes in yeast and human. *Appl. Microbiol. Biotechnol.* 105, 8059–8072. doi: 10.1007/s00253-021-11612-4
- Uchida, Y., Hayashi, H., Washio, T., Yamasaki, R., Kato, S., and Oikawa, T. (2014). Cloning and characterization of a novel fold-type I branched-chain amino acid aminotransferase from the hyperthermophilic archaeon *Thermococcus* sp. CKU-1. *Extremophiles* 18, 589–602. doi: 10.1007/s00792-014-0642-0
- Ura, H., Harata, K., Matsui, I., and Kuramitsu, S. (2001). Temperature dependence of the enzyme-substrate recognition mechanism. *J. Biochem.* 129, 173–178. doi: 10.1093/oxfordjournals.jbchem.a002829
- Ward, D. E., De Vos, W. M., and Van Der Oost, J. (2002). Molecular analysis of the role of two aromatic aminotransferases and a broad-specificity aspartate aminotransferase in the aromatic amino acid metabolism of *Pyrococcus furiosus*. *Archaea* 1, 133–141. doi: 10.1155/2002/959031
- Ward, D. E., Kengen, S. W., Van Der Oost, J., and De Vos, W. M. (2000). Purification and characterization of the alanine aminotransferase from the hyperthermophilic archaeon *Pyrococcus furiosus* and its role in alanine production. *J. Bacteriol.* 182, 2559–2566. doi: 10.1128/JB.182.9.2559-2566.2000
- Yokooji, Y., Sato, T., Fujiwara, S., Imanaka, T., and Atomi, H. (2013). Genetic examination of initial amino acid oxidation and glutamate catabolism in the hyperthermophilic archaeon *Thermococcus kodakaraensis*. *J. Bacteriol.* 195, 1940–1948. doi: 10.1128/JB.01979-12
- Yokooji, Y., Tomita, H., Atomi, H., and Imanaka, T. (2009). Pantoate kinase and phosphopantothenate synthetase, two novel enzymes necessary for CoA biosynthesis in the archaea. *J. Biol. Chem.* 284, 28137–28145. doi: 10.1074/jbc.M109.009696
- Yoshida, A., Tomita, T., Atomi, H., Kuzuyama, T., and Nishiyama, M. (2016). Lysine biosynthesis of *Thermococcus kodakaraensis* with the capacity to function as an ornithine biosynthetic system. *J. Biol. Chem.* 291, 21630–21643. doi: 10.1074/jbc.M116.743021
- Zheng, R. C., Hachisuka, S. I., Tomita, H., Imanaka, T., Zheng, Y. G., Nishiyama, M., et al. (2018). An ornithine ω -aminotransferase required for growth in the absence of exogenous proline in the archaeon *Thermococcus kodakaraensis*. *J. Biol. Chem.* 293, 3625–3636. doi: 10.1074/jbc.RA117.001222
- Zheng, R. C., Lu, X. F., Tomita, H., Hachisuka, S. I., Zheng, Y. G., and Atomi, H. (2021). TK1211 encodes an amino acid racemase towards leucine and methionine in the hyperthermophilic archaeon *Thermococcus kodakaraensis*. *J. Bacteriol.* 203, e00617–e00620. doi: 10.1128/JB.00617-20
- Zillig, W., Holz, I., Janekovic, D., Schafer, W., and Reiter, W. D. (1983). The archaeobacterium *Thermococcus celer* represents, a novel genus within the thermophilic branch of the archaeobacteria. *Syst. Appl. Microbiol.* 4, 88–94. doi: 10.1016/S0723-2020(83)80036-8



OPEN ACCESS

EDITED BY

Arthur Charles-Orszag,
University of California,
San Francisco,
United States

REVIEWED BY

Zhenfeng Zhang,
Institute of Microbiology (CAS),
China
Haruyuki Atomi,
Kyoto University,
Japan

*CORRESPONDENCE

Swati Saha

✉ ss5gp@yahoo.co.in;

✉ ssaha@south.du.ac.in

SPECIALTY SECTION

This article was submitted to
Biology of Archaea,
a section of the journal
Frontiers in Microbiology

RECEIVED 18 December 2022

ACCEPTED 27 February 2023

PUBLISHED 15 March 2023

CITATION

Gulati P, Singh A, Goel M and Saha S (2023) The
extremophile *Picrophilus torridus* carries a DNA
adenine methylase M.Ptol that is part of a Type
I restriction-modification system.
Front. Microbiol. 14:1126750.
doi: 10.3389/fmicb.2023.1126750

COPYRIGHT

© 2023 Gulati, Singh, Goel and Saha. This is an
open-access article distributed under the terms
of the [Creative Commons Attribution License
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). 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 extremophile *Picrophilus torridus* carries a DNA adenine methylase M.Ptol that is part of a Type I restriction-modification system

Pallavi Gulati¹, Ashish Singh¹, Manisha Goel² and Swati Saha^{1*}

¹Department of Microbiology, University of Delhi South Campus, New Delhi, India, ²Department of Biophysics, University of Delhi South Campus, New Delhi, India

DNA methylation events mediated by orphan methyltransferases modulate various cellular processes like replication, repair and transcription. Bacteria and archaea also harbor DNA methyltransferases that are part of restriction-modification systems, which serve to protect the host genome from being cleaved by the cognate restriction enzyme. While DNA methylation has been exhaustively investigated in bacteria it remains poorly understood in archaea. *Picrophilus torridus* is a euryarchaeon that can thrive under conditions of extremely low pH (0.7), and thus far no reports have been published regarding DNA methylation in this extremophile. This study reports the first experimentation examining DNA methylation in *P. torridus*. We find the genome to carry methylated adenine (m6A) but not methylated cytosine (m5C) residues. The m6A modification is absent at GATC sites, indicating the absence of an active Dam methylase even though the *dam* gene has been annotated in the genome sequence. Two other methylases have also been annotated in the *P. torridus* genome sequence. One of these is a part of a Type I restriction-modification system. Considering that all Type I modification methylases characterized to date target adenine residues, the modification methylase of this Type I system has been examined. The genes encoding the S subunit (that is responsible for DNA recognition) and M subunit (that is responsible for DNA methylation) have been cloned and the recombinant protein purified from *E.coli*, and regions involved in M-S interactions have been identified. The M.Ptol enzyme harbors all the motifs that typify Type I modification methylases, and displays robust adenine methylation in *in vitro* assays under a variety of conditions. Interestingly, magnesium is essential for enzyme activity. The enzyme displays substrate inhibition at higher concentrations of AdoMet. Mutational analyses reveal that Motif I plays a role in AdoMet binding, and Motif IV is critical for methylation activity. The data presented here lays the foundation for further research in the area of DNA methylation and restriction-modification research in this most unusual microorganism.

KEYWORDS

DNA adenine methylase, Type I restriction-modification system, modification methylase, *Picrophilus torridus*, acidophile, extremophile

1. Introduction

DNA methylation typically occurs at adenine and cytosine residues. While methylation at adenine occurs exocyclically at the sixth position (N⁶-methyladenine: m6A), methylation at cytosine occurs either exocyclically at the fourth position (N⁴-methylcytosine: m4C), or endocyclically at the fifth position (5-methylcytosine: m5C). The primary DNA methylation mark in eukaryotes is m5C, while in bacteria the m6A methylation mark is the most widely prevalent. The marking of DNA by methylation has two distinctive roles: one, it modulates DNA interactions with proteins, thereby having an impact on DNA replication, DNA mismatch repair and gene expression, and second, it protects the bacterial and archaeal genome from self-destruction by their own restriction systems (Pingoud et al., 2014; Anton and Roberts, 2021).

The enzymes that mediate DNA methylation in bacteria and archaea use S-adenosyl methionine (AdoMet) as the methyl-group donor and are either orphan methyltransferases or a component of restriction-modification systems. Orphan methyltransferases (which do not have cognate restriction endonucleases) are associated with methylation events that regulate DNA-related transactions like replication and transcription. The Dam methylase which targets adenine methylation in the sequence 5'-GATC-3' is the best known of these (Adhikari and Curtis, 2016). Methylases that are a component of restriction-modification (R-M) systems are designed to protect the genome from being attacked by cognate restriction endonucleases harbored by the organism. R-M systems are linked to the cell's defense against incoming DNA, with the restriction endonuclease (RE) cleaving the foreign DNA that is not methylated at its specific recognition sequence. The host genome remains resistant to cleavage due to the recognition sites on the genome being methylated by the cognate modification methylase (MTase). Such methyltransferases can be a component of three types of R-M systems: Types I-III (Loenen et al., 2014b).

The Types I-III R-M systems differ in their subunit structure and organization, recognition sequence and cleavage site, and cofactor requirements. In the widely found Type II R-M systems, the modification methylase and the restriction endonuclease functions reside in independent enzymes that recognize the same palindromic sequence, with the endonuclease cleaving DNA within the (non-methylated) recognition sequence or just beyond. Type I and Type III R-M systems are complex multisubunit enzymes, with the modification methylase and restriction endonuclease functions residing in the same enzyme albeit different subunits (Rao et al., 2014; Loenen et al., 2014a). Type III R-M systems comprise two subunits: Mod and Res. The Mod is responsible for the recognition of the short asymmetric recognition sequence on DNA as well as its methylation, while the Res mediates DNA cleavage 25–27 bp downstream of the site. The Type I R-M systems comprise three subunits: S, M, and R, typically existing as pentamers of 1S, 2M, and 2R subunits. They recognize a specific bipartite DNA sequence *via* their S (specificity) subunit, modify the DNA at the recognition sequence *via* their M (modification) subunits, and cleave DNA through their R (restriction) subunits hundreds of base pairs away. While the restriction endonuclease exists in active form only as part of the multisubunit 1S2M2R enzyme, the modification methylase can also exist in active form independent of the R subunits, as a complex with the S subunit which it depends on for DNA sequence recognition (M2S1).

While DNA methylation and its effects have been extensively investigated in bacteria and eukaryotes, it remains a poorly understood area in archaea. The primary methylation marks in archaea are m6A and m4C, with m5C remaining undetected so far in these organisms. Single molecule real time (SMRT) sequencing has led to the analysis of the DNA methylomes of several bacteria and archaea. In a landmark effort, Blow et al. (2016) analyzed the DNA methylomes of 230 bacterial and archaeal species. Their study identified over 800 methylated motifs, emphasizing the extent of diversity among these groups of organisms (Blow et al., 2016). Other studies include the analysis of the DNA methylome of the hyperthermophilic archaeon *Thermococcus onnurineus* where over 2,000 sites of m6A were detected in addition to several sites of m4C (Lee et al., 2016), and the identification of DNA methylation at over 99% of the motifs 5'-GATC-3' and 5'-AGCT-3' in the genome of the marine thaumarchaeote *Candidatus Nitrososmarinus catalina* SPOT01 at the adenine and cytosine, respectively (Ahlgren et al., 2017). The DNA methylome of the hyperthermoacidophile *Sulfolobus acidocaldarius* was found to carry the m6A mark in the 5'-GATC-3' motif occurring in two different genome contexts: 5'-AGATCC-3' and 5'-GGATCT/C-3' (Couturier and Lindas, 2018), and the m4C mark was also detected at the 5'-GGCC-3' motif that was previously identified as the target site of the SuiI/M.SuiI restriction modification system (Prangishvili et al., 1985; Grogan, 2003).

Several R-M systems have been annotated in archaeal genomes but very few have been characterized. Among the earliest to be purified and studied were the Type II restriction endonuclease *ThaI* from *Thermoplasma acidophilum* (McConnell et al., 1978), three Type II restriction endonucleases from *Methanococcus aeolicus* (Schmid et al., 1984), a Type II R-M system from *Methanobacterium wolfei* (Lunnen et al., 1989), three Type II R-M systems from *Methanobacterium thermoformicicum* strains (Nolling and de Vos, 1992a,b), the PspG1 Type II R-M system from a *Pyrococcus* species (Morgan et al., 1998; Pingoud et al., 2003) and the SuiI/M.SuiI Type II R-M system in *Sulfolobus acidocaldarius* (Prangishvili et al., 1985; Grogan, 2003). Other Type II R-M systems that have been studied in archaea are PabI/M.PabI in *Pyrococcus abyssi* (Ishikawa et al., 2005; Watanabe et al., 2006) a type II restriction endonuclease SuiI from *Sulfolobus islandicus* (Suzuki and Kurosawa, 2016) and two Type IIG R-M systems TkoI and TkoII in *Thermococcus kodakarensis* (Zatopek et al., 2021). The only Type I R-M system studied in archaea to date is in *Haloferax volcanii* (Ouellette et al., 2018).

The present study was initiated with the aim of examining DNA methylation in the thermoacidophilic archaeon *Picrophilus torridus*. Originally isolated from a solfataric field in northern Japan, the *Picrophilus* species are the most acidophilic organisms isolated to date, growing optimally at a pH of 0.7 at 55–60°C, with an intracellular pH of ~4.6 (Schleper et al., 1995). When the 1.55 Mb genome of *Picrophilus torridus* was sequenced (Futterer et al., 2004) it was found to be AT-rich (64%) even though 92% of the genome was coding sequence. Our study found that the genome carried m6A methylation, but Dam-mediated methylation was not detectable. The investigation was taken forward with the examination of the modification methylase of the Type I R-M system annotated in its genome sequence, as all Type I R-M systems characterized to date demonstrate m6A methylation activity. The recombinant protein (M.PtoI) was found to mediate DNA adenine methylation under a wide range of conditions *in vitro*, and residues critical to methylation activity were identified.

2. Materials and methods

2.1. *Picrophilus torridus* cultures and genomic DNA isolation

Picrophilus torridus (strain DSM 9790) was cultured as described earlier (Arora et al., 2014). Briefly, cells were grown at 55°C in liquid culture with aeration, in medium of pH 1.0 comprising yeast extract (0.2%, Difco, United States), glucose (1%, Sigma, United States), potassium dihydrogen phosphate (0.3%), magnesium sulphate (0.05%), calcium chloride (0.025%) and ammonium sulphate (0.02%). For growth analysis, cultures were initiated with 1% (v/v) inoculum in media of different pH, from a starter culture at pH 1.0 that was at OD₆₀₀ ~ 1.5. Growth was monitored by measuring OD₆₀₀ every 24 h over a period of 14 days. Three experiments were set up in parallel and the average values of the three are plotted graphically, with error bars representing standard deviation. Genomic DNA was isolated from exponentially growing *Picrophilus torridus* cultures as described earlier (Arora et al., 2014). Briefly, harvested cells were resuspended in TEN buffer [20 mM Tris-Cl (pH 8.0), 1 mM EDTA, 100 mM NaCl], lysed by the addition of sodium lauroylsarcosine (1.6%) and Triton X-100 (0.12%), the lysate extracted with phenol, and genomic DNA precipitated using ethanol.

2.2. Cloning of M.PtoI genes

The ~1.3 kb gene encoding the S subunit was amplified off *Picrophilus torridus* genomic DNA using primers PtS-F (5'-TAAGATCTCATGAATAAAAAGGATTATAAT-3') and PtS-R (5'-TACTCGAGATCACCACTTATTTTCAC-3'), with the proof-reading enzyme Phusion DNA polymerase (NEB, United States). The amplicon was cloned into pUC19, followed by subcloning into the BglII-XhoI sites of the pETDuet-1 vector (Novagen, United States), thus obtaining the pET-Duet/PtS clone. The ~1.7 kb gene encoding the M subunit was similarly amplified using the primers PtM-F (5'-TAGAATTCGATGGCAGATTAAACTGG-3') and PtM-R (5'-TAGCGGCCGCTTC ATAGTAACCAAG-3'), the amplicon cloned into pUC19, and then subcloned into the EcoRI-NotI sites of pET-Duet and pET-Duet/PtS, thus obtaining plasmids pET-Duet/PtM and pET-Duet/PtSM, respectively.

2.3. Creation of PtM mutants

All mutageneses were carried out through PCR using Phusion DNA polymerase. For creating PtM mutants the plasmid pUC/PtM was used as template. The C-terminal deletion mutant was created by PCR using the PtM-F primer in combination with PtM_{Δ537-576}-R (5'-TAGCGGCCGCTTATTGTCTATATATAAAG-3'). The ~1.6 kb amplicon thus obtained was cloned into pJET vector followed by subcloning into the EcoRI-NotI sites of pETDuet/PtS, yielding the plasmid pETDuet/PtSM_{Δ537-576}. The NPPW and xxGxxG mutants were created by overlap PCR. To mutate the NPPW motif, the N-terminal part of the gene was amplified using primers PtM-F and PtM-N360A-W363A-R (5'-CTGGTTCGCTGGAGGCGCCG CAAC-3') while the C-terminal part of the gene was amplified using primers PtM-N360A-W363A-F (5'-GTTGCGGCGCCTCCAGCGAACCAG-3') and

PtM-R. The full length amplicon obtained by amplification with PtM-F and PtM-R primers was cloned into pJET vector followed by subcloning into the EcoRI-NotI sites of pETDuet/PtS, yielding the plasmid pETDuet/PtSM_{N360A-W363A}. To mutate the xxGxxG motif, the N-terminal part of the gene was amplified using primers PtM-F and PtM-G284S-R (5'-CCTGCAGTTGAACATGCCGG-3') while the C-terminal part of the gene was amplified using primers PtM-G284S-F (5'-CCGGCATGTTCAACTGCAGG-3') and PtM-R. The full length amplicon obtained using primers PtM-F and PtM-R was cloned into pJET vector followed by subcloning into the EcoRI-NotI sites of pETDuet/PtS, yielding the plasmid pETDuet/PtSM_{G284S}.

2.4. Creation of PtS mutant

M.PtoI/S_{211-222(A-G)} was created using Overlap PCR. The N-terminal fragment of the mutant amplicon was produced using primers PtS-F and PtS_{211-222(A-G)}-R (5'-GGCCCCGGCCCC CAGCTCCACCTGCCCCCT GTCCTGCCAACTGTATTTCTCT-3') and the C-terminal fragment of the mutant amplicon was produced using primers PtS_{211-222(A-G)}-F (5'-GCAGGAGCAGGGGCGAGGTGGAGCTGGGGCCGGG GCCTTCTGAAGGCATC-3') and PtS-R. The full-length mutated gene was amplified using PtS-F and PtS-R and ligated into the BglII-XhoI sites of pET-Duet vector to create plasmid pET-Duet/PtS_{211-222(A-G)}. This was followed by cloning the PtM gene into the EcoRI-NotI sites of pETDuet/PtS_{211-222(A-G)} to create plasmid pETDuet/PtS_{211-222(A-G)}M.

2.5. Expression and purification of recombinant M.PtoI

The M.PtoI enzymes were expressed in *E. coli* BL21 Codon Plus cells (Stratagene, United States) that had been transformed with their pET-Duet clones. Expression of the wild type or mutant proteins was induced in cells growing exponentially (OD₆₀₀ = 0.6–0.8) at 37°C using 1 mM IPTG. Cells were allowed to grow for a further 3 h at 37°C before harvesting by centrifugation and resuspension in lysis buffer (100 mM Tris.Cl pH8, 150 mM NaCl, 10% glycerol) containing protease inhibitors and lysozyme. This was followed by incubation of the cell suspension on ice for 20 min and lysis by sonication. After clarification of the lysates by high speed centrifugation they were subjected to cobalt-affinity chromatography using TALON metal affinity resin (Clontech). The lysates were loaded on the column resin, the column extensively washed (with 100 mM Tris.Cl (pH 8), 1 M NaCl, 10% glycerol), and the bound proteins eluted using 100 mM Tris.Cl (pH 8), 250 mM imidazole, 10% glycerol. The M subunit of M.PtoI was similarly purified from bacteria transformed with pET-Duet/PtM.

2.6. *Picrophilus torridus* genomic DNA methylation analysis

Dot blot assay was used to analyze genomic DNA that was isolated from *P. torridus* cultures as described above. For this, 1 µg genomic DNA was digested with EcoRI and then denatured with 0.3 N sodium hydroxide for 30 min. This was followed by neutralization with 1 M

ammonium acetate before spotting on Hybond N⁺ nylon membrane (GE Healthcare, United States) using the Bio-Dot apparatus (Bio-Rad Laboratories, United States). The membrane was baked at 80°C for 2 h before probing with anti-m6A antibody (Sigma Aldrich, Cat. No. ABE572) or anti-m5C antibody (EpiGentek, Cat.No. A-1014-010). This was done by blocking the baked dot blots with 10% skim milk (Difco Laboratories) for an hour at room temperature, washing the blots twice with 1X PBS-T, and then incubating them overnight at 4°C with the anti-m6A or anti-m5C antibodies (1:1000 dil in 1X PBS). Following three washes with 1X PBS-T the blots were incubated with HRP-labelled secondary antibody (1:10000 dil, Jackson Laboratory, United States) for an hour and developed using a chemiluminescence method.

2.7. DNA methylation assay

DNA methylation assays were performed using a dot blot-based assay with lambda DNA (as the substrate) that had been isolated from phage particles hosted by an *E.coli dam⁻ dcm⁻* strain (DNA purchased from Thermo, United States; Cat no: SD0021). Reactions (10 µl) were typically carried out in a Tris-acetate based buffer [30 mM Tris-acetate (pH 5.0), 10 mM potassium acetate, 10 mM magnesium acetate, 100 µg/ml BSA] containing 100 µM S-adenosyl methionine (AdoMet; purchased from NEB, United States), 12 mM 2-mercaptoethanol, 20 mM sodium chloride, 1 µg DNA and 1.4 µM M.PtoI, at 55°C for 1 h. Reactions were stopped with 100 mM EDTA. Reactions were analyzed by dot blot assay as described above. The blots were quantified using ImageJ analysis, and the ratios of m6A with reference to *E.coli* genomic DNA spotted on the same blot was determined. Each experiment was done thrice. The graphs depict average values of three experiments with error bars representing standard deviation.

2.8. CD spectroscopy analysis

CD spectroscopy was carried out as earlier (Arora et al., 2014). Briefly, proteins (100 µg in 10 mM potassium phosphate buffer (pH 5.8)) were analyzed with a Jasco J-815 spectropolarimeter using cuvette of 0.1 cm path length. CD spectra were recorded over 250–190 nm at room temperature, in 1 nm steps at a scan speed of 200 nm/min. The presented spectra are the average of 20 scans. The data is depicted as the mean residue ellipticity (MRE).

2.9. Homology modelling and structural analysis

The structures of the M and S subunits of M.PtoI were predicted using the Phyre2 (protein homology/analogy recognition engine version 2.0) online server (Kelley et al., 2015).¹ The reliability of the predicted models were evaluated using the QMean and LG scores

obtained (Benkert et al., 2009).² In both cases the models that were considered demonstrated high coverage against their respective templates. All structures were visualized and illustrations created using PyMOL³ (Mooers, 2016).

3. Results

3.1. *Picrophilus torridus* genome carries methylated adenine residues

The study was initiated with examining the growth pattern of the organism in media of different pH. The data obtained revealed that while *P. torridus* grew comparably at pH 0.7 and 1.0, growth was considerably slower at pH 2.0 and no growth was discernible at pH 3–5 (Figure 1A). To determine if DNA methylation exists in this microorganism under the extreme conditions in which it thrives, genomic DNA was isolated from *P. torridus* cells grown in media whose pH ranged over 0.7 to 2 and the DNA analyzed by dot blot assay as described above. The results obtained clearly demonstrated that the *Picrophilus torridus* genome carries methylated adenine residues (m6A) regardless of the pH at which it is growing, although displaying lower levels of methylation at pH 0.7 (Figure 1B). No m5C methylation was detectable (Figure 1C), and we were unable to assess m4C methylation due to lack of quality antibodies.

A previous study by Koike et al. (2005) had established the presence of the *dam* gene in several archaea including *P. torridus* by computational analysis, but Dam-mediated adenine methylation was not experimentally verified in *P. torridus*. To check the methylation status of the *P. torridus* genome at Dam target sites (5'-GATC-3') we exploited the property of the DpnI restriction enzyme that cleaves methylated but not unmethylated GATC sites. Genomic DNA isolated from *Picrophilus* cells grown in medium of pH 1.0 was thus subjected to DpnI digestion, and it was observed that while *E.coli* genomic DNA was cleaved by the enzyme the *P. torridus* genomic DNA was not (Figure 1D), nor was *S.pombe* genomic DNA (which lacks m6A methylation as evident from Figure 1B). These data indicate that while the *P. torridus* genome carries the m6A mark the organism does not harbor Dam-mediated adenine methylation.

An examination of the *P. torridus* genome sequence (KEGG database)⁴ revealed that in addition to the *dam* gene, the euryarchaeon has genes encoding a Type II R-M system and a Type I R-M system. All Type I modification methylases characterized to date are m6A methylases, and thus we investigated the modification methylase of the *P. torridus* Type I R-M system. The three genes of the *P. torridus* Type I R-M system encoding for R, M and S subunits lie in a single operon on the lower strand of the circular genome (Figure 1E). The investigation of the modification methylase (henceforth referred to as M.PtoI) commenced with cloning the genes encoding M and S subunits.

¹ <http://www.sbg.bio.ic.ac.uk/phyre2>

² <https://proq.bioinfo.se/ProQ/ProQ.html>

³ <http://pymol.org>

⁴ www.genome.jp/kegg

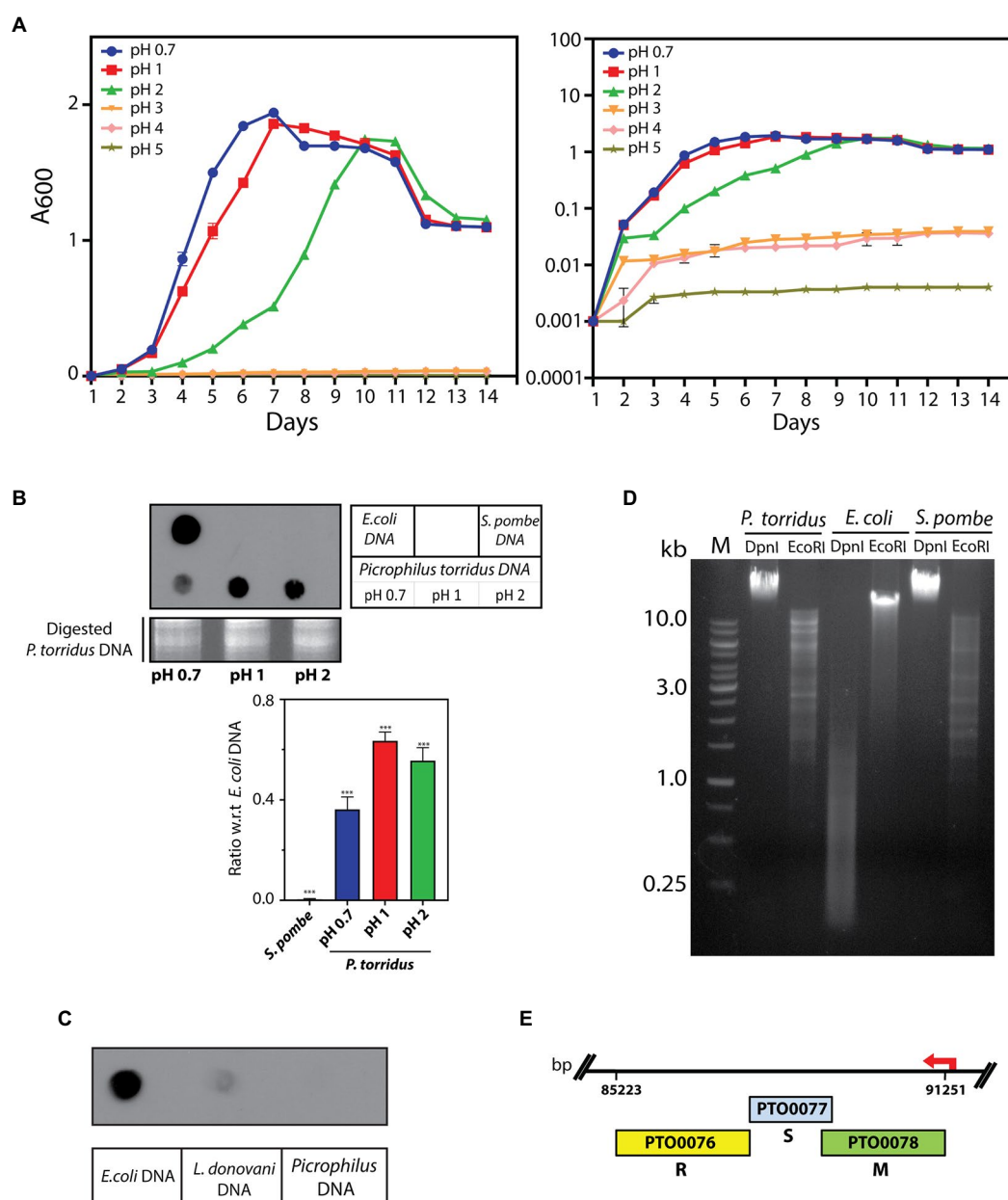


FIGURE 1

(A) Analysis of growth of *P. torridus* at different pH. Cultures were initiated with 1% (v/v) inoculum from a starter culture at pH 1.0 that was at OD₆₀₀~1.5. Growth was monitored every 24h. Graph represents average values of three experiments and error bars represent standard deviation. Left panel: data plotted against linear y-axis. Right panel: data plotted against log y-axis. (B) Dot blot analysis of *P. torridus* genomic DNA for m6A mark. 1μg of genomic DNA was used in each case. Upper left panel: blot probed with anti-m6A antibody (1:1000 dil). Loading control: agarose gel analysis of input genomic DNA digested with EcoRI. Upper right panel: loading scheme of dot blot. Lower panel: quantification of *P. torridus* genomic DNA signal as a ratio of *E. coli* genomic DNA signal, done using ImageJ software. Bars represent average of three experiments. Error bars represent standard deviation. Student's two-tailed *t*-test was applied, ****p*<0.0005. (C) Dot blot analysis of *P. torridus* genomic DNA for m5C mark. 1μg of genomic DNA was used in each case. Upper panel: blot probed with anti-m5C antibody (1:1000 dil). Lower panel: loading scheme of dot blot. (D) Analysis of 5'-GA^{m6}TC in *P. torridus* genomic DNA. In each case, 1μg of genomic DNA was treated with restriction enzyme. M: molecular weight marker. (E) Line diagram of *P. torridus* Type I RM system: PTO0076: gene for R subunit, PTO0077: gene for S subunit, PTO0078: gene for M subunit.

3.2. M.Ptol carries all the motifs that typify Type I modification methylases

The M.PtoI genes (designated herewith as PtM and PtS for M and S subunit genes respectively) were cloned by amplification using primers designed against the published PTO0078 and PTO0077

sequences (encoding M and S subunits respectively), as described in Materials and Methods. The pUC clones obtained were sequenced for verification. The sequences of the M and S subunits of M.PtoI were compared with those of typical Type I modification methylases EcoKI and EcoR124I as well as the identified archaeal Hvo MTase (HVO_2270–71), using Clustal Omega and BLASTp analyses (Sievers

et al., 2011).⁵ While the M subunit shared ~25–28% sequence identity over a coverage of 55–85% with the M subunits of the other enzymes (Figure 2), the S subunit shared between 23–25% identity over a coverage of 60–93% with their corresponding S subunits (Supplementary Figure S1). A comparison of the sequence of M subunit with the sequences of modification methylases of typical Type II systems using the same tools revealed that although all of them harbor the characteristic motifs discussed below (Supplementary Figure S2), overall there was no significant sequence identity (data not shown), reflecting the differences in the structure and subunit composition of Type I and Type II modification methylases.

m6A and m4C methylases are typified by the presence of nine conserved motifs, identified more than 25 years ago by analyzing the sequences of over 40 DNA MTases (Malone et al., 1995). By examining these motifs in the context of published structures of MTases it is now known that Motifs I–III and X are involved in the binding of the methyl group donor S-adenosyl methionine (AdoMet). It has been suggested that Motif I and motif X form a pocket that lodges the methionine group of AdoMet while motifs II and III are involved in interactions with the ribose and adenine moieties of the same. On the basis of structural as well as biochemical analyses it is also understood that Motifs IV–VIII are part of the active site pocket and are involved in catalysis of the methyl-transfer reaction, while the Target Recognition Domain (TRD) is responsible for recognition of the specific DNA (Malone et al., 1995; Bheemanaik et al., 2006). In Type I enzymes where DNA recognition is not a part of the functions of the M subunit the TRD lies in the S subunit. The recognition sequence of Type I enzymes is bipartite, and accordingly, the S subunits harbor two TRD domains. The MTases have been classified into six categories: α , β , γ , δ , ϵ , and ζ based on the order of these signature motifs in their primary sequence (Bujnicki, 2002). The majority of the m6A and m4C methylases fall in the α , β , and γ categories. All the Type I modification methylases as well as several Type II methylases are γ MTases, whose order of the motifs is X-I-II-III-IV-V-VI-VII-VIII, and M.PtoI conforms to this as seen in Figure 2. No clearcut motif II and motif III were identifiable in M.PtoI.

3.3. The structures of the M and S subunits of M.PtoI are broadly conserved with those of other Type I modification methylases

While the structures of the individual subunits of Type I R-M systems have been solved in case of several enzymes, the holoenzyme (M2S1) structures of only two Type I modification methylases are available to date. The EcoKI modification methylase structures (pdb IDs: 2Y7C and 2Y7H) were proposed based on single particle electron microscopy constructions at 18 Å resolution (Kennaway et al., 2009). The more recent EcoR124I modification methylase and restriction enzyme structures were proposed based on single particle cryo-EM at 4.54 Å resolution (Gao et al., 2020). The basic architecture of the M and S subunits of the two MTases and the interfaces of interaction among the different subunits are largely conserved between the two enzymes. The structures of the M and S subunits of M.PtoI were

predicted by modelling using Phyre2. The modelled structures were viewed using PyMOL, which allowed us to superimpose the obtained structures on their respective templates to analyze structural alignments.

The most reliable model obtained for the M subunit used the EcoR124I enzyme as template (pdb ID: 7BTQ; Gao et al., 2020). The predicted structure of the M subunit displayed 27% identity with the structure of the M subunits in the 7BTQ cryo-EM structure (4.54 Å), over a coverage of ~83% (residues 89–153 and 164–569). As seen in Figure 3, seven motifs that were aligned in the primary sequence (Figure 2) were also aligned structurally. The M subunits of Type I RM systems are typified by a C-terminal helix that is involved in interactions with the S subunit, and the M subunit of M.PtoI too harbored this conserved helical C-terminal domain (Figure 3).

The most reliable structure obtained for the S subunit used the S subunit of the *Methanocaldococcus jannaschii* Type I R-M system (pdb ID: 1YF2; Kim et al., 2005) as template. The S subunit of M.PtoI showed 30% identity with the 1YF2 X-ray crystal structure of the *M. jannaschii* S subunit (2.4 Å), over ~97% coverage (amino acids 6–106 and 121–437). The S subunit of Type I systems harbors two TRDs (target recognition domains) that are oriented inverted with respect to each other. Each TRD carries a globular domain and an alpha-helical dimerization domain (Loenen et al., 2014a). The globular domains (TRDI and TRDII in Figure 4) are responsible for DNA binding, with the N-terminal one binding to the 5' segment (specific sequence) of the bipartite recognition site and the C-terminal one binding to the 3' segment (specific sequence) of the same. The alpha-helical domains of the two TRDs (named central conserved region or CCR, and distal conserved region or DCR; see Figure 4) associate with each other such that the two globular domains are held apart at a fixed distance, thus fixing the length of the non-specific spacer sequence in the bipartite recognition site. The architecture of the S subunit of M.PtoI was predicted to be similar, with two TRDs each harboring a globular domain and an alpha helical domain (Figure 4).

The S and M subunits of Type I enzymes characteristically interact with each other (either in the M2S1 or in the R2M2S1 complex) through a four-helix bundle that comprises the C-terminal helical domains of the two M subunits (depicted as CTD in Figure 3) and the two alpha-helical domains of the S subunit (depicted as CCR and DCR in Figure 4). As these regions are conserved in the M and S subunits of M.PtoI, the structure of the M.PtoI holoenzyme (M2S1) was predicted using template-based docking, *via* superimposition of the predicted 3D structures of M and S of M.PtoI (Figures 3, 4) on the M.EcoR124I holoenzyme structure (pdb:7BTQ). The M.PtoI docked structure obtained predicted the two M subunits of the M2S1 holoenzyme to interact with each other *via* their N-terminal domains, and the M and S subunits to interact with each other *via* the four-helix bundle (Figure 5).

The role of the C-terminal helical domain of M in mediating interactions with the S subunit was experimentally validated by creating a C-terminal deletion mutant of M: M.PtoI/M $_{\Delta 537-576}$, and analyzing it for loss of interactions with S. For this, the recombinant M.PtoI proteins were purified from *E.coli*. Recombinant M.PtoI (full length) was expressed in *E.coli* using the pET-Duet expression vector which allows the simultaneous expression of two proteins from two separate T7lac promoters, with one of the two proteins being expressed in fusion with the His tag at the N-terminus. The M and S subunits were thus simultaneously expressed in BL21 Codon Plus

⁵ www.blast.ncbi.nlm.nih.gov

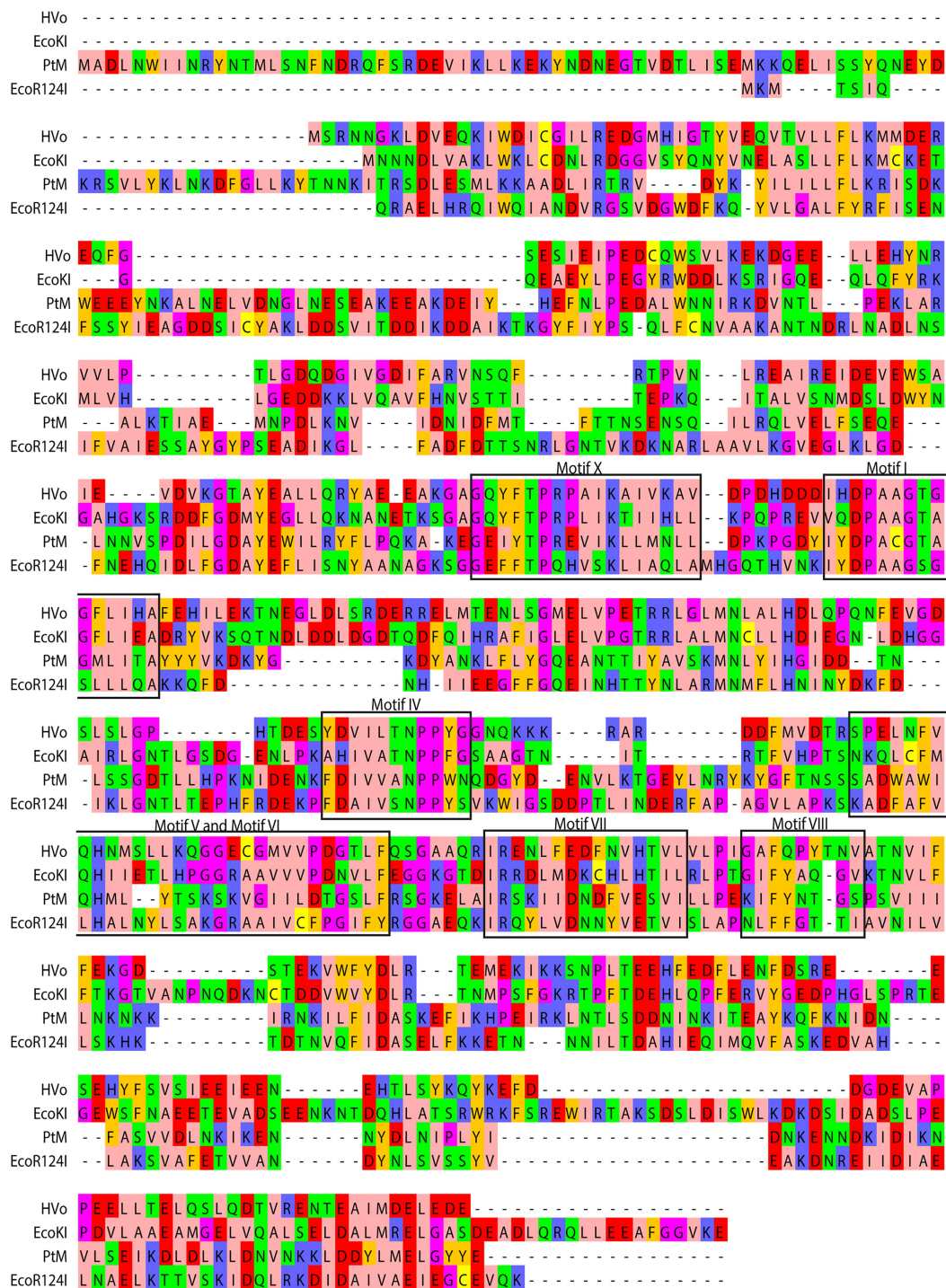


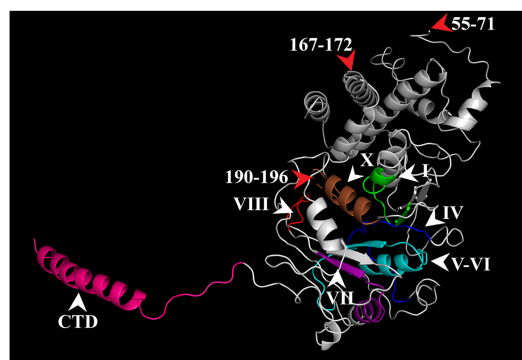
FIGURE 2

Comparison of derived amino acid sequence of M subunit of M.PtoI with the sequences of M subunits of other Type I modification methylases: Clustal Omega analysis viewed using Jalview multiple alignment editor (Sievers et al., 2011). The conserved motifs are demarcated in black rectangular boxes. Colors are indicative of the physicochemical properties of the amino acids. Pink, aliphatic/hydrophobic; orange/ochre, aromatic; purple, glycine/proline; dark blue, basic; green, hydrophilic; red, acidic; yellow, cysteine.

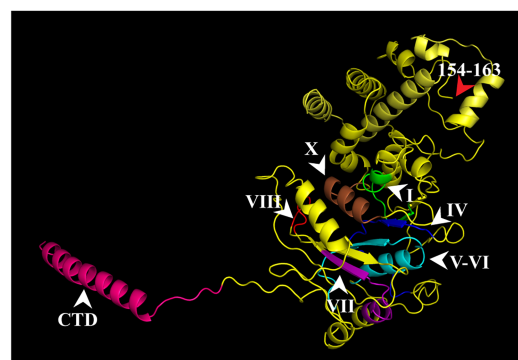
E. coli cells from the pETDuet/PtSM clone (as described in Methods) such that M was His-tagged. The cell lysates were subjected to cobalt affinity-based chromatography. SDS-PAGE analysis revealed the eluate to carry both subunits in stoichiometric amounts, leading us to conclude that the M.PtoI holoenzyme (M2S1) assembles in the host

bacterium, allowing the S subunit to co-purify with the His-tagged M (Figure 6A). In creating the C-terminal deletion of M, amino acids 537–576 were excluded (mutagenesis detailed in Methods) and the M_{Δ537-576} co-expressed with S in BL21 Codon Plus cells from the pETDuet/PtSM_{Δ537-576} clone. When these cell lysates were put through

PtM	89	SDLESMLKKAADLI	RTRVDY	----	KYILILFLKRI	SDKWEEEYNKALNELVDNGLNESEAKKEEAKDEIYHEFNL
7BTQ	10	AELHRQIWQIANDV	RGSVDGWDFKQYV	LGALFYRF	ISENFSSYIEAGTDDIKDDAIKTKGYFIYPSQLFCN	----
PtM	160	PEDALWNNIRKDVNTL	LPEKLARALKTIAE	-----	MNPDLKNVIDNIDFMTFTTNS	EQILRQLVELFSEQELN
7BTQ	95	-----	VAAKANTNDRLNAD	LNSIFVAIESSAYGYPSEADIKGLFADF	TTSNRLGNTVKDKNARLA	AVLKGV
Motif X						
PtM	229	NVSPDIL	GDAYEWILRYFLPQAK	EGE	IYTPREVIKLLMNL	----
7BTQ	165	KLGI	DLFGDAYEFLISNY	-----	GGEFFTPQHVSKLIAQLAMHGQTHV	NKI
Motif I						
PtM	302	KDYANKLFLY	GQEANTTIYAVSKM	NLYIHGIDDTN	--LSSGDTLLHPKNIDENK	FDIVVANPPWNQDGYDENVLK
7BTQ	244	---	IEEGFFGQE	INHTTYNLARMNMF	LHNINYDKFDIKL	NTLTPHFRDEKPFDAIVSNPPYSVKWIGSDD--P
Motif IV						
Motif V and Motif VI						
PtM	375	TGEYLNRYKYG--	FTNSSADAWIQHML	--YTSKSKVGI	ILDTGSLFRSGKEL	IRSKIIDNDFVESVILLPEK
7BTQ	316	TLINDERFAPAGVLAPKS	KADFAFVLHALNYLSAK	GRAAIVCFPGIFY	RGAEQKIRQYLV	DNVYETVISLAPN
Motif VII						
Motif VIII						
PtM	446	IF	YNTGSPSVIIILNKNK	KIRNKILFIDASKEFI	KHPEIRKLNTLSDDNINK	ITEAYKQFKNIDNFA
7BTQ	391	LF	FGTTIAVNILVLSKHKT	DTN-VQFIDASELFK	---ETNNIL	DAHIEQIMQVFASKEDVAHLAKSVAFETV
CTD						
PtM	521	KEN	NYDLNIPLYIDN	KENNDKIDIKNVLSEIKDL	DLKLDNVNKKLDDYL	
7BTQ	462	VAN	DYNSVSSVVEAK	DNREIIDIAELNAELKTTVSK	IDQLRKDIDATV	



M subunit of EcoR124I (7BTQ)



M subunit of M.PtoI

FIGURE 3

Structural analysis of M subunit of M.PtoI: Upper panel: Alignment of sequence of M subunit of M.PtoI with M subunit of EcoR124I, using Phyre2. The boxes demarcate the positions of the motifs conserved between the two subunits. CTD: C-terminal domain. Lower left panel: Ribbon representation derived from cryo-EM structure (4.54Å) of EcoR124I [pdb ID: 7BTQ reported by Gao et al. (2020)]. Lower right panel: Ribbon representation of 3D model of M subunit of M.PtoI, modelled against the M subunit of EcoR124I (7BTQ) as template using Phyre2. CTD: C-terminal domain. Roman numerals indicate motif positions. Arabic numerals depict amino acid residues that were excluded from the models.

TALON beads the eluate carried only the truncated M subunit, clearly demonstrating loss of interaction with S in M.PtoI/M_{Δ537-576} (Figure 6B, left panel, fourth lane). The deletion did not cause any gross structural changes in the M subunit based on a comparison of the CD spectroscopy profiles of purified M (Figure 6B, right panel, fourth lane) and M.PtoI/M_{Δ537-576} (Figure 6C). Taken together, these results indicate that the C-terminal helical domain of M is functionally conserved in M.PtoI.

The role of the central helical region of the S subunit in mediating M-S interactions was examined by creating a mutant S subunit where 12 amino acids of the central conserved region (CCR) were replaced by a 12 amino acid stretch of alternate glycine and alanine residues to create M.PtoI/S_{211-222(A-G)} (described in “Methods”). When the recombinant M.PtoI/S_{211-222(A-G)} protein was purified from *E. coli* using

cobalt affinity chromatography it was found that there was a substantial increase in the ratio of M:S in the eluate fractions (Figure 6D, fourth lane), implying that the replaced region played a significant role in mediating M-S interactions. The difference in CD spectroscopy profiles of the wild type and mutant M.PtoI proteins most likely reflects the differences in M:S stoichiometry in the two proteins (Figure 6E).

M-S interactions in wild type and mutant M.PtoI proteins were also checked using purified proteins. For this, M and S subunits were separately expressed in *E. coli* cells from respective pET/Duet clones carrying either M or S genes. As all M subunits were tagged with His at their N-terminus, they were immobilized on Co-affinity resin. Whole cell lysates harboring recombinant S subunit (wild type or mutant) were added to the reaction, and after washing off the unbound

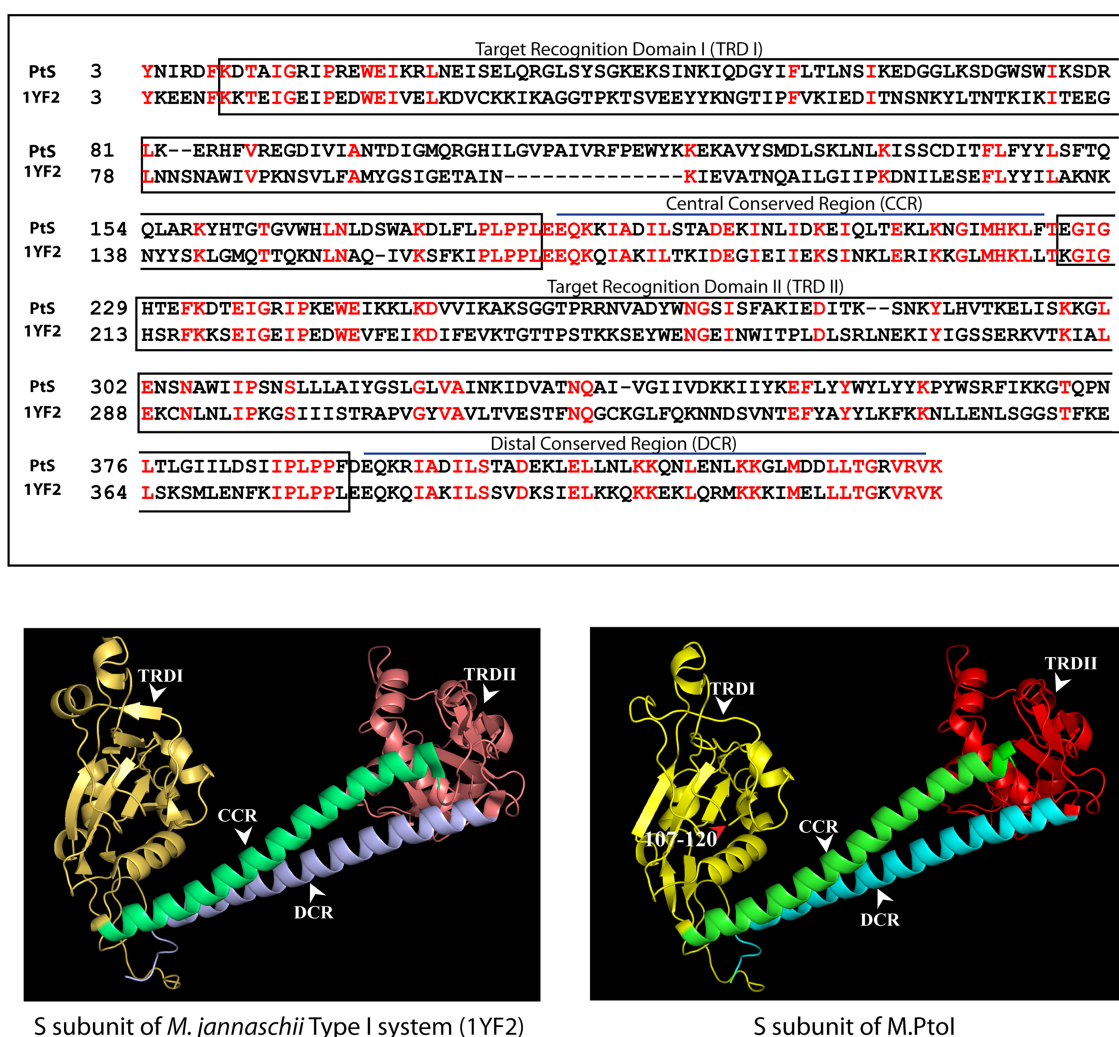


FIGURE 4

Structural analysis of S subunit of M.PtoI: Upper panel: Alignment of sequence of S subunit of M.PtoI with S subunit of *Methanocaldococcus jannaschii* Type I R-M system, using Phyre2. The boxes demarcate the positions of the domains conserved between the two subunits. Lower left panel: Ribbon representation derived from X-ray crystallography structure (2.4Å) of *M. jannaschii* Type I RM system's S subunit [pdb ID: 1YF2, reported by Kim et al. (2005)]. Lower right panel: Ribbon representation of 3D model of S subunit of M.PtoI, modelled against the 1YF2 S subunit structure using Phyre2. Arabic numerals depict amino acid residues that were excluded from the models. TRDI and TRDII: Target Recognition Domain I and Target Recognition Domain II representing the globular domains of the N-terminal and C-terminal TRDs, respectively. CCR: Central Conserved Region, DCR: Distal Conserved Region.

fractions, the bound fractions (M subunits plus any interacting S subunits) were eluted using imidazole and analyzed. As seen in [Supplementary Figure S3](#), while S subunit interacts with wild type M subunit (lane 5), it fails to interact with mutant M_{Δ537-576} subunit (lane 6). The mutant S_{211-222(A-G)} subunit interacts with M subunit to a significantly lesser extent than wild type S (compare lane 7 with lane 5). These data confirm the role of the C-terminus of M subunit and residues 211–222 of S subunit in maintaining M-S interactions.

3.4. M.PtoI methylates DNA at adenine residues *in vitro*

The ability of the purified M.PtoI to methylate DNA at adenine residues was analyzed using unmethylated lambda DNA isolated from phage that had been grown in a *dam⁻dcm⁻* *E. coli*

strain. The reactions were analyzed in dot blot assays using anti-m6A antibody, as described in Methods. Activity assays were initiated at 55°C for 1 h in Tris-acetate buffers whose pH ranged over 3 to 7, in the presence of sodium chloride ranging over 0–100 mM. As seen in [Figure 7A](#) M.PtoI was able to methylate lambda DNA (producing m6A) at pH ranging from 4 to 7, with optimum activity being detected at pH 5, which is near the organism's intracellular pH of 4.6. Higher salt concentrations were detrimental to activity, with optimal activity seen at 0–20 mM NaCl. Thus, all further assays were carried out in Tris-acetate buffer (pH 5) containing 20 mM NaCl. The optimal concentration of magnesium ions that supported the DNA methylation reaction was determined over 0–50 mM magnesium acetate. It was found that while magnesium was necessary for the methylation reaction, concentrations of magnesium over 5–25 mM supported the reaction more or less equivalently, with

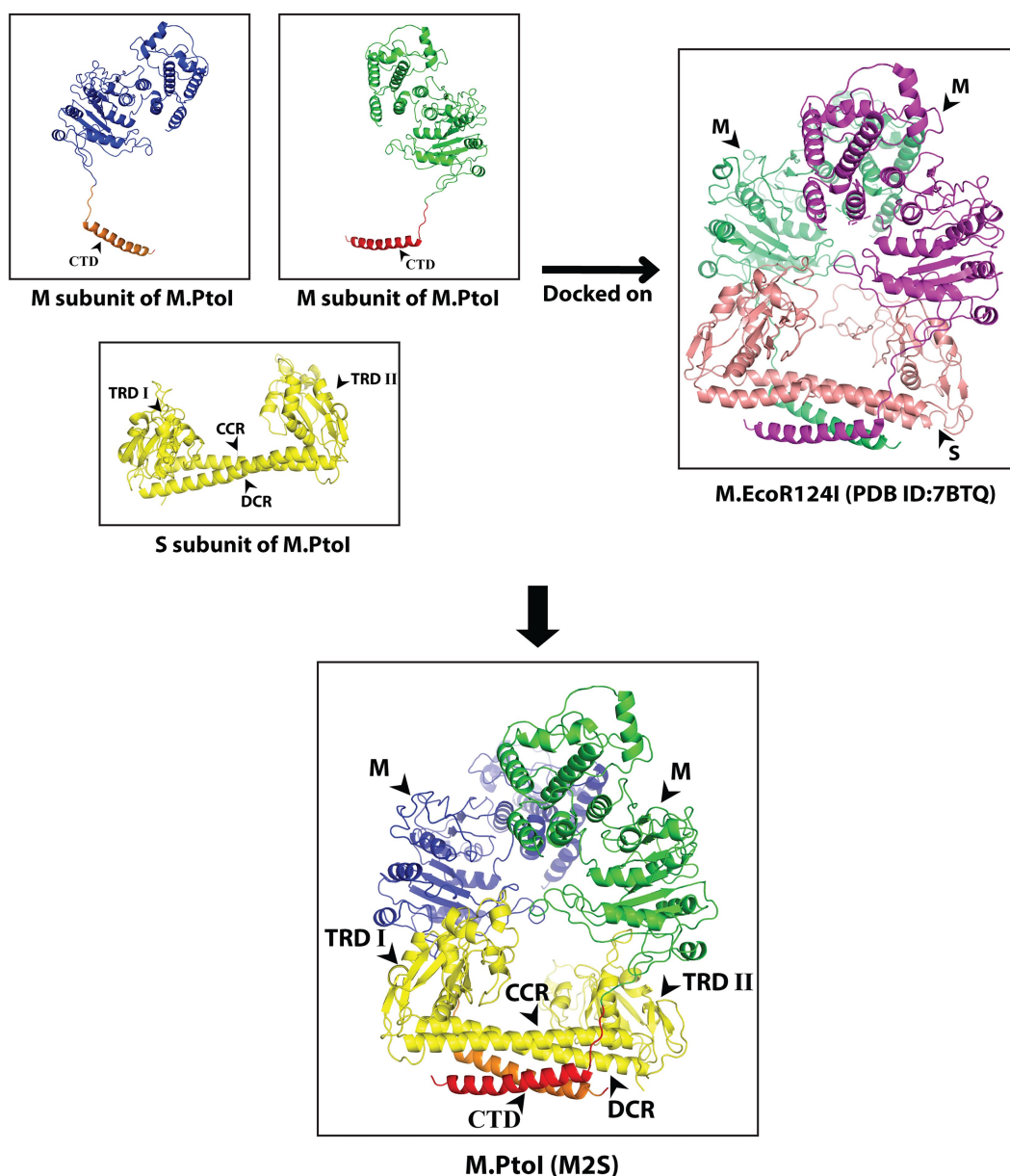


FIGURE 5

Predicted structure of M.PtoI: Predicted 3D structures of M and S subunits of M.PtoI were superimposed on the 7BTQ M.EcoR124I holoenzyme structure. M.PtoI structure was thus obtained using template-based docking. M: M subunit, CTD: C-terminal domain, TRDI and TRDII: Target Recognition Domain I and II respectively, CCR: Central Conserved Region, DCR: Distal Conserved Region.

a higher concentration of 50 mM being inhibitory (Figure 7B). The methylation reaction displayed an initial lag before proceeding almost linearly for approximately the first 45 min and plateaued beyond 60 min (Figure 7C). The ability of M.PtoI to methylate DNA was concentration-dependent, increasing almost linearly between 200 and 600 nM before plateauing (Figure 7D). High concentrations of enzyme (2,000 nM and beyond) proved to be detrimental, perhaps due to protein aggregation occurring at such high concentrations. In carrying out the assays at different temperatures ranging over 37–70°C it was found that M.PtoI was active at temperatures between 37–65°C with maximal activity at 55 and 60°C, in keeping with the conditions in which it thrives (Figure 7E). AdoMet supported the methylation reaction over

concentrations varying from 10–500 μM, with 50–100 μM being optimal (Figure 7F). Interestingly, the enzyme displayed substrate inhibition at higher AdoMet concentration of 1 mM. Furthermore, the enzyme was able to methylate DNA, although to a much lesser extent, even in the absence of exogenously added AdoMet, suggesting that endogenous AdoMet remains bound to the M subunit through the purification process. This is not unusual for such enzymes, with similar findings in multiple enzymes such as in BpmI, BseMII, and EcoR124II (Dreier et al., 1996; Jurenaite-Urbanaviciene et al., 2001; Bath et al., 2002). Unsurprisingly, the addition of ATP had no effect on methylation activity (data not shown). Taken together, the results presented in Figure 7 establish M.PtoI as an active m6A MTase.

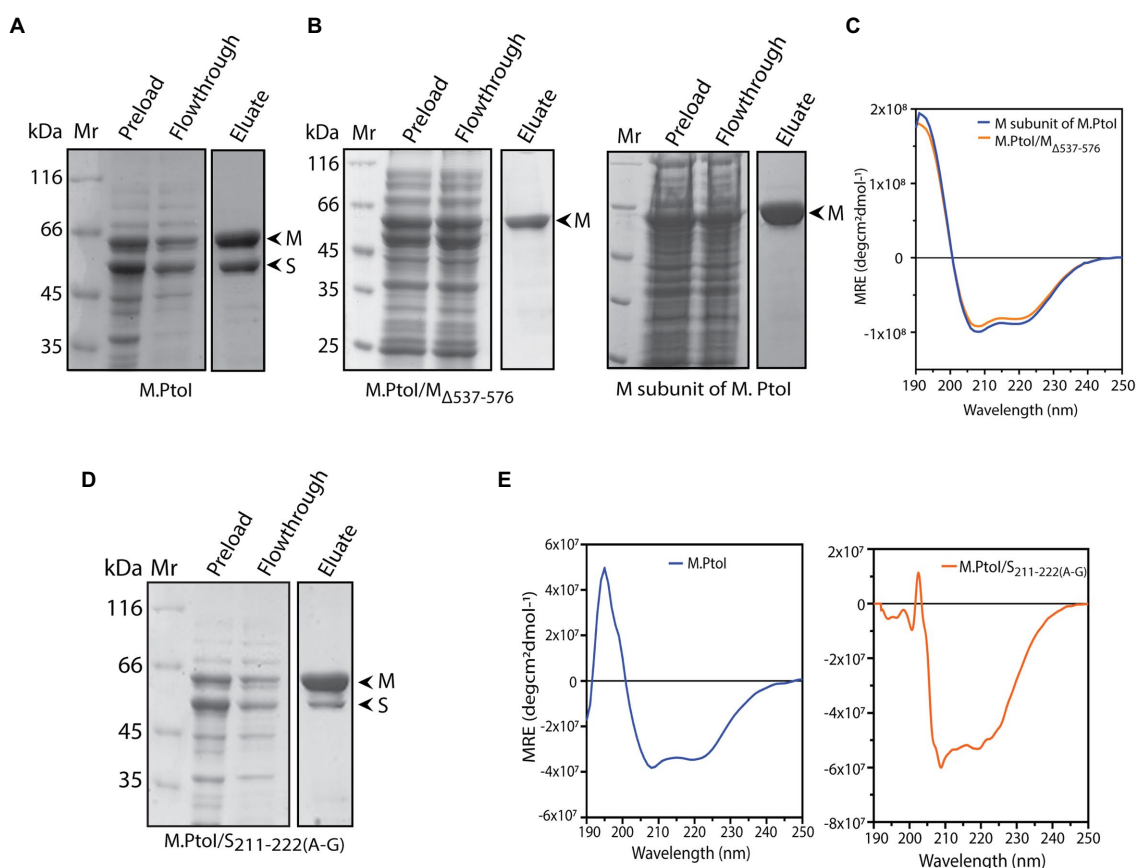


FIGURE 6

(A,B,D) Analysis of purified recombinant M.PtoI proteins. Coomassie-staining of SDS-PAGE. Mr.: molecular weight marker, preload: bacterial whole cell lysates after induction of expression of recombinant proteins, flowthrough: unbound protein fraction after incubation with cobalt affinity beads, eluate: protein fraction after elution with 250mM imidazole. Arrowheads indicate subunits in eluate fractions. (A) Wild type. Eluate: M and S subunits (molecular weights 63kDa and 48kDa respectively). (B) Left panel: M.PtoI/M Δ 537-576. Eluate: M Δ 537-576 subunit only (molecular weights ~58kDa). Right panel: M subunit of M.PtoI. Eluate: M subunit (molecular weight 63kDa). (C) CD spectra of M.PtoI M subunit and M.PtoI/M Δ 537-576 depicted as a measure of mean residue ellipticity (MRE). (D) M.PtoI/S₂₁₁₋₂₂₂(A-G). Eluate: M and S subunits (molecular weights 63kDa and 48kDa respectively). (E) CD spectra of M.PtoI (left panel) and M.PtoI/S₂₁₁₋₂₂₂(A-G) (right panel) depicted as a measure of mean residue ellipticity (MRE).

3.5. The NPPW and xxGxxG motifs in M.PtoI are critical for methylation activity

Type I modification methylases, like most adenine MTases, are typified by two primary domains they harbor in their M subunits: the AdoMet binding domain (motifs X, I-III in Figures 2, 3), and the catalytic domain that is responsible for the transfer of the methyl group from the donor to the nucleotide which is to be methylated (motifs IV-VIII in Figures 2, 3). Two motifs that are of singular importance are the AdoMet-binding motif I (FxGxxG) and the catalytic motif IV (NPPY/F/W). The catalytic motif IV lies in a pocket on the surface of the M subunit and at the time of methylation the specific adenine residue flips out of the duplex DNA substrate into the pocket where it stacks with the aromatic residue as well as forms hydrogen bonds with the first amino acid of the motif (which could be D, S, or N), stabilizing it for receiving the methyl group which is directly transferred to the target amino group on the adenine (Bheemanaik et al., 2006). Based on the sequence alignment with other Type I MTases (Figure 2), the FxGxxG motif in M is located at residues 282–287 and the NPPW motif lies at residues 360–363 (Figure 8A). The importance of these two motifs for the

methyltransferase activity of M.PtoI was assessed by developing mutant enzymes and analyzing their activity. Accordingly, the glycine residue at position 284 was mutated to serine, and the asparagine and tryptophan residues at positions 360 and 363 were mutated to alanine (described in Methods). The mutated M subunits were co-expressed with S subunit in *E.coli* cells from pETDuet/PtSM_{G284S} and pETDuet/PtSM_{N360A-W363A}, respectively. The mutant enzymes M.PtoI/M_{G284S} and M.PtoI/M_{N360A-W363A} were purified (Figure 8B) and assessed for any gross structural changes using CD spectroscopy; no gross structural changes were detectable (Figure 8C).

The two mutant enzymes were assessed for their ability to methylate DNA in Tris-acetate buffer at pH 5 over sodium chloride concentrations ranging from 0–50mM. M.PtoI/M_{N360A-W363A} was found to be virtually inactive under all tested conditions, while M.PtoI/M_{G284S} was less active than wild type enzyme at all tested conditions (Figure 8D). The activity of M.PtoI/M_{G284S} was compared with that of the wild type M.PtoI at different concentrations of AdoMet. It was observed that while the wild type enzyme supported DNA methylation more or less equivalently over 50–500μM AdoMet, the mutant enzyme while displaying an overall lower activity showed maximal support of DNA methylation at 100–200μM AdoMet (Figure 8E).

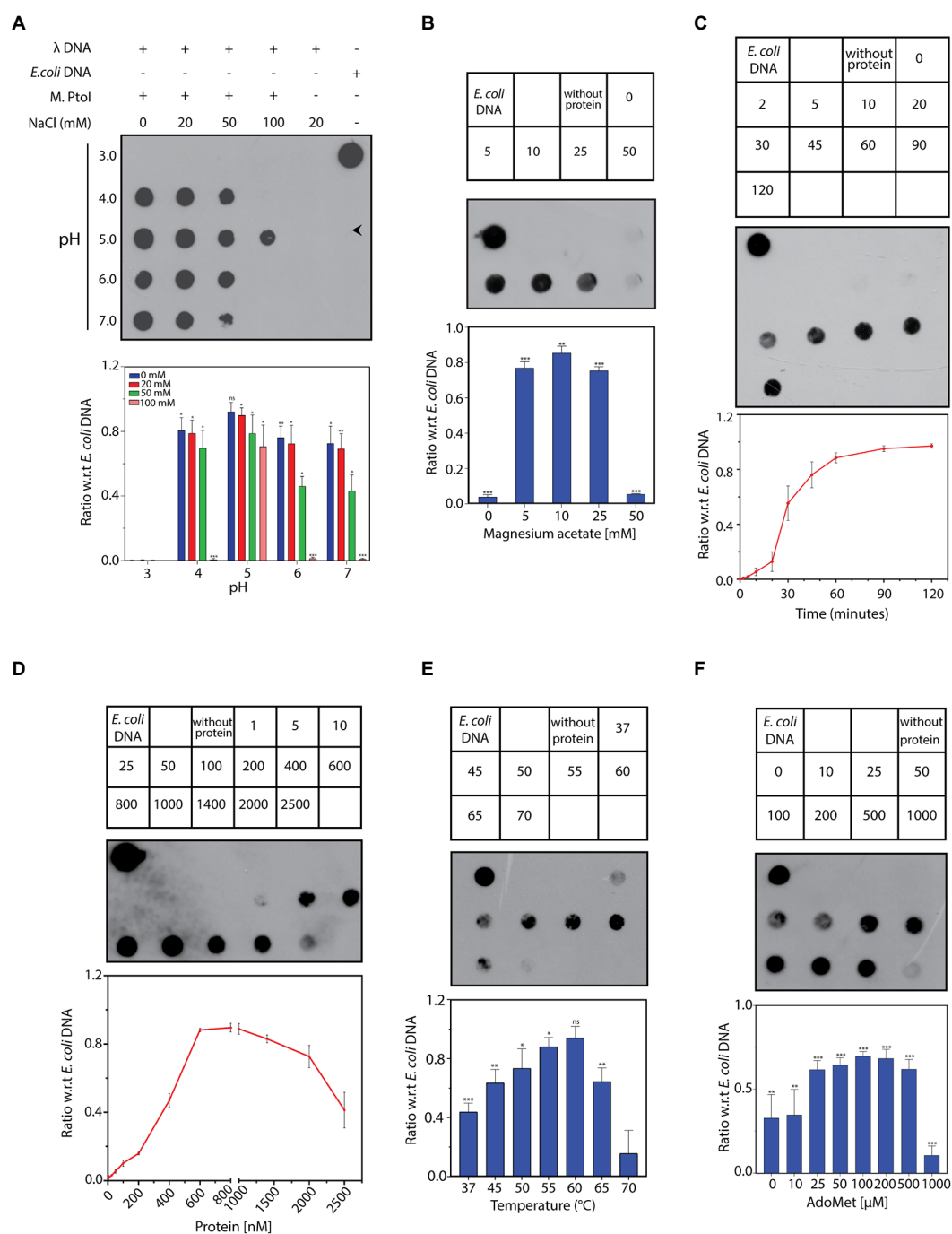


FIGURE 7
(A–F) Dot blot analyses of methylation assays performed with M.Ptol. Assays were performed with 1 μg lambda DNA substrate, in Tris-acetate based buffer (pH 5) carrying 10 mM magnesium acetate, 100 μM S-adenosyl methionine (AdoMet), 20 mM sodium chloride, and 1.4 μM M.Ptol, at 55°C, unless otherwise specified. 1 μg *E. coli* genomic DNA was used as control on every dot blot. Upper panels in each figure: loading scheme of dot blot. Middle panel: Dot blot of assay. Lower panel: quantification of signals from methylation reactions as a ratio of *E. coli* genomic DNA signal, done using ImageJ software. Data represent average of three experiments. Error bars represent standard deviation. Student's two-tailed *t*-test was applied, **p* < 0.05, ***p* < 0.005, ****p* < 0.0005, ns-not significant. (A) Reaction as a function of pH and salt concentration: Activity was assayed over pH 3–7 in presence of 0–100 mM sodium chloride, using 1.4 μM M.Ptol at 55°C for 1 h. (B) Reaction as a function of magnesium ion concentration: Activity was assayed over 0–50 mM magnesium acetate, using 1.4 μM M.Ptol at 55°C for 1 h. Numbers in the boxes above the blot represent magnesium acetate concentrations (mM). Reaction without protein had 10 mM magnesium acetate. (C) Progress curve of the reaction: Activity was monitored over 2–120 min using 1.4 μM M.Ptol, at 55°C. Numbers in the boxes above the blot represent incubation time (min). Reaction without protein was incubated for 120 min. (D) Reaction as a function of enzyme concentration: Activity was assayed with 1–2,500 nM M.Ptol, at 55°C for 45 min. Numbers in the boxes above the blot represent enzyme concentrations (nM). (E) Reaction as a function of temperature: Activity was assayed over 37–70°C, using 600 nM M.Ptol for 45 min. Numbers in the boxes above the blot represent reaction temperature. Reaction without protein was incubated at 55°C. (F) Reaction as a function of AdoMet concentration: Activity was assayed over 0–1,000 μM S-adenosyl methionine (AdoMet), using 400 nM M.Ptol at 55°C for 1 h. Numbers in the boxes above the blot represent AdoMet concentrations (μM). Reaction without protein had 100 μM AdoMet.

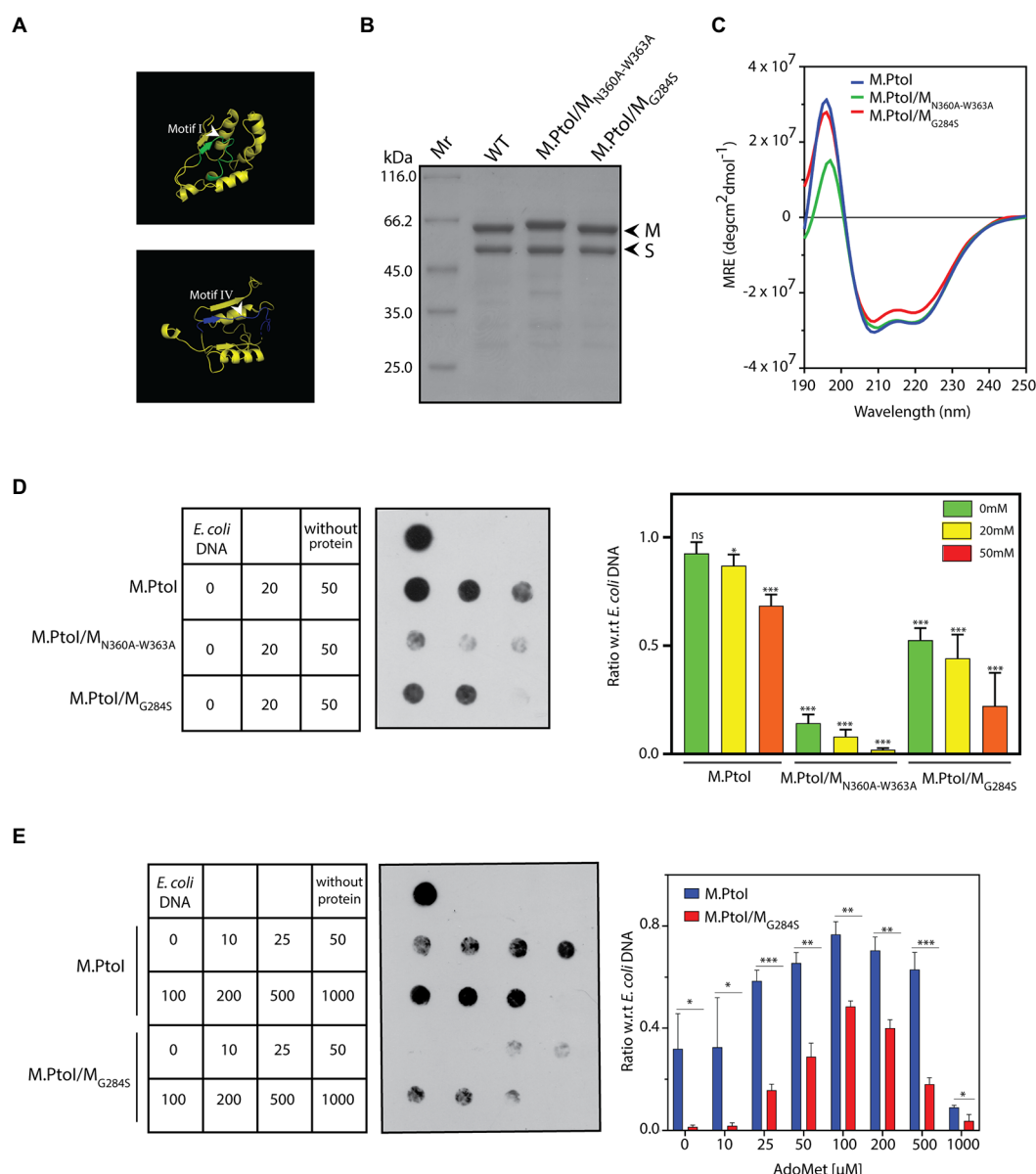


FIGURE 8

(A) Ribbon representation of motif I and motif IV derived from the predicted structure of M subunit of M.PtoI. (B) Coomassie stained SDS-PAGE analysis of purified recombinant M.PtoI/M_{N360A-W363A} and M.PtoI/M_{G284S}. Mr.: molecular weight marker. Arrowheads depict M and S subunits (63kDa and 48kDa respectively). (C) CD spectra of M.PtoI/M_{N360A-W363A} and M.PtoI/M_{G284S} depicted as a measure of mean residue ellipticity (MRE). (D,E) Dot blot analyses of methylation assays with mutant M.PtoI proteins. Assays were performed as described in the legend to Figure 7 unless otherwise specified. Left panel: loading scheme of dot blot. Middle panel: Dot blot of assay. Right panel: quantification of signals from methylation reactions as a ratio of *E. coli* genomic DNA signal, done using ImageJ software. Data represent average of three experiments. Error bars represent standard deviation. Student's two-tailed *t*-test was applied, **p*<0.05, ***p*<0.005, ****p*<0.0005, ns-not significant. (D) Reactions with M.PtoI, M.PtoI/M_{N360A-W363A} and M.PtoI/M_{G284S}. Activity assayed as a function of sodium chloride concentration (0–50mM) at pH 5. Numbers in boxes represent NaCl concentrations (mM). Reaction without protein had 20mM NaCl. (E) Reactions with M.PtoI and M.PtoI/M_{G284S}. Activity assayed as a function of AdoMet concentration (0–1,000μM). Numbers in boxes represent AdoMet concentrations (μM). Reaction without protein had 100μM AdoMet.

These data establish that the NPPW and xxGxxG motifs in M.PtoI are critical to M.PtoI activity.

4. Discussion

DNA methylation is widely prevalent in bacteria and archaea, and known to modulate various cellular processes as well as exist as part

of the organisms' defense mechanisms against foreign DNA. *Picrophilus torridus* grows in conditions of extreme pH, thriving at pH 0.7 to 2 (Figure 1A), and to date no other microorganism growing under these extreme conditions has been isolated. We initiated the study with examining whether the *Picrophilus* genome is methylated, and found it to carry the m6A modification mark but not the m5C mark (Figures 1B,C). Interestingly, the extent of adenine methylation was significantly lower when the organism was

grown at pH 0.7 as compared to when grown at pH 1 or 2. The commonest mediator of adenine methylation in prokaryotes is the Dam methylase. The *E. coli* Dam methylase regulates the timing of replication origin firing *via* methylation of the GATC sites in the *OriC* region. Dam-mediated GATC methylation also modulates methyl-directed mismatch repair (Adhikari and Curtis, 2016), and has been identified as a regulator of gene expression in *E. coli*, *Yersinia*, and *Salmonella typhimurium* (Falker et al., 2007; Brunet et al., 2020; Keceli Oguz et al., 2022). Widely prevalent across bacteria, some archaea also harbor the Dam. *Halobacterium saccharovororum* and *Methanobacterium* strain Ivanov have been reported to be Dam⁺ as early as 1984 (Barbeyron et al., 1984), and the *in vitro* methylation activity of the Dam methylase of *Pyrococcus horikoshii* has also been examined (Maynard-Smith et al., 2011). Koike et al. (2005) identified the Dam methylase in multiple archaea species through annotation of genome sequences. Their experimental analysis found *Thermoplasma volcanium*, *Thermoplasma acidophilum*, and *Pyrococcus* species OT3 genomes to carry methylated 5'-GATC sequences, while *Sulfolobus solfataricus* and *Sulfolobus shibatae* did not show evidence of 5'-GATC methylation. A subsequent study by Couturier and Lindas (2018) identified the methylation of 5'-GATC sites in *Sulfolobus acidocaldarius*. While the Dam methylase was identified by sequence annotation in *P. torridus* in the study by Koike et al. (2005), 5'-GATC methylation was not experimentally verified in that study. Our results indicated lack of 5'-GATC methylation in the *P. torridus* genome (Figure 1D), suggesting the absence of an active Dam in this organism and implicating the possible role of an R-M system in mediating adenine methylation in the organism.

The genes encoding the components of the single Type I R-M system identified through sequence annotation of the *P. torridus* genome lie in a cluster, on the lower strand of the genome between positions 85,223 and 91,251 (Figure 1E), and the modification methylase component of this system, named M.PtoI, was investigated for its ability to methylate adenine residues. The structure of M.PtoI and the interacting interfaces of the M and S subunits were predicted to be conserved with those of other Type I R-M systems (Figures 3–5), and the predictions were experimentally verified in part by the creation and analysis of suitable M and S mutant proteins (Figure 6), which confirmed that the M subunits interacted with the central conserved region (CCR) of the S subunit through their C-terminal helical domains (CTDs). The ability of M.PtoI to methylate DNA was assessed using the recombinant protein in *in vitro* assays. Activity was found to be optimal at pH and temperature conditions that reflected the organism's lifestyle and physiological state (Figures 7A,E). The M.PabI Type II modification methylase of *Pyrococcus abyssi* has been found to methylate DNA at 95°C as much as at 65°C although optimally at 85°C, and demonstrated optimal activity at pH 5.8 to 6.7, similarly reflecting the growth and intracellular conditions of the hyperthermophilic organism (Watanabe et al., 2006). Magnesium ions were found to be essential for M.PtoI activity (Figure 7B). Magnesium ions (but not calcium or manganese ions) have been found to be essential for M.EcoP15I methyltransferase activity as well. CD spectroscopy analysis of M.EcoP15I titrated against increasing concentrations of magnesium revealed a secondary structure alteration in the enzyme in response to magnesium ions, suggesting that magnesium ion-induced conformational changes preceded the catalysis reaction. Mutation of the magnesium binding motif resulted in enzyme inactivity (Bist and Rao, 2003). Earlier studies with EcoBI

as well as MmeI methyltransferases indicated that the presence of magnesium ions, though not essential to the methylation reaction, stimulated it (Lautenberger and Linn, 1972; Tucholski et al., 1995). M.PtoI was found to be active over AdoMet concentrations 25–500 μM, exhibiting substrate inhibition at 1 mM AdoMet. The purified protein most likely harbored endogenous AdoMet as it exhibited methylation activity even when AdoMet was not added to the reaction (Figure 7F), a feature of a few other methyltransferases as well (Dreier et al., 1996; Jurenaite-Urbanaviciene et al., 2001; Bath et al., 2002).

M.PtoI harbored all the motifs typifying adenine methyltransferases (Figure 2). The functional roles of these motifs have been uncovered by a combination of structural and biochemical studies of other DNA methyltransferases. The methylation reaction is mediated by a base-flipping mechanism, wherein the target adenine residue flips out of the double helix into the catalytic pocket (Goedecke et al., 2001). Motif IV (NPPW) is a primary constituent of the catalytic pocket, and the interaction of the flipped target adenine with the asparagine residue is crucial for catalysis. The data from structural and biochemical studies lead us to believe that hydrogen bonding between the N⁶-amino group of the adenine (which serves as the donor) and the asparagine side chain of the NPPW motif possibly polarizes the amino group, promoting the direct transfer of the methyl group from AdoMet to the adenine N⁶ position (Pogolotti et al., 1988; Labahn et al., 1994; Goedecke et al., 2001; Bheemanaik et al., 2006). The tryptophan residue at the last position of the motif can be replaced with any other aromatic residue without any significant impact on catalytic activity, but replacement with other residues has a negative impact on activity (Willcock et al., 1994; Ahmad et al., 1995; Pues et al., 1999). In keeping with these findings, the ability of the M.PtoI/M_{N360A-W363A} motif IV mutant to methylate DNA was severely compromised (Figure 8D). Structural studies of AdoMet-bound enzyme as well as in-depth mutational analyses coupled to biochemical assays have together revealed that motif I plays a major role in AdoMet binding, though it is not directly involved in catalysis *per se* (Labahn et al., 1994; Willcock et al., 1994; Ahmad et al., 1995; Roth et al., 1998; Goedecke et al., 2001). It is possible that the M.PtoI/M_{G284A} mutant displays weaker binding of AdoMet, as the mutant (unlike the wild type enzyme) does not appear to be harboring endogenous AdoMet when purified from *E. coli* since no catalysis is detected at 0 μM AdoMet (Figure 8E). Collectively, the data suggest that the functions of motif I and motif IV may be conserved in M.PtoI.

To date, only one Type I R-M system has been investigated in archaea: that of *Haloferax volcanii*. While several MTases have been annotated in the *Haloferax* genome sequence, a combination of single molecule real time sequencing and gene knockout analyses have demonstrated that only two motifs are methylated in the genome, one of which is methylated at a cytosine residue *via* a Type II enzyme and the second of which (5'-GCAM6BN6VTGC-3') is methylated by a Type I adenine modification methylase (Ouellette et al., 2015, 2018). The Type I modification methylase harbors all the motifs that typify these enzymes. Interestingly and unusually, adenine methylation was detected only in the first half of the bipartite recognition sequence, suggesting that either the second half of the sequence does not exhibit methylation on the lower strand, or methylation does not occur equivalently on the two strands and thus remains undetected during sequencing.

Taken together, the data presented here establish that even organisms growing in such extreme conditions (as *Picrophilus torridus* does) exhibit DNA methylation. We find evidence for m6A but not m5C methylation in *P. torridus*, and the absence of GATC methylation indicates that the organism lacks an active Dam. Our results suggest that the Type I R-M system annotated in the *P. torridus* genome sequence is active and mediates m6A methylation. We were unable to analyze the *in vivo* role of M.PtoI as *Picrophilus* species are not amenable to genetic manipulations. Attempts to carry out complementation analyses in *E. coli dam*[−] strain (a kind gift from the Lab Collection of Manjula Reddy at CSIR-CCMB, India) did not succeed as the M and S subunits did not get expressed (from the tetracycline-driven promoter system in pASKIBA43PLUS) in this strain, which is deficient in *argU*, *ileY*, and *leuW* tRNA genes (Supplementary Figure S4). Future studies will be directed toward identifying the target site of M.PtoI and characterizing the PtoI restriction enzyme (M2R2S1).

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

SS and MG designed the project. PG and AS performed the research and prepared the figures. PG, AS, MG, and SS analyzed the data. SS wrote the paper. All authors contributed to the article and approved the submitted version.

References

- Adhikari, S., and Curtis, P. D. (2016). DNA methyltransferases and epigenetic regulation in bacteria. *FEMS Microbiol. Rev.* 40, 575–591. doi: 10.1093/femsre/fuw023
- Ahlgren, N. A., Chen, Y., Needham, D. M., Parada, A. E., Sachdeva, R., Trinh, V., et al. (2017). Genome and epigenome of a novel marine Thaumarchaeota strain suggest viral infection, phosphorothioation DNA modification and multiple restriction systems. *Environ. Microbiol.* 19, 2434–2452. doi: 10.1111/1462-2920.13768
- Ahmad, I., Krishnamurthy, V., and Rao, D. N. (1995). DNA recognition by the EcoP15I and EcoPI modification methyltransferases. *Gene* 157, 143–147. doi: 10.1016/0378-1119(95)00671-R
- Anton, B. P., and Roberts, R. J. (2021). Beyond restriction modification: Epigenomic roles of DNA methylation in prokaryotes. *Annu. Rev. Microbiol.* 75, 129–149. doi: 10.1146/annurev-micro-040521-035040
- Arora, J., Goswami, K., and Saha, S. (2014). Characterization of the replication initiator Orc1/Cdc6 from the archaeon *Picrophilus torridus*. *J. Bacteriol.* 196, 276–286. doi: 10.1128/JB.01020-13
- Barbeyron, T., Kean, K., and Forterre, P. (1984). DNA adenine methylation of GATC sequences appeared recently in the *Escherichia coli* lineage. *J. Bacteriol.* 160, 586–590. doi: 10.1128/jb.160.2.586-590.1984
- Bath, A. J., Milsom, S. E., Gormley, N. A., and Halford, S. E. (2002). Many type II restriction endonucleases interact with two recognition sites before cleaving DNA. *J. Biol. Chem.* 277, 4024–4033. doi: 10.1074/jbc.M108441200
- Benkert, P., Kunzli, M., and Schwede, T. (2009). QMEAN server for protein model quality estimation. *Nucleic Acids Res.* 37, W510–W514. doi: 10.1093/nar/gkp322
- Bheemanaik, S., Reddy, Y. V., and Rao, D. N. (2006). Structure, function and mechanism of exocyclic DNA methyltransferases. *Biochem. J.* 399, 177–190. doi: 10.1042/BJ20060854
- Bist, P., and Rao, D. N. (2003). Identification and mutational analysis of Mg²⁺ binding site in EcoP15I DNA methyltransferase: involvement in target base eversion. *J. Biol. Chem.* 278, 41837–41848. doi: 10.1074/jbc.M307053200
- Blow, M. J., Clark, T. A., Daum, C. G., Deutschbauer, A. M., Fomenkov, A., Fries, R., et al. (2016). The Epigenomic landscape of prokaryotes. *PLoS Genet.* 12:e1005854. doi: 10.1371/journal.pgen.1005854
- Brunet, Y. R., Bernard, C. S., and Cascales, E. (2020). Fur-dam regulatory interplay at an internal promoter of the Enterococcal *Escherichia coli* type VI secretion scII gene cluster. *J. Bacteriol.* 202:20. doi: 10.1128/JB.00075-20
- Bujnicki, J. M. (2002). Sequence permutations in the molecular evolution of DNA methyltransferases. *BMC Evol. Biol.* 2:3. doi: 10.1186/1471-2148-2-3
- Couturier, M., and Lindas, A. C. (2018). The DNA methylome of the hyperthermoacidophilic crenarchaeon *Sulfolobus acidocaldarius*. *Front. Microbiol.* 9:137. doi: 10.3389/fmicb.2018.00137
- Dreier, J., MacWilliams, M. P., and Bickle, T. A. (1996). DNA cleavage by the type IC restriction-modification enzyme EcoR124II. *J. Mol. Biol.* 264, 722–733. doi: 10.1006/jmbi.1996.0672
- Falkner, S., Schilling, J., Schmidt, M. A., and Heussipp, G. (2007). Overproduction of DNA adenine methyltransferase alters motility, invasion, and the lipopolysaccharide O-antigen composition of *Yersinia enterocolitica*. *Infect. Immun.* 75, 4990–4997. doi: 10.1128/IAI.00457-07
- Futterer, O., Angelov, A., Liesegang, H., Gottschalk, G., Schleper, C., Schepers, B., et al. (2004). Genome sequence of *Picrophilus torridus* and its implications for life around pH 0. *Proc. Natl. Acad. Sci. U. S. A.* 101, 9091–9096. doi: 10.1073/pnas.0401356101
- Gao, Y., Cao, D., Zhu, J., Feng, H., Luo, X., Liu, S., et al. (2020). Structural insights into assembly, operation and inhibition of a type I restriction-modification system. *Nat. Microbiol.* 5, 1107–1118. doi: 10.1038/s41564-020-0731-z
- Goedecke, K., Pignot, M., Goody, R. S., Scheidig, A. J., and Weinhold, E. (2001). Structure of the N6-adenine DNA methyltransferase M.TaqI in complex with DNA and a cofactor analog. *Nat. Struct. Biol.* 8, 121–125. doi: 10.1038/84104
- Grogan, D. W. (2003). Cytosine methylation by the SmaI restriction-modification system: implications for genetic fidelity in a hyperthermophilic archaeon. *J. Bacteriol.* 185, 4657–4661. doi: 10.1128/JB.185.15.4657-4661.2003

Acknowledgments

We thank the Central Instrumentation Facility, University of Delhi South Campus for DNA sequencing and the use of CD spectrometer. We thank Manjula Reddy (CSIR-CCMB, Hyderabad, India) for providing us with an *E. coli dam*[−] strain from her Lab Collection.

Conflict of interest

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.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2023.1126750/full#supplementary-material>

- Ishikawa, K., Watanabe, M., Kuroita, T., Uchiyama, I., Bujnicki, J. M., Kawakami, B., et al. (2005). Discovery of a novel restriction endonuclease by genome comparison and application of a wheat-germ-based cell-free translation assay: PabI (5'-GTA/C) from the hyperthermophilic archaeon *Pyrococcus abyssi*. *Nucleic Acids Res.* 33:e112. doi: 10.1093/nar/gni113
- Jurenaite-Urbanaviciene, S., Kazlauskienė, R., Urbelyte, V., Maneliene, Z., Petruseyte, M., Lubys, A., et al. (2001). Characterization of BseMII, a new type IV restriction-modification system, which recognizes the pentanucleotide sequence 5'-CTCAG(N)(10/8). *Nucleic Acids Res.* 29, 895–903. doi: 10.1093/nar/29.4.895
- Keceli Oguz, S., Has, E. G., Akcelik, N., and Akcelik, M. (2022). Phenotypic impacts and genetic regulation characteristics of the DNA adenine methylase gene (dam) in *Salmonella typhimurium* biofilm forms. *Res. Microbiol.* 174:103991. doi: 10.1016/j.resmic.2022.103991
- Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., and Sternberg, M. J. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* 10, 845–858. doi: 10.1038/nprot.2015.053
- Kennaway, C. K., Obarska-Kosinska, A., White, J. H., Tuszyńska, I., Cooper, L. P., Bujnicki, J. M., et al. (2009). The structure of M.EcoKI type I DNA methyltransferase with a DNA mimic antirestriction protein. *Nucleic Acids Res.* 37, 762–770. doi: 10.1093/nar/gkn988
- Kim, J. S., Degiovanni, A., Jancarik, J., Adams, P. D., Yokota, H., Kim, R., et al. (2005). Crystal structure of DNA sequence specificity subunit of a type I restriction-modification enzyme and its functional implications. *Proc. Natl. Acad. Sci. U. S. A.* 102, 3248–3253. doi: 10.1073/pnas.0409851102
- Koike, H., Yokoyama, K., Kawashima, T., Yamasaki, T., Makino, S., Clowney, L., et al. (2005). GATC methylation by dam methylase in bacteria: its roles and possible transcription regulation by an FRP. *Proc. Jpn. Acad.* 81, 278–290. doi: 10.2183/pjab.81.278
- Labahn, J., Granzin, J., Schluckebier, G., Robinson, D. P., Jack, W. E., Schildkraut, I., et al. (1994). Three-dimensional structure of the adenine-specific DNA methyltransferase M.Taq I in complex with the cofactor S-adenosylmethionine. *Proc. Natl. Acad. Sci. U. S. A.* 91, 10957–10961. doi: 10.1073/pnas.91.23.10957
- Lautenberger, J. A., and Linn, S. (1972). The deoxyribonucleic acid modification and restriction enzymes of *Escherichia coli* B. I. Purification, subunit structure, and catalytic properties of the modification methylase. *J. Biol. Chem.* 247, 6176–6182. doi: 10.1016/S0021-9258(19)44779-0
- Lee, S. H., Kim, M. S., Lee, J. H., Kim, T. W., Bae, S. S., Lee, S. M., et al. (2016). Adaptive engineering of a hyperthermophilic archaeon on CO and discovering the underlying mechanism by multi-omics analysis. *Sci. Rep.* 6:22896. doi: 10.1038/srep22896
- Loenen, W. A., Dryden, D. T., Raleigh, E. A., and Wilson, G. G. (2014a). Type I restriction enzymes and their relatives. *Nucleic Acids Res.* 42, 20–44. doi: 10.1093/nar/gkt847
- Loenen, W. A., Dryden, D. T., Raleigh, E. A., Wilson, G. G., and Murray, N. E. (2014b). Highlights of the DNA cutters: a short history of the restriction enzymes. *Nucleic Acids Res.* 42, 3–19. doi: 10.1093/nar/gkt990
- Lunnen, K. D., Morgan, R. D., Timan, C. J., Krzycki, J. A., Reeve, J. N., and Wilson, G. G. (1989). Characterization and cloning of MwoI (GCN7GC), a new type-II restriction-modification system from *Methanobacterium wolfei*. *Gene* 77, 11–19. doi: 10.1016/0378-1119(89)90354-5
- Malone, T., Blumenthal, R. M., and Cheng, X. (1995). Structure-guided analysis reveals nine sequence motifs conserved among DNA amino-methyltransferases, and suggests a catalytic mechanism for these enzymes. *J. Mol. Biol.* 253, 618–632. doi: 10.1006/jmbi.1995.0577
- Maynard-Smith, M. D., Mckelvie, J. C., Wood, R. J., Harmer, J. E., Ranasinghe, R. T., Williams, C. L., et al. (2011). Direct and continuous fluorescence-based measurements of *Pyrococcus horikoshii* DNA N-6 adenine methyltransferase activity. *Anal. Biochem.* 418, 204–212. doi: 10.1016/j.ab.2011.07.023
- McConnell, D. J., Searcy, D. G., and Sutcliffe, J. G. (1978). A restriction enzyme Tha I from the thermophilic mycoplasma *Thermoplasma acidophilum*. *Nucleic Acids Res.* 5, 1729–1739. doi: 10.1093/nar/5.6.1729
- Mooers, B. H. M. (2016). Simplifying and enhancing the use of PyMOL with horizontal scripts. *Protein Sci.* 25, 1873–1882. doi: 10.1002/pro.2996
- Morgan, R., Xiao, J., and Xu, S. (1998). Characterization of an extremely thermostable restriction enzyme, PspGI, from a *Pyrococcus* strain and cloning of the PspGI restriction-modification system in *Escherichia coli*. *Appl. Environ. Microbiol.* 64, 3669–3673. doi: 10.1128/AEM.64.10.3669-3673.1998
- Nolling, J., and De Vos, W. M. (1992a). Characterization of the archaeal, plasmid-encoded type II restriction-modification system MthTI from *Methanobacterium thermoformicicum* THF: homology to the bacterial NgoPII system from *Neisseria gonorrhoeae*. *J. Bacteriol.* 174, 5719–5726. doi: 10.1128/jb.174.17.5719-5726.1992
- Nolling, J., and De Vos, W. M. (1992b). Identification of the CTAG-recognizing restriction-modification systems MthZI and MthFI from *Methanobacterium thermoformicicum* and characterization of the plasmid-encoded mthZIM gene. *Nucleic Acids Res.* 20, 5047–5052. doi: 10.1093/nar/20.19.5047
- Ouellette, M., Gogarten, J. P., Lajoie, J., Makkay, A. M., and Papke, R. T. (2018). Characterizing the DNA methyltransferases of *Haloflex volcanii* via bioinformatics, gene deletion, and SMRT sequencing. *Genes (Basel)* 9:129. doi: 10.3390/genes9030129
- Ouellette, M., Jackson, L., Chimileski, S., and Papke, R. T. (2015). Genome-wide DNA methylation analysis of *Haloflex volcanii* H26 and identification of DNA methyltransferase related PD-(D/E)XK nuclease family protein HVO_A0006. *Front. Microbiol.* 6:251. doi: 10.3389/fmicb.2015.00251
- Pingoud, V., Conzelmann, C., Kinzebach, S., Sudina, A., Metelev, V., Kubareva, E., et al. (2003). PspGI, a type II restriction endonuclease from the extreme thermophile *Pyrococcus* sp.: structural and functional studies to investigate an evolutionary relationship with several mesophilic restriction enzymes. *J. Mol. Biol.* 329, 913–929. doi: 10.1016/S0022-2836(03)00523-0
- Pingoud, A., Wilson, G. G., and Wende, W. (2014). Type II restriction endonucleases—a historical perspective and more. *Nucleic Acids Res.* 42, 7489–7527. doi: 10.1093/nar/gku447
- Pogolotti, A. L. Jr., Ono, A., Subramaniam, R., and Santi, D. V. (1988). On the mechanism of DNA-adenine methylase. *J. Biol. Chem.* 263, 7461–7464. doi: 10.1016/S0021-9258(18)68520-5
- Prangishvili, D. A., Vashakidze, R. P., Chelidze, M. G., and Gabriadze, I. (1985). A restriction endonuclease Sual from the thermoacidophilic archaeobacterium *Sulfolobus acidocaldarius*. *FEBS Lett.* 192, 57–60. doi: 10.1016/0014-5793(85)80042-9
- Pues, H., Bleimling, N., Holz, B., Wolcke, J., and Weinhold, E. (1999). Functional roles of the conserved aromatic amino acid residues at position 108 (motif IV) and position 196 (motif VIII) in base flipping and catalysis by the N6-adenine DNA methyltransferase from *Thermus aquaticus*. *Biochemistry* 38, 1426–1434. doi: 10.1021/bi9818016
- Rao, D. N., Dryden, D. T., and Bheemanaik, S. (2014). Type III restriction-modification enzymes: a historical perspective. *Nucleic Acids Res.* 42, 45–55. doi: 10.1093/nar/gkt616
- Roth, M., Helm-Kruse, S., Friedrich, T., and Jeltsch, A. (1998). Functional roles of conserved amino acid residues in DNA methyltransferases investigated by site-directed mutagenesis of the EcoRV adenine-N6-methyltransferase. *J. Biol. Chem.* 273, 17333–17342. doi: 10.1074/jbc.273.28.17333
- Schleper, C., Puehler, G., Holz, I., Gambacorta, A., Janekovic, D., Santarius, U., et al. (1995). Picrophilus gen. Nov., fam. Nov.: a novel aerobic, heterotrophic, thermoacidophilic genus and family comprising archaea capable of growth around pH 0. *J. Bacteriol.* 177, 7050–7059. doi: 10.1128/jb.177.24.7050-7059.1995
- Schmid, K., Thomm, M., Laminet, A., Laue, F. G., Kessler, C., Stetter, K. O., et al. (1984). Three new restriction endonucleases Mael, MaelII and MaelIII from *Methanococcus aeolicus*. *Nucleic Acids Res.* 12, 2619–2628. doi: 10.1093/nar/12.6.2619
- 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
- Suzuki, S., and Kurosawa, N. (2016). Disruption of the gene encoding restriction endonuclease Sual and development of a host-vector system for the thermoacidophilic archaeon *Sulfolobus acidocaldarius*. *Extremophiles* 20, 139–148. doi: 10.1007/s00792-016-0807-0
- Tucholski, J., Skowron, P. M., and Podhajska, A. J. (1995). MmeI, a class-IIS restriction endonuclease: purification and characterization. *Gene* 157, 87–92. doi: 10.1016/0378-1119(94)00787-5
- Watanabe, M., Yuzawa, H., Handa, N., and Kobayashi, I. (2006). Hyperthermophilic DNA methyltransferase M.PabI from the archaeon *Pyrococcus abyssi*. *Appl. Environ. Microbiol.* 72, 5367–5375. doi: 10.1128/AEM.00433-06
- Willcock, D. F., Dryden, D. T., and Murray, N. E. (1994). A mutational analysis of the two motifs common to adenine methyltransferases. *EMBO J.* 13, 3902–3908. doi: 10.1002/j.1460-2075.1994.tb06701.x
- Zatopek, K. M., Burkhart, B. W., Morgan, R. D., Gehring, A. M., Scott, K. A., Santangelo, T. J., et al. (2021). The Hyperthermophilic restriction-modification Systems of *Thermococcus kodakarensis* protect genome integrity. *Front. Microbiol.* 12:657356. doi: 10.3389/fmicb.2021.657356



OPEN ACCESS

EDITED BY

Michel Geovanni Santiago-Martínez,
University of Connecticut,
United States

REVIEWED BY

José Luis Aguirre-Noyola,
National Autonomous University of Mexico,
Mexico
Fernando Medina Ferrer,
University of California,
Berkeley, United States

*CORRESPONDENCE

Carmen Pire
✉ carmen.pire@ua.es

SPECIALTY SECTION

This article was submitted to
Biology of Archaea,
a section of the journal
Frontiers in Microbiology

RECEIVED 27 November 2022

ACCEPTED 21 February 2023

PUBLISHED 16 March 2023

CITATION

Miralles-Robledillo JM, Martínez-Espinosa RM
and Pire C (2023) Analysis of the external
signals driving the transcriptional regulation of
the main genes involved in denitrification in
Haloferax mediterranei.
Front. Microbiol. 14:1109550.
doi: 10.3389/fmicb.2023.1109550

COPYRIGHT

© 2023 Miralles-Robledillo, Martínez-Espinosa
and Pire. 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.

Analysis of the external signals driving the transcriptional regulation of the main genes involved in denitrification in *Haloferax mediterranei*

Jose María Miralles-Robledillo¹, Rosa María Martínez-Espinosa^{1,2}
and Carmen Pire^{1,2*}

¹Biochemistry and Molecular Biology Division, Agrochemistry and Biochemistry Department, Faculty of Sciences, University of Alicante, Alicante, Spain, ²Multidisciplinary Institute for Environmental Studies "Ramón Margalef", University of Alicante, Alicante, Spain

Haloferax mediterranei is the model microorganism for the study of the nitrogen cycle in haloarchaea. This archaeon not only assimilate *N*-species such as nitrate, nitrite, or ammonia, but also it can perform denitrification under low oxygen conditions, using nitrate or nitrite as alternative electron acceptors. However, the information currently available on the regulation of this alternative respiration in this kind of microorganism is scarce. Therefore, in this research, the study of haloarchaeal denitrification using *H. mediterranei* has been addressed by analyzing the promoter regions of the four main genes of denitrification (*narGH*, *nirK*, *nor*, and *nosZ*) through bioinformatics, reporter gene assays under oxic and anoxic conditions and by site-directed mutagenesis of the promoter regions. The results have shown that these four promoter regions share a common semi-palindromic motif that plays a role in the control of the expression levels of *nor* and *nosZ* (and probably *nirK*) genes. Regarding the regulation of the genes under study, it has been concluded that *nirK*, *nor*, and *nosZ* genes share some expression patterns, and therefore their transcription could be under the control of the same regulator whereas *nar operon* expression displays differences, such as the activation by dimethyl sulfoxide with respect to the expression in the absence of an electron acceptor, which is almost null under anoxic conditions. Finally, the study with different electron acceptors demonstrated that this haloarchaea does not need complete anoxia to perform denitrification. Oxygen concentrations around 100 μM trigger the activation of the four promoters. However, a low oxygen concentration *per se* is not a strong signal to activate the promoters of the main genes involved in this pathway; high activation also requires the presence of nitrate or nitrite as final electron acceptors.

KEYWORDS

haloarchaea, *Haloferax mediterranei*, denitrification, transcriptional regulation, promoter activity

1. Introduction

Denitrification is one of the main pathways of the nitrogen cycle. This process encompasses a series of redox reactions in which nitrate (NO₃⁻) can be reduced sequentially to nitrite (NO₂⁻) by the respiratory nitrate reductase (encoded by the *narGH* genes), nitric oxide (NO) by the respiratory nitrite reductase (encoded by the *nirK* gene), nitrous oxide (N₂O) by the nitric oxide

reductase (encoded by the *nor* gene), and finally, dinitrogen (N_2) by the nitrous oxide reductase (encoded by the *nosZ* gene) (Richardson et al., 2009). This process shows a close relationship with climate change, especially due to two of their intermediates that participate directly or indirectly in ozone layer depletion: NO and N_2O (Ehhalt et al., 2001; Richardson et al., 2009; Cavigelli et al., 2012; Myhre et al., 2013). NO is considered a precursor of greenhouse gasses, whereas N_2O shows a potent greenhouse effect, being 300 times as potent as carbon dioxide (CO_2) at heating the atmosphere (Ehhalt et al., 2001; Cavigelli et al., 2012; Myhre et al., 2013; Masson-Delmotte et al., 2019). Due to this reason, worldwide efforts in mitigating their release into the atmosphere have caused an increase in the number of publications that correlate denitrification with climate change (Masson-Delmotte et al., 2019; Miralles-Robledillo et al., 2021).

N_2O emissions mostly come from agricultural practices due to soil microbial metabolism in fertilized soils (Cavigelli et al., 2012; Zhang et al., 2016; Feng et al., 2018; Muntean et al., 2019; Olivier and Peters, 2019). Therefore, many studies about denitrification focus on these systems aiming to improve the use of N-species as fertilizers (Shcherbak et al., 2014; Wang et al., 2018). Nevertheless, other less studied and underestimated systems in terms of N_2O emissions are saline and hypersaline environments such as submarine salt domes, salted lagoons, or salt marshes (Torregrosa-Crespo et al., 2018). The exact extent of these ecosystems on our planet is not known, although it has been described that they are increasing due to climate change and anthropogenic activities contributing to the expansion of arid and semiarid ecosystems (Delgado-Baquerizo et al., 2013; Medhaug et al., 2017; Torregrosa-Crespo et al., 2018; Schuler et al., 2019). Moreover, intakes of nitrate, nitrite, and ammonium in these areas are rising due to anthropogenic activities (Firestone et al., 1980; Conrad, 1996; Kool et al., 2011; Martínez-Espinosa et al., 2011).

Nowadays, studies carried out in these saline/hypersaline environments have been mainly related to the ecology of living beings that can tolerate the presence of high salt concentrations as well as to the characterization of microbial population interactions. It has been reported that in hypersaline environments the microbial population which is most prevalent is the haloarchaea, the *Haloferacaceae* family being the best well-studied (Oren, 2002). Some important representatives of this family are *Haloferax*, *Haloarcula*, and *Natronomonas* genera due to their possible biotechnological applications and their important role in the biogeochemical cycles of these ecosystems (Bonete et al., 2008; Giani et al., 2019; Miralles-Robledillo et al., 2021). In this context, the *Haloferax* genus is the best characterized from the biochemical point of view and it has been chosen as a model organism in many studies in order to decipher how N-cycle works in hypersaline environments (Martínez-Espinosa et al., 2007; Torregrosa-Crespo et al., 2016).

Haloferax mediterranei is the model archaeon for dissimilatory nitrogen metabolism studies due to its metabolic versatility. Not only can it assimilate nitrate, nitrite, and ammonia, but also shows the capacity for denitrification, being capable of reducing completely nitrate to dinitrogen using this route as alternative respiration to standard aerobic respiration (Esclapez et al., 2014; Oren, 2017; Torregrosa-Crespo et al., 2019). This adaptation is advantageous for the cells since saline systems show low oxygen solubility and therefore oxygen concentration is usually lower than in mesophilic environments (Oren, 2016). Over the years, haloarchaeal denitrification has been explored from different points of view: efforts have focused on the purification of the main enzymes and more recently on bioinformatic analyzes and the physiological characterization of denitrification (Lledó et al., 2004; Esclapez et al., 2013; Torregrosa-Crespo et al., 2017, 2019). However, unlike in bacteria, little is known about the genetic regulation of denitrification in this type of microorganism, which represents a major knowledge gap about this pathway (Hattori et al., 2016; Torregrosa-Crespo et al., 2020a; Koyanagi et al., 2021). Furthermore, there are no homologous genes in haloarchaea for bacterial regulators related to denitrification such as FnrP, NNR, or NarR, which makes it more difficult to identify the genetic network controlling this alternative respiration (Spiro, 2016).

Torregrosa-Crespo et al. (2020a) studied gas emissions due to the growth of *H. mediterranei* under anaerobic growth in the presence of nitrate and nitrite. In addition, they also performed a transcriptomic analysis of the four main denitrification genes (*narGH*, *nirK*, *nor*, and *nosZ*) under denitrification conditions (presence of nitrate and anaerobiosis) (Torregrosa-Crespo et al., 2020a). Nevertheless, to understand haloarchaeal denitrification in depth, it is essential to explore different conditions that could provide information about denitrification activation as well as about the nature of the signals that can modulate this alternative respiration in hypersaline environments. Thus, in this study *H. mediterranei* anaerobic growth and denitrification promoter activation have been analyzed under the presence of different final alternative electron acceptors (NO_3^- , NO_2^- , and DMSO) and without them.

This work has delved into the regulation of denitrification in *H. mediterranei* through bioinformatic and experimental analysis of the promoter regions of the genes encoding the four main enzymes of denitrification. Common regulatory sequences have been found in their promoter regions and promoter activities have been tested under different growth conditions aiming to complete the regulatory model for this respiratory pathway in haloarchaea.

2. Materials and methods

2.1. Reporter gene assay: Promoter cloning and β -galactosidase enzymatic assay

Reporter gene assays were performed using the β -galactosidase (*bgaH*) gene from *Haloferax lucentense* as a reporter gene cloned in an expression plasmid. The plasmid used (pVA513) was a modified version of the plasmid pMDS132 with an extra *NcoI* restriction site at the start ATG of the *bgaH* gene (this plasmid was kindly provided by Dr. Mike Dyall-Smith, University of Melbourne, Australia) (Holmes and Dyall-Smith, 2000). The vector map is displayed in Supplementary Figure S1A.

Abbreviations: *narGH*, nitrate reductase; *nirK*, nitrite reductase; *nor*, nitric oxide reductase; *nosZ*, nitrous oxide reductase; *narp*, promoter region of nitrate reductase operon; *nirp*, promoter region of the nitrite reductase gene; *norp*, promoter region of the nitric oxide reductase gene; *nosp*, promoter region of the nitrous oxide reductase operon; OD_{600 nm}, Optical density value at 600 nanometers; NEDA, N-(1-naphthyl)-ethylenediamine-dihydrochloride; Abs, Absorbance; DMSO, Dimethyl sulfoxide; MOPS, (3-(N-morpholino) propanesulfonic acid; ONPG, ortho-nitro-phenol β -D-galactoside; ONP, o-nitrophenol.

TABLE 1 List of primers and genes whose promoter regions were amplified.

Gene (Location)	Sequence	Position	Amplicon size (bp)
<i>narGH</i> (pHME322)	GTCGCCTCCAAGCTTCTTCCGC	20220–20198	420
	AAGATGGCCATGGTCTCTCGCCTCAT	19801–19826	
<i>nirK</i> (chromosome)	CGGGCAACAAGCTTCGGTCACG	151234–151255	267
	ACGTCCGTCCCATGGTTGTTGATAGCAT	151500–151473	
<i>nor</i> (chromosome)	GGGAGCTCATCAAAGCTTGAAGTAACGG	145162–154189	159
	TTCGCCATGGTCTTACGCTTGAGTTCCAT	145320–145292	
<i>nosZ</i> (pHME322)	CGTTGACTAAGCTTCCGGGTGAGCAC	4829–4804	501
	CGTTCCTCCATGGTATGTTTCCCTGCCAT	4329–4359	

TABLE 2 Strains of *H. mediterranei* used in reporter gene assays.

Strain	Promoter under study	Genotype
pVA513 + <i>narp</i>	<i>nar</i> promoter (<i>narp</i>)	<i>narp::bgaH</i> + <i>nov</i> *
pVA513 + <i>nirp</i>	<i>nir</i> promoter (<i>nirp</i>)	<i>nirp::bgaH</i> + <i>nov</i> *
pVA513 + <i>norp</i>	<i>nor</i> promoter (<i>norp</i>)	<i>norp::bgaH</i> + <i>nov</i> *
pVA513 + <i>nosp</i>	<i>nos</i> promoter (<i>nosp</i>)	<i>nosp::bgaH</i> + <i>nov</i> *

**nov*: novobiocin resistance.

Promoter regions of the studied genes (Table 1; Supplementary Figures S2, S3) were amplified by PCR from *H. mediterranei* R4 ATCC 33500 genomic DNA (genome consists of one chromosome of 2,948,887 bp and three megaplasms, pHME132 of 131,975 bp, pHME322 of 321,907 bp, and pHME505 of 504,704 bp) (Rodríguez-Valera et al., 1980; DasSarma et al., 2019). For these amplifications, all the spaces between two coding regions were considered promoter regions ensuring that the possible TATA and BRE boxes are inside this amplified region. Restriction sites for *NcoI* and *HindIII* enzymes were added to the primers used for subsequent cloning into the corresponding vector. PCR products were purified using the GFX PCR DNA and Gel Band Purification Kit from Cytiva and cloned into pJET1.2 linearized plasmid using CloneJET PCR Cloning Kit (Thermo Scientific) (Vector map is shown in Supplementary Figure S1B). *Escherichia coli* JM110 was transformed by the heat-shock method with pJET1.2 plasmid containing promoter regions to produce a high amount of unmethylated plasmid and facilitate restriction (Asif et al., 2017). The pJET1.2 constructions were linearized by the *NcoI* and *HindIII* restriction enzymes, and the inserts were purified with the E.Z.N.A.® Plasmid DNA Mini Kit II (Omega Bio-tek) and ligated in the pVA513 plasmid using T4 DNA ligase (Thermo Scientific). *Escherichia coli* JM110 were transformed with pVA513 constructs by the heat-shock method (Asif et al., 2017). The plasmid was isolated from *E. coli* JM110 with the E.Z.N.A.® Plasmid DNA Mini Kit II and transformed into *H. mediterranei* R4 ATCC 33500 using the standard PEG-mediated transformation of haloarchaea (Dyall-Smith, 2009). The correct vector construction was verified by sequencing the plasmid isolated from *H. mediterranei*. Sequencing was carried out by the Research Support Services of the University of Alicante using the Sanger method. PCR product purification, plasmid purification, and ligation were carried out following manufacturer instructions. Transformed strains of *H. mediterranei* R4 ATCC 33500 and its respective genotypes are shown in Table 2.

Protein extracts from these cultures were prepared as follows for β -galactosidase activity measurement. An aliquot of culture was extracted and centrifuged (13,000 g, 4°C, 5 min). The volume of the aliquot was taken according to the optical density value at 600 nm (OD₆₀₀ nm) of the cultures, ranging from 15 mL for lower cell density values to 5 mL for higher values. Pellet was weighed and resuspended in bgaH buffer (2.5 M NaCl, 50 mM Tris-HCl, 10 μ M MnCl₂, 0.1% (v/v) β -mercaptoethanol pH 7.2) at 30% (w/v). This mixture was sonicated and centrifuged again (13,000 g, 4°C, 10 min) to obtain the protein extract (supernatant). The protein concentration of extracts was measured by Bradford assay to calculate specific activity (Bradford, 1976). The β -galactosidase activity was assayed as described previously (Dyall-Smith, 2009): 25 μ L to 400 μ L of protein extracts were diluted to a final volume of 950 μ L using bgaH buffer and heated at 40°C for 3 min. The volume of the extract was adjusted based on β -galactosidase activity. After this time, 50 μ L ortho-nitro-phenol β -D-galactoside (ONPG) was added at a final concentration of 0.4 μ g/ μ L to the reaction. Finally, the release of o-nitrophenol (ONP) from ONPG was monitored for 5 min following the increase in absorbance at 405 nm. Measurements were done in triplicate. One unit of enzyme activity was defined as the amount of enzyme necessary to form 1 μ mol of ONP per minute under the assay conditions. Specific activity (U/mg of protein) was calculated from the initial slope of the absorbance vs. time plot, using the following equation (ϵ : molar extinction coefficient of ONP):

$$\frac{U}{mg} = \frac{\frac{\Delta A_{405}}{\min} \cdot \text{Reaction vol. (L)} \cdot 10^6}{\epsilon_{ONP} \left(M^{-1} \cdot cm^{-1} \right) \cdot \text{Extract vol. (mL)} \cdot \text{Protein conc.} \left(\frac{mg}{mL} \right)}$$

2.2. Strain used, culture media, and cell growth calculations

Haloferax mediterranei R4 transformed strains were grown in defined media: 27.75 mM glucose, 1 mM Pi (K₂HPO₄/KH₂PO₄), 15 mM NH₄Cl, 0.03 mM FeCl₃, 2.67 M NaCl, 0.16 M MgSO₄·7H₂O, 0.133 M MgCl₂·6H₂O, 53.33 mM KCl, 1.6 mM NaHCO₃, 4.53 mM NaBr, 6.6 mM CaCl₂, 253.8 mM novobiocin and buffered with 0.1 M (3-(N-morpholino) propanesulfonic acid (MOPS) (Rodríguez-Valera et al., 1980; DasSarma et al., 2019). Media were

supplemented with different electrons acceptor depending on the assay: 10 mM KNO₃, 2 mM NaNO₂, and 15 mM dimethyl sulfoxide (DMSO). These concentrations were selected based on previous studies (Oren and Trüper, 1990; Torregrosa-Crespo et al., 2019, 2020b). Cultures were set up at pH 7.3 using a NaOH solution and 42°C for oxic and anoxic conditions. Cellular growth was monitored spectrophotometrically by measuring the OD_{600 nm}. Growth-specific velocity (μ) was calculated by obtaining the slope of the semi-log graph in the exponential growth phase during anaerobiosis. The slope value was the average value for all the growth curves under the same condition, regardless of the strain used. Cultures and measurements were done in triplicate.

2.3. Aerobic growth: O₂, NO₃⁻, NO₂⁻ and promoter activation monitoring

Cell cultures of the different strains used in this study were grown in defined media supplemented with NO₃⁻ and set up on Duran bottles at 42°C on a rotary shaker at 170 rpm agitation rate to ensure high oxygen exchange between media and air chamber. Air chambers of these cultures were continuously renewed when taking samples ensuring the maximum air saturation in the media during the growth. Dissolved oxygen concentration in the culture media was measured using oxygen-dependent luminescence sensor spots (OXSP5, PyroScience). Spots were affixed to the inner side of flasks and were noninvasively connected to an optical oxygen meter (FireStingGO2, Pyroscience) for measurements. Dissolved oxygen concentration and oxygen solubility calculations were based on the parameterized model proposed by Geng and Duan for saltwater and brines (Geng and Duan, 2010). Extracellular NO₃⁻ and NO₂⁻ concentrations were also determined in these cultures following spectrophotometric assays as previously described (Griess, 1879; Snell and Snell, 1944).

For NO₃⁻ monitorization, 1 mL of cell culture was centrifuged at 13,000 g for 2 min. The supernatant was recovered and diluted to 1:100 with water. Subsequently, 20 μ L of HCL 1 N was added to the supernatant and this mixture was incubated at room temperature for 10 min. Finally, the sample was measured spectrophotometrically by measuring the absorbance (Abs) value at 220 nm and 275 nm (this value is used to correct for interferences with organic matter). Measurements were done in triplicate. The NO₃⁻ content is calculated using a standard curve and the following equation:

$$Abs = Abs_{220nm} - 2 \cdot Abs_{275nm}$$

Extracellular NO₂⁻ was measured as follows: 1 mL of cell culture was centrifuged at 13,000 g for 2 min and the supernatant was recovered and diluted by convenience to 1:100 or 1:200 with water. After that, 950 μ L of water were mixed with 50 μ L of the sample. 1 mL of sulphanilamide solution (0.01 g/mL sulphanilamide in 3 N HCL) and 1 mL of N-(1-naphthyl)-ethylenediamine-dihydrochloride (NEDA) solution (0.02 mg/mL NEDA in water) were added to the mixture and incubated at room temperature for 20 min. Finally, the absorbance value at 540 nm of the sample was measured spectrophotometrically and NO₂⁻ content was determined using a standard curve. Measurements were done in triplicate.

2.4. Anaerobic growth with different electron acceptors and promoter activation

Before anaerobic growth, cell cultures were grown aerobically in defined media supplemented with different alternative electron acceptors or without alternative electron acceptors other than O₂ until the OD_{600 nm} reached 0.2. Then, the cultures were transferred to Duran glass bottles with a bromobutyl rubber closure that does not allow air exchange and keeps a small air chamber (< 10% culture volume). Samples of these cultures were extracted by using sterile needles until the cultures reached the stationary phase of growth.

2.5. Regulatory motifs: Search and mutagenesis

The sequences of the promoter regions of the genes encoding the four main enzymes of denitrification (*narGH*, *nirK*, *nor*, and *nosZ*) were obtained from NCBI Genome DataBase (NCBI Genome RefSeq ID: GCF_005406325.1). These promoter regions were analyzed using the free software MEME from the online tool MEME-Suite 5.1.1¹ to find possible common regulatory motifs (Bailey et al., 2015). In addition to this analysis, FIMO software was used to find this motif in the whole *H. mediterranei* R4 genome (NCBI Genome RefSeq ID: GCF_005406325.1), and BEDtools was used to filter the intergenic motifs found using the sub-command *closest* (Quinlan and Hall, 2010; Bailey et al., 2015). Default parameters were set for MEME and FIMO search. In these analyzes, all the spaces between two coding regions were considered promoter regions, ensuring that possible TATA and BRE boxes are inside this region.

One putative regulatory motif shared by *narp*, *nirp*, *norp*, and *nosp* was found and selected for site-directed mutagenesis. Site-directed mutagenesis was carried out using QuikChange II XL Site-Directed Mutagenesis Kit (Agilent) in pVA513 + *narp*, pVA513 + *nirp*, pVA513 + *norp*, and pVA513 + *nosp* plasmids. Table 3 displays the different mutations performed in the conserved motif of the four promoters. The effect of mutations in promoter activity was monitored by measuring β -galactosidase activity after growing the cells anaerobically for 24 h in defined media without an electron acceptor (see section 2.4).

2.6. Promoter activity and cell growth comparisons: Statistical analysis

The identification of significant differences in the maximum specific activity of the different promoters among different culture conditions was performed as follows: Firstly, the variance of the different samples was compared using the Bartlett's test (5 culture conditions and 4 promoters), and the specific activity values were represented in box and whiskers graphs (Supplementary Figure S4). Groups separated on the basis of variance, were analyzed separately by a two-way ANOVA test followed by a *post hoc* Tukey's test to

¹ <http://meme-suite.org/>

TABLE 3 Strains of *H. mediterranei* used in reporter gene assays.

Promoter	Motif	Mutation	Strain
<i>nosp</i>	CGAACATGTTTCG	<u>T</u> GAACATGTTTCG	pVA513 + <i>nosp</i> MUT1
		CGAAC <u>G</u> CGTTTCG	pVA513 + <i>nosp</i> MUT6 + 7
		CGAACATGTTTC <u>A</u>	pVA513 + <i>nosp</i> MUT12
		CGAACATGTTT <u>A</u>	pVA513 + <i>nosp</i> MUT11 + 12
<i>narp</i>	CAACAATATTTCG	<u>T</u> AACAATATTTCG	pVA513 + <i>narp</i> MUT1
		<u>T</u> GACAATATTTCG	pVA513 + <i>narp</i> MUT1 + 2
<i>nirp</i>	CGAATATGTTTCG	<u>T</u> GAATATGTTTCG	pVA513 + <i>nirp</i> MUT1
<i>norp</i>	CGAACATGTTTCG	<u>T</u> GAACATGTTTCG	pVA513 + <i>norp</i> MUT1

Mutated nucleotides are marked in bold and underlined.

perform multiple comparisons between culture conditions and promoters within the same group. Alpha was set at 0.05 for all comparisons.

Calculations for differences in specific activity values among promoters under the same culture condition were carried out using a one-way ANOVA test followed by a Dunnett's multiple comparison test. Alpha was set at 0.05 for all comparisons.

Statistical analysis for the site-directed mutagenesis experiments was performed using a Student's *t*-test comparing the specific activity of the different mutations against the control to look for significant differences (value of $p < 0.05$).

Bartlett's test and box and whiskers graph were carried out using R 4.2.2, the rest of the graphs, ANOVA test, and multiple comparison tests were performed using GraphPad Prism 8 software.

3. Results and discussion

3.1. Denitrification promoters are activated in anaerobiosis under the presence of different electron acceptors

In order to monitor the activity of denitrification promoters in anaerobiosis under the presence of different electron acceptors, the β -galactosidase specific activity was measured, as described previously, from protein extracts of *H. mediterranei* cells transformed with the four plasmids (pVA513 + *narp*, pVA513 + *nirp*, pVA513 + *norp*, and pVA513 + *nosp*), and grown in anoxic conditions in presence of alternative electron acceptor (DMSO, nitrate, and nitrite), and in the absence of electron acceptor. The variance of the maximum β -galactosidase specific activity values was analyzed, showing significant differences according to Bartlett's test (value of $p < 0.0001$). These values were plotted in [Supplementary Figure S4](#), displaying two different groups: Condition with high specific activity values (anaerobic cultures supplemented with nitrate or nitrite) and condition with low specific activity values (anaerobic cultures without electron acceptor, and anaerobic cultures supplemented with DMSO). These two groups presented clear differences between them and comparisons inside each group were analyzed separately ([Supplementary Tables S1–S4](#)).

The condition without an electron acceptor was analyzed under anoxic conditions to explore if the sole absence of oxygen is enough for the activation of denitrification promoters. As expected,

H. mediterranei shows almost no growth under this condition ([Figure 1](#)). Nonetheless, low specific activity has been detected for *narp*, *nirp*, *norp*, and *nosp*, compared with the cultures supplemented with nitrate or nitrite (see below). Furthermore, the *nar* promoter seems to be the least responsive to anaerobiosis, showing almost no activity at 45 h compared with the rest of the promoters ([Figure 1](#); [Supplementary Tables S5, S6](#)). These results indicate that there is probably a transcriptional regulator that triggers the promoter activation of *narp*, *nirp*, *norp*, and *nosp* under micro-oxic conditions, but an inducing molecule may be necessary to produce high activation. In the haloarchaeon *Haloferax volcanii*, this regulator has been proposed to be the NarO transcriptional regulator found upstream of the *nar* operon ([Hattori et al., 2016](#)). However, neither the gene encoding this regulator nor homologs have been found in the *H. mediterranei* genome.

The growth and activation of promoters under the presence of an electron acceptor other than O_2 were first assessed in media supplemented with DMSO ([Figure 2](#)). This sulfur compound can be used by *H. mediterranei* (and many other haloarchaea) for anaerobic respiration due to the action of the DMSO reductase enzyme, which reduces DMSO to dimethyl sulfide (DMS) ([Qi et al., 2015](#)). These compounds, together with dimethylsulphoniopropionate (DMSP) are key components of the marine sulfur cycle, and some studies pointed out that saline environments are associated with high concentrations of them ([Yoch, 2002](#); [Asher et al., 2015](#)). This association derives mainly from the plant and microbial diversity present in saline environments, which are major sources of sulfur compounds such as these ([Yoch, 2002](#)).

The growth of cells in DMSO-supplemented media showed a diauxic form with the typical three phases once cells enter anoxic conditions: lag phase in which cells are adapting to the new condition, exponential phase, and finally a stationary phase, with an estimated growth-specific velocity in the exponential phase of 0.0085 h^{-1} ([Figure 2](#); [Supplementary Figure S5](#); [Supplementary Table S7](#)). Regarding promoter activation ([Figure 2](#)), no significant differences for the maximum specific activities reached were found compared to cultures without electron acceptors, except for the *narp* ([Supplementary Tables S1, S2](#)). This promoter displayed a major activation when DMSO was present in the culture media compared with the cultures without electron acceptors ([Table 4](#); [Supplementary Tables S1, S2](#)). This activation occurred approximately at the beginning of the exponential phase under anoxic conditions. DMSO respiration is also activated under anoxic

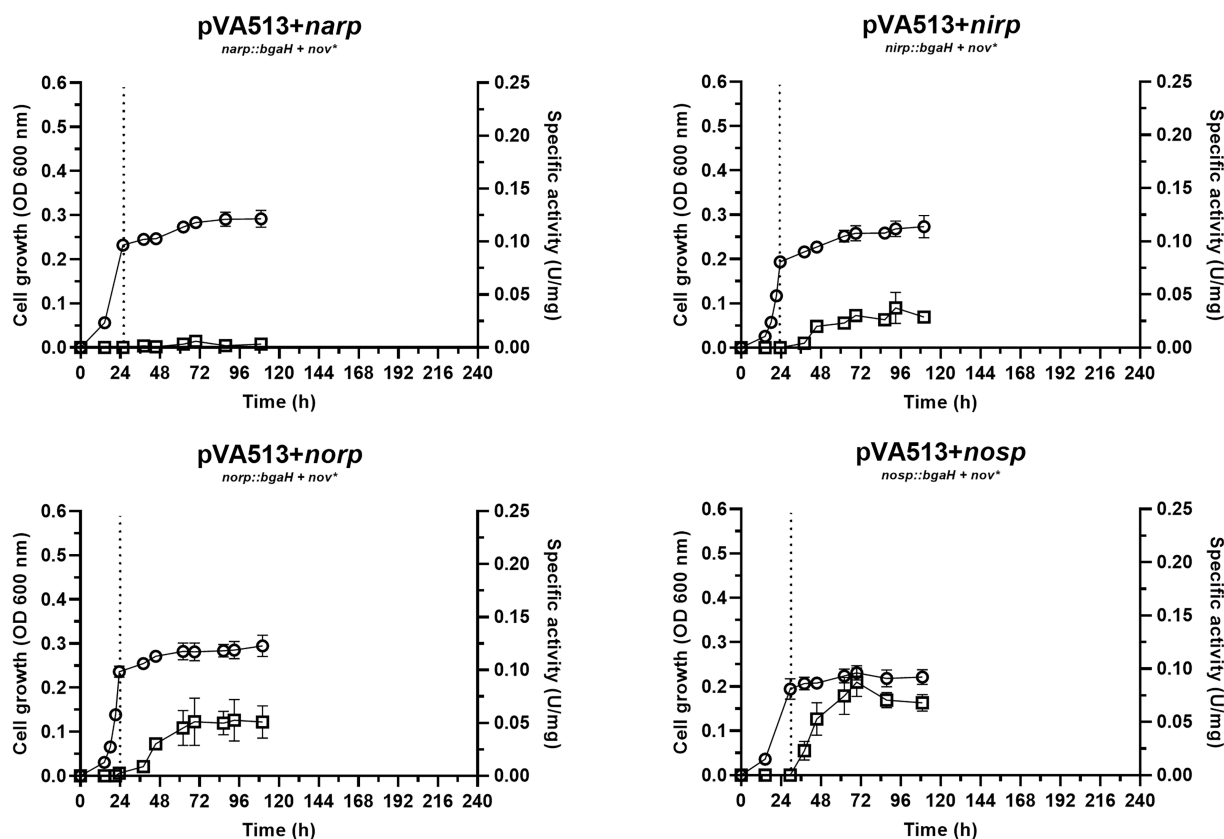


FIGURE 1

Haloferax mediterranei R4 anaerobic growth curves (circles) and promoter activity (squares) of the four main genes of denitrification (*narGH*, *nirK*, *nor*, and *nosp*) without the availability of alternative electron acceptors. Promoter activity was measured using the β -galactosidase specific activity. Dotted lines represent the switch from oxic conditions to anoxic. Titles of each graph represent the specific strain used for that study. pVA513 refers to the plasmid used that carries the reporter gene (*bgaH* gene). *narp* (nitrate reductase promoter), *nirp* (nitrite reductase promoter), *norp* (nitric oxide reductase promoter), and *nosp* (nitrous oxide reductase promoter) refer to the promoter cloned in the plasmid. *Nov*: novobiocin resistance.

conditions in presence of DMSO in *H. volcanii* cells, but, in this case, it is repressed in the presence of nitrate by a shared regulator (NarO), which acts as an activator of *nar* operon and a nitrate dependent repressor of *dms* operon (Koyanagi et al., 2021). In *H. mediterranei*, the behavior observed indicates that transcription of *nar* operon is activated in the presence of DMSO (compared with cultures without electron acceptor) and this activation has not been reported so far (Figure 2; Supplementary Tables S1, S2). Therefore, the reason explaining *narp* activation under this condition is unclear, opening new questions about the connection between DMSO respiration and denitrification.

The next final electron acceptor tested was nitrate (Figure 3). Similar denitrifying conditions (anaerobiosis and presence of nitrate) were analyzed in a previous study in complex media with yeast extract as a nutrient source, but not in defined media (Torregrosa-Crespo et al., 2020a). In the present study, temperature and media composition were adjusted to the optimal growth conditions. As a result of this optimization, the cultures reached an estimated growth-specific velocity of 0.0123 h^{-1} under denitrifying conditions (Supplementary Figure S6; Supplementary Table S7).

In terms of promoter activation, *nirp* and *norp* displayed the highest specific activity (Table 4; Supplementary Figure S4). These data are in line with transcriptomics studies in which *nirK* and *nor*

genes showed higher RNA copy numbers than *nar* and *nosp* (Torregrosa-Crespo et al., 2020a).

Finally, nitrite presence as an alternative electron acceptor was tested under the same anoxic conditions (Figure 4). Denitrification using nitrite as the final electron acceptor does not have the same yield reached in the presence of nitrate, due to both, nitrite toxicity and lower proton motive force connected to ATP production. In terms of bioenergetic, nitrate reduction can move 2H^+ to the pseudo periplasm, which are not translocated when the reduction starts from nitrite (Martinez-Espinosa et al., 2007). Growth data revealed that the estimated specific growth rate of the cells in media supplemented with nitrite was 0.0089 h^{-1} (Supplementary Figure S7; Supplementary Table S7), which is slightly lower than the one shown in the presence of nitrate. It is relevant to notice that there is almost no lag phase when cells are transferred to low oxygen conditions, may be due to faster adaptation of the cells to the anaerobic growth under this situation.

This faster adaptation could be explained by the fact that *narp* is less responsive than *nirp* promoter. Notably, *nirp* has shown significantly higher activation than *narp* promoter under the presence of nitrate and non-significant (value of $p = 0.0599$) but higher activation under the presence of nitrite (Supplementary Figure S4; Supplementary Tables S3, S4). Furthermore, *nirp* activation is also faster (reaching higher specific activity values before *narp*), therefore

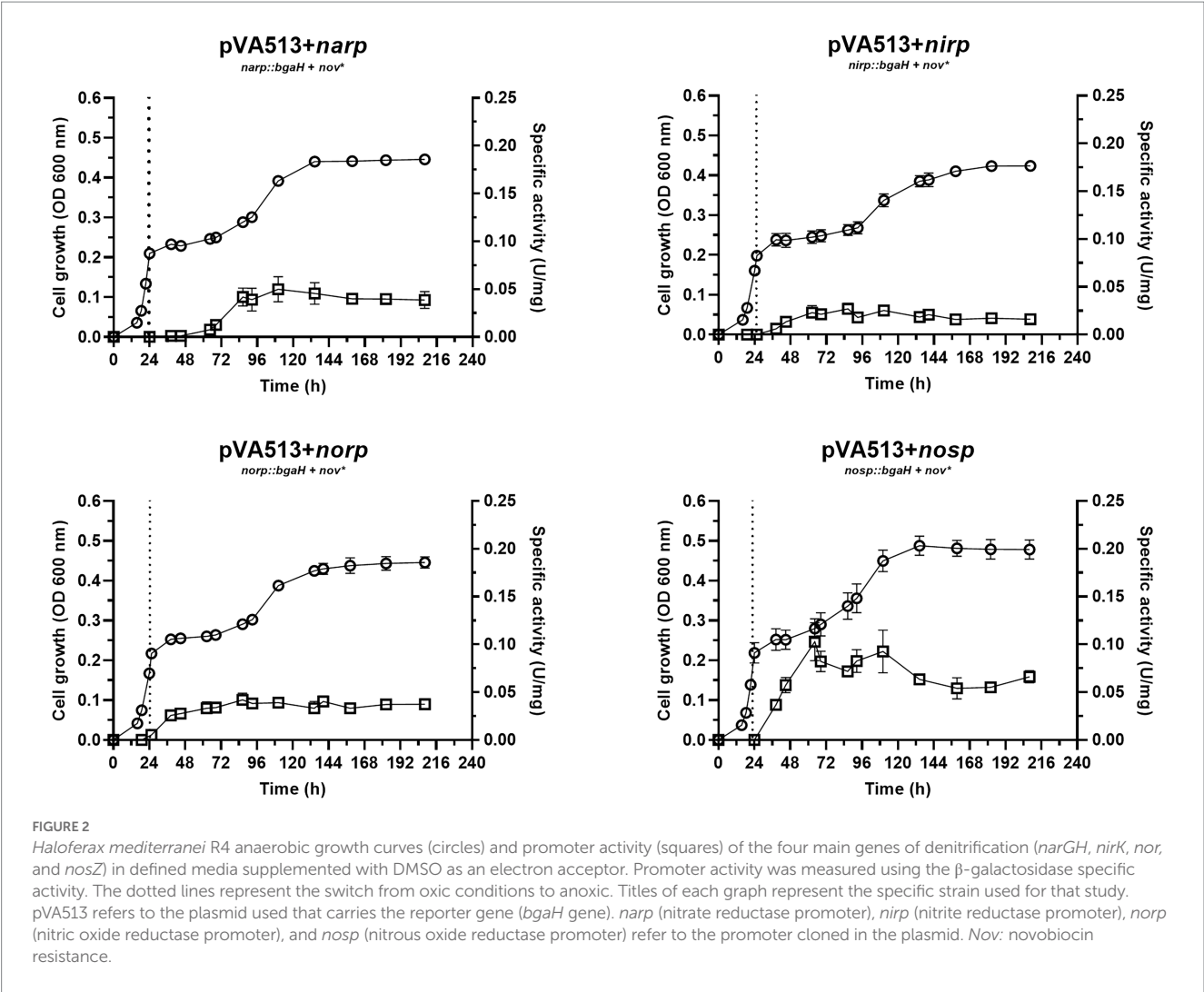


TABLE 4 Maximum specific activity values (U/mg) reached in anaerobic cultures containing different final electron acceptors.

Electron acceptor	Condition	<i>narp</i>	<i>nirp</i>	<i>norp</i>	<i>nosp</i>
No EA*	Anoxic	0.0058 ± 0.0009	0.037 ± 0.015	0.05 ± 0.02	0.087 ± 0.013
DMSO	Anoxic	0.050 ± 0.013	0.027 ± 0.006	0.042 ± 0.007	0.10 ± 0.02
Nitrate	Anoxic	1.0 ± 0.2	3.4 ± 0.8	3.7 ± 0.2	2.2 ± 0.6
	Oxic	0.0345 ± 0.0044	0.0189 ± 0.0073	0.08525 ± 0.0042	0.0657 ± 0.011
Nitrite	Anoxic	1.3 ± 0.3	2.6 ± 0.2	2.5 ± 0.7	3.2 ± 0.6

Narp (nitrate reductase promoter), *nirp* (nitrite reductase promoter), *norp* (nitric oxide reductase promoter), and *nosp* (nitrous oxide reductase promoter) refer to the promoter cloned in the plasmid pVA513. *No electron acceptor other than remaining O₂.

nir gene transcription should occur earlier than *nar* (Figures 1–4; Table 4). Another interesting result obtained in these cultures is that maximum promoter activity is observed around 60 h of cell growth, sooner than in the presence of nitrate or DMSO. This distinctive feature can be explained by exploring the transcriptomic profiling of denitrification genes in *H. mediterranei* performed in previous studies together with gas kinetics (Torregrosa-Crespo et al., 2020a). In those studies, when cells are grown directly using nitrite instead of nitrate, a greater accumulation of nitric oxide is observed (compared with the use of nitrate) (Torregrosa-Crespo et al., 2020a). Nitric oxide is a very

reactive molecule, and its accumulation is toxic to the cells. This toxicity increases transcript levels of denitrification genes to rapidly consume it when this molecule appears (Torregrosa-Crespo et al., 2020a). Hence, this accumulation of nitric oxide produced by the direct consumption of nitrite could lead to higher activation of the promoters, producing the observed peaks, which may be involved in cellular detoxification strategies. Thus, it is known that in bacteria, NO is not only an intermediate of denitrification, but also it is a signaling molecule that can alter gene expression and the same behavior is expected in haloarchaea according to these results (Poole, 2005).

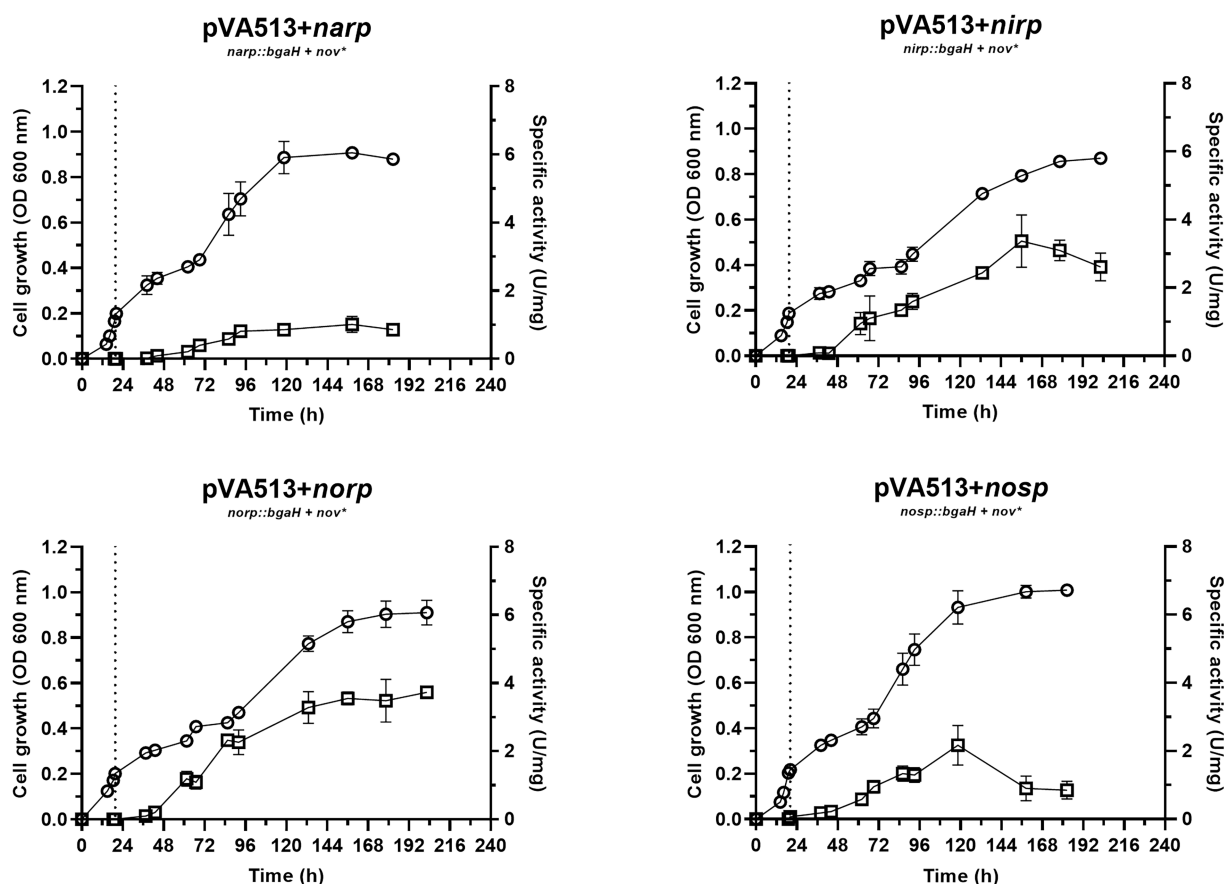


FIGURE 3

Haloflex mediterranei R4 anaerobic growth curves (circles) and promoter activity (squares) of the four main genes of denitrification (*narGH*, *nirK*, *nor*, and *nosZ*) in defined media supplemented with nitrate as an electron acceptor. Promoter activity was measured using the β -galactosidase specific activity. The dotted lines represent the switch from oxic conditions to anoxic. Titles of each graph represent the specific strain used for that study. pVA513 refers to the plasmid used that carries the reporter gene (*bgaH* gene). *narp* (nitrate reductase promoter), *nirp* (nitrite reductase promoter), *norp* (nitric oxide reductase promoter), and *nosp* (nitrous oxide reductase promoter) refer to the promoter cloned in the plasmid. Nov: novobiocin resistance.

Interestingly, another result obtained from this condition was the activation of *narp*. This activation reaches similar levels than in the presence of nitrate, despite NarGH does not have an electron acceptor to reduce under this condition. Due to the similar levels of *narp* activation under nitrite and nitrate conditions, a shared or similar activation could be taking place in both situations. Similar behavior has been described previously in the NarR regulator of *Paracoccus*, which is related to Nar activation and is induced either by nitrate or nitrite (Wood et al., 2001; Spiro, 2016). However, there are no homologs of NarR in *H. mediterranei*. Another explanation is that *narp* could respond to nitrite or nitric oxide and not to nitrate. Therefore, a basal activation of the promoter would be needed to transcript and translate NarGH, triggering the high activation of *narp* when nitrite or NO are present. Figure 1 shows low activation of *narp* when no electron acceptor is present. However, the exact reason for the behavior of this promoter is unknown, but this result together with the pattern observed under the presence of DMSO, show that *narp* has a different regulation than the other denitrification promoters.

Considering all the conditions analyzed, it can be deduced that the absence of oxygen itself is not such a strong signal for the activation of the denitrification response. Previous studies suggested

that at least *narG* and *nosZ* genes were mostly responsive to hypoxia but observing Figures 1–4 it is inferred that the presence of an electron acceptor is crucial for a high activation of the denitrification promoters in *H. mediterranei*, as can be seen by comparing promoter activity in media supplemented with nitrate or nitrite with those in absence of electron acceptor (Table 4; Supplementary Figure S2; Torregrosa-Crespo et al., 2020a).

3.2. Promoter activation is initiated at low oxygen concentration

The presence of denitrification genes in haloarchaeal genomes represents an evolutive adaptation to their environments, where oxygen solubility is very low (Oren, 2016; Miralles-Robledillo et al., 2021). The next step in this study was to analyze whether low oxygen conditions are necessary for the activation of denitrification. Aerobic cultures supplemented with nitrate, the electron acceptor that showed a higher growth-specific velocity value, were set up. These cultures were incubated under continuous agitation, and their air chamber was continuously renewed each time samples were taken, to ensure the

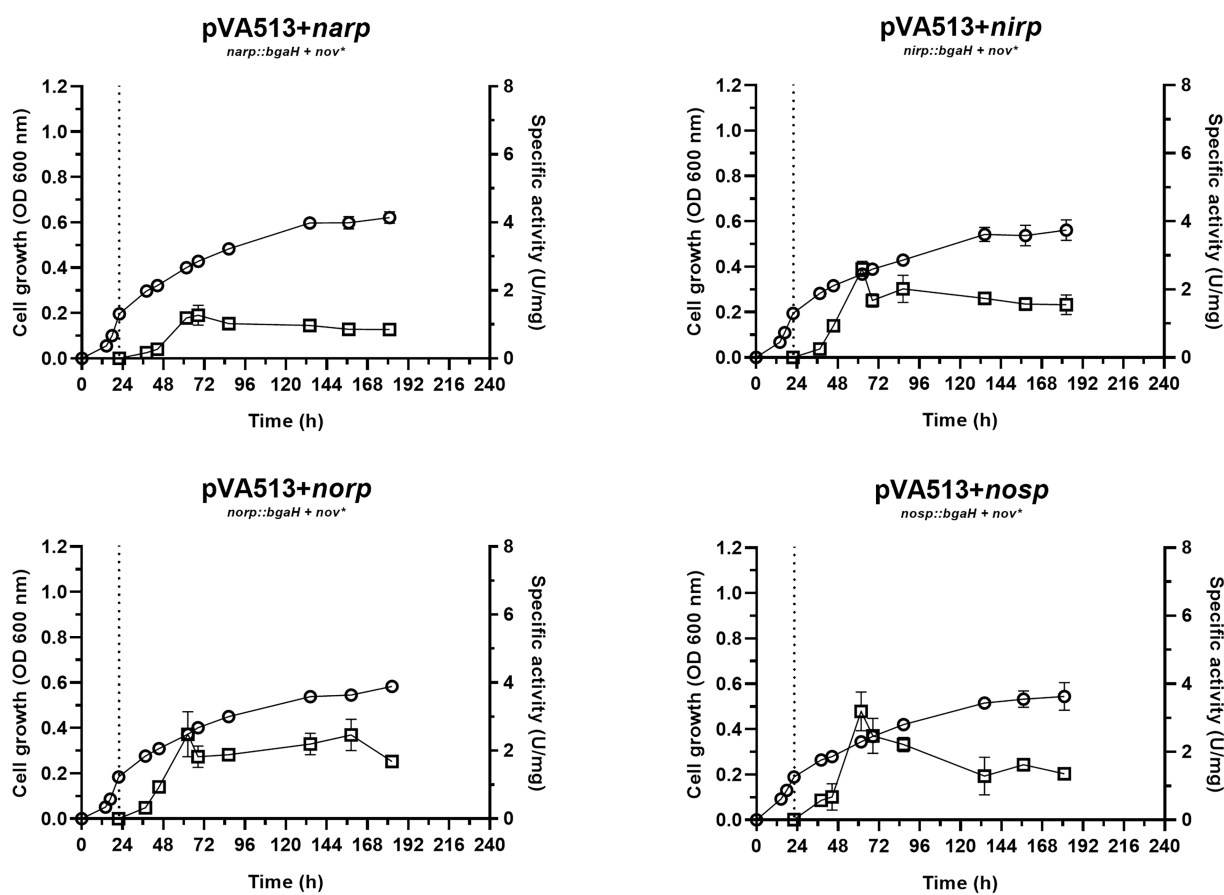


FIGURE 4

Haloferax mediterranei R4 anaerobic growth curves (circles) and promoter activity (squares) of the four main genes of denitrification (*narGH*, *nirK*, *nor*, and *nosZ*) in defined media supplemented with nitrite as an electron acceptor. Promoter activity was measured using the β -galactosidase specific activity. The dotted lines represent the switch from oxic conditions to anoxic. Titles of each graph represent the specific strain used for that study. pVA513 refers to the plasmid used that carries the reporter gene (*bgaH* gene). *narp* (nitrate reductase promoter), *nirp* (nitrite reductase promoter), *norp* (nitric oxide reductase promoter), and *nosp* (nitrous oxide reductase promoter) refer to the promoter cloned in the plasmid. *Nov*: novobiocin resistance.

oxygen exchange with the media. Figure 5 displays the results obtained for each strain.

Reported data on oxygen concentration in hypersaline environments present variations, as dissolved oxygen concentration changes depending on parameters such as salinity, water movement, temperature, or depth at which the sample is taken. Therefore, oxygen concentration values in these environments vary between 2.5 and 250 μM (Rodríguez-Valera et al., 1985; Pavlova et al., 1998; Máthé et al., 2014; Oren, 2016; El-Mezayen et al., 2018). The experimental design here described showed a maximum peak of dissolved oxygen concentration around 360 μM at $t=0$ (this high value is probably given due to the continuous agitation) and displayed a decrease reaching 25–50 μM as minimum values at $\approx 48\text{h}$, as the cell culture grew. This variation is caused by cell growth, as the oxygen demand increases, and because *H. mediterranei* is an exopolysaccharide-producing microorganism. This exopolysaccharide increases the viscosity of the medium contributing to the decrease in oxygen solubility, despite the continuous agitation of the medium (Antón et al., 1988; Boán et al., 1998; Chen et al., 2020). However, this oxygen concentration seems to be sufficient to maintain exponential growth throughout the experiment and to reach high $\text{OD}_{600\text{nm}}$ values compared to cultures under fully anoxic conditions.

Data showed that during the first 27 h of growth, there is enough dissolved oxygen to sustain aerobic respiration without denitrification because there is almost no nitrate consumption and almost no promoter activation. Then, the critical oxygen concentration of $\approx 100\mu\text{M}$ was reached and nitrate consumption by denitrification starts, followed by nitrite production. From this point, the promoter's activity began to increase and nitrate, and oxygen were consumed rapidly as cell density was increasing. Nitrite production reached its maximum value at around 42.5 h, with a maximum concentration of 8.2–8.7 μM . This nitrite was subsequently consumed until cell growth reached the stationary phase. It is also noticeable that the peak of maximum activity of the *norp* appears in coordination with the maximum nitrite values, and therefore maximum production of nitric oxide. Stationary cell growth is reached at $\approx 74\text{h}$, and all measured values were constant due to the decrease in metabolic activity and cell death. Finally, the maximum values of activity reached in this experiment are of the same order as those obtained under anoxic conditions supplemented with DMSO and without an electron acceptor (Table 4). Therefore, under no circumstances these measurements reached the values obtained with nitrate or nitrite in anoxia (Table 4; Supplementary Figure S4).

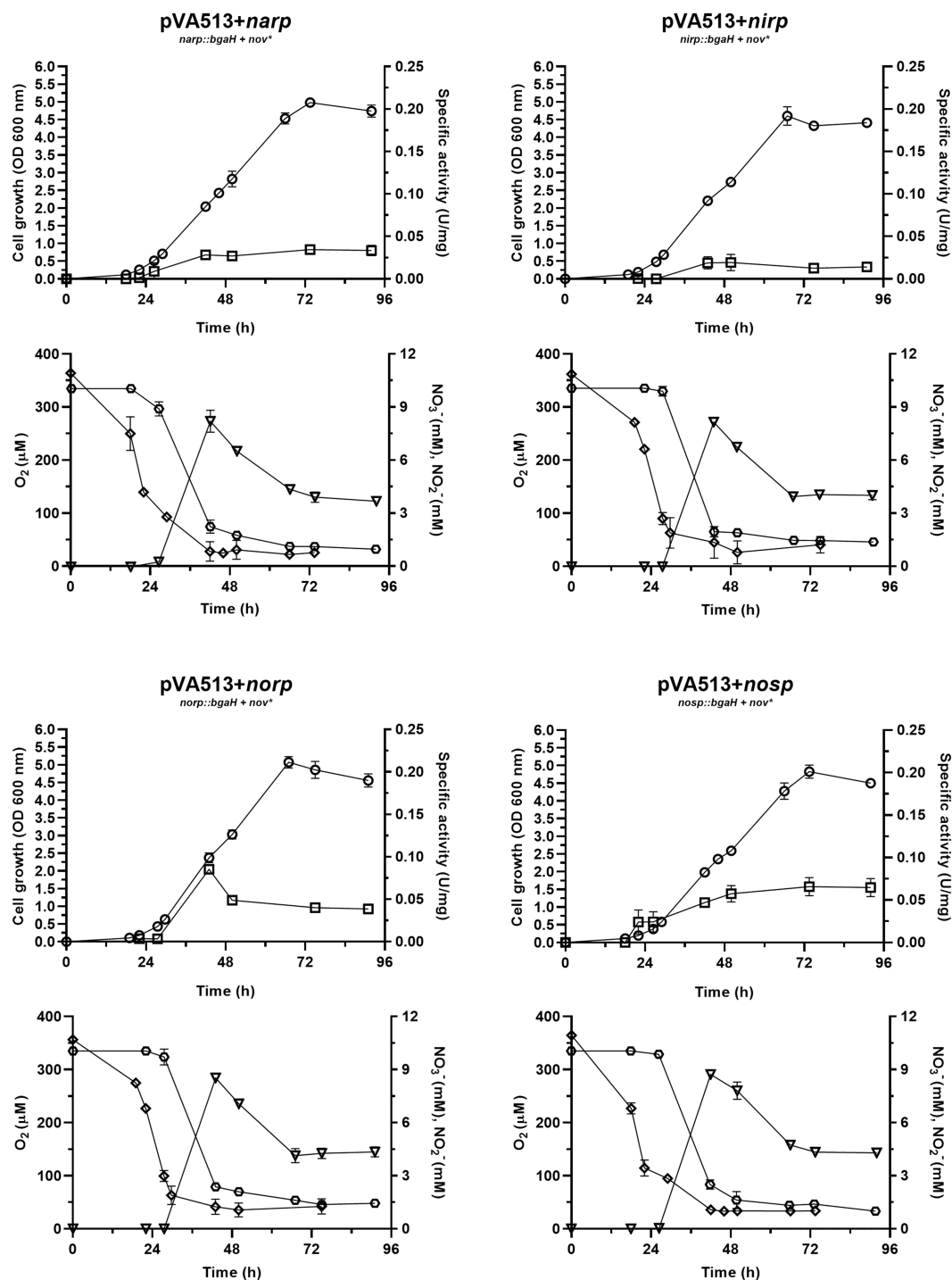


FIGURE 5

Haloferax mediterranei R4 aerobic growth curves (circles) and promoter activity (squares) of the four main genes of denitrification (*narGH*, *nirK*, *nor*, and *nosZ*) in defined media supplemented with nitrate as electron acceptor. Promoter activity was measured using the β -galactosidase specific activity. Below each growth graph, it is displayed the measurements of different parameters of the media: oxygen (diamond), nitrate (hexagon), and nitrite (triangles). Titles of each graph represent the specific strain used for that study. pVA513 refers to the plasmid used that carries the reporter gene (*bgaH* gene). *narp* (nitrate reductase promoter), *nirp* (nitrite reductase promoter), *norp* (nitric oxide reductase promoter), and *nosp* (nitrous oxide reductase promoter) refer to the promoter cloned in the plasmid. Nov: novobiocin resistance.

Considering these data together with those observed under anoxic conditions, it is concluded that the activation of denitrification is mainly due to two factors that need to be combined: low oxygen availability and the presence of electron acceptors. Separately, these

two elements (Figures 1, 5) cannot produce a high activation of the denitrification promoters (compared to the activation reached in anoxic media supplemented with nitrate or nitrite) but combined, the response of the main denitrification promoters is produced and the

growth under denitrifying conditions is supported. The first regulatory model for denitrification was proposed by Torregrosa-Crespo and co-workers in 2020 (Torregrosa-Crespo et al., 2020a); the results here described complement it regarding the regulatory signal network that is controlling denitrification promoters (Torregrosa-Crespo et al., 2020a). An updated version of the denitrification model in haloarchaea is displayed in Figure 6. This version has added new connections between the intermediates of the pathway, such as the activation of both promoters *narp* and *nirp* in the presence of NO_2^- and the demonstration of the activation of *narp* in the presence of NO_3^- . In addition, this model has linked the sulfur cycle with denitrification due to the activation of the *narp* by DMSO and has opened new questions about the possible role of NO in *narp* activation. To summarize, when oxygen concentration drops below $\approx 100 \mu\text{M}$ a small activation of denitrification response is observed if an electron acceptor is available for respiration (Figures 2–5). Under these circumstances oxygen and nitrate are concomitantly consumed, indicating that aerobic respiration should be occurring simultaneously with denitrification supporting an exponential growth, and that denitrification in *H. mediterranei* may not require a dramatic decrease in O_2 concentration (Chen and Strous, 2013; Ji et al., 2015; Yang et al., 2020). However, it is worth mentioning that the maximum activity of the promoter is reached under completely anoxic conditions, as shown in section 3.1. Finally, it is also noticed an activity peak of the *norp* promoter coordinated with nitrite production under this culture condition, probably due to the subsequent formation of nitric oxide.

3.3. A candidate regulatory motif is shared by all promoter regions

Genetic regulation of the denitrification pathway is not well-known in haloarchaea, whereas in bacteria there is a wide range of small-RNAs and transcriptional regulators identified alongside their mechanisms of action (Qu et al., 2016; Gaimster et al., 2018, 2019; Durand and Guillier, 2021). To better understand how promoter activation is produced in haloarchaeal denitrification, a search for common regulatory motives was carried out employing bioinformatic

methods. MEME software identified a shared motif in the four promoter regions studied: CGAAYATDKTYG (Bailey et al., 2015). This motif is semi-palindromic and is especially well conserved in *nirp*, *norp*, and *nosp* promoters, whereas in *narp* has more variations (Figure 7). The identified motif has different positions in each promoter region. In *nirp*, *norp* and *nosp* promoters is located at 79, 66, and 69 nucleotides upstream of the ATG initiation codon (of the gene in the case of *nirK* and *nor* and of the first gene of the operon in the case of *narGH* and *nosZ*) respectively, but in *nar* promoter is found at 319 nucleotides upstream of the initiation codon (Supplementary Figures S2, S3). Therefore, *nirp*, *norp*, and *nosp* promoters not only share a more conserved sequence, but also a similar position of it.

A similar motif has been also found in the *narGH*, *nirK*, *nor*, and *nosZ* promoter regions of other haloarchaea from *Haloferax*, *Halomicrobium*, *Halogeometricum*, *Halorubrum*, *Halorhabdus*, and *Haloarcula* genera (Hattori et al., 2016). However, it should be noted that in *H. mediterranei* this motif has also been found in 83 promoter regions of different genes (Supplementary Table S8). These promoter regions include the promoters of the main denitrification genes (or its operons), and genes encoding some proteins previously reported as accessory candidates involved in haloarchaeal denitrification. These accessory proteins are a multicopper oxidase domain-containing protein (NCBI gene ID: E6P09_RS00725) that could be involved in the electron transfer during nitrite/nitric oxide reduction, a hypothetical protein (NCBI gene ID: E6P09_RS00790) that is hypothesized to act as electron donor or as an oxygen scavenger, and an halocyanin (NCBI gene ID: E6P09_RS00740) which is suggested to play a role in the electron transfer during nitrite and/or nitric oxide reduction (Torregrosa-Crespo et al., 2020b). In addition, it is also noticeable that there are a high number of promoter regions of genes that encode transporters, transcriptional regulators from different families, and a truncated hemoglobin that would be proposed as accessory proteins of the denitrification metabolism. To clarify the role of this semi-palindromic sequence in the promoter regions of the main denitrification genes here studied, a site-directed mutagenesis experiment had been designed (Figures 8A–D). Firstly, the *nosp* motif, which is perfectly semi-palindromic, was mutated at different positions

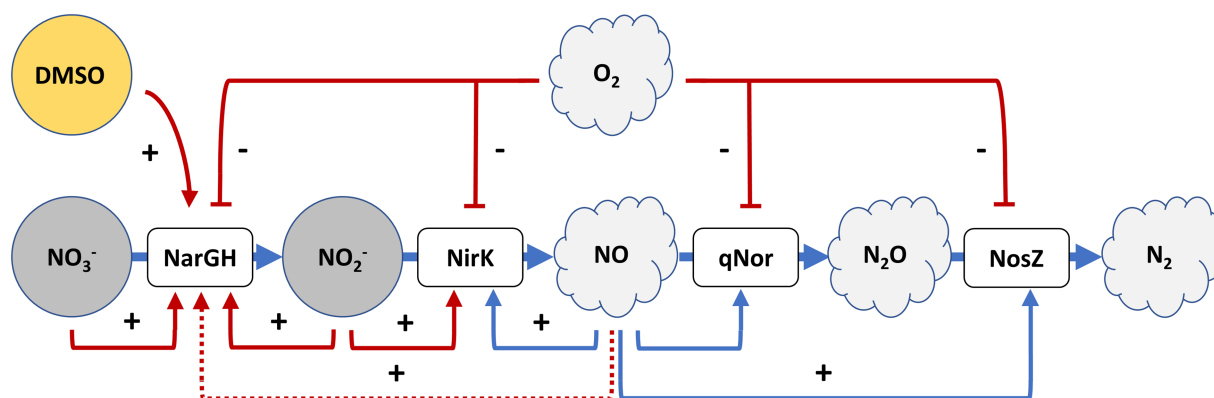


FIGURE 6

A tentative model for denitrification regulatory signals in *H. mediterranei* R4. In red are the signals that have been assessed in this paper, that complement the first proposed model. The activation of respiratory nitrate reductase transcription by NO has been proposed in this paper (dot line), but it should be analyzed with further studies. Adapted from Torregrosa-Crespo et al. (2020a).

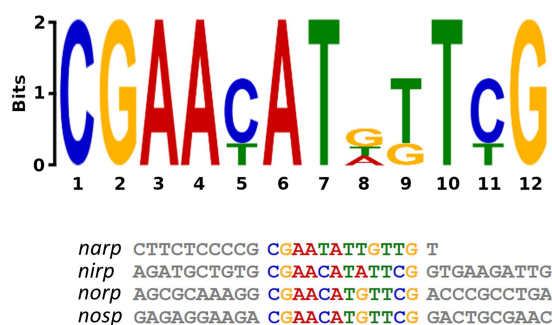


FIGURE 7

Semi-palindromic motif found in the promoter regions of the four main genes of denitrification. *Nosp*: nitrous oxide reductase promoter; *norp*: nitric oxide reductase promoter; *nirp*: nitrite reductase promoter; *nar*: nitrate reductase promoter. In *nirp*, *nor* and *nosp* promoters the motif is located at 79, 66, and 69 nucleotides upstream of the ATG initiation codon (of the gene in the case of *nirK* and *nor* and of the first gene of the operon in the case of *narGH* and *nosZ*) respectively, and in *nar* promoter is found at 319 nucleotides upstream of the initiation codon.

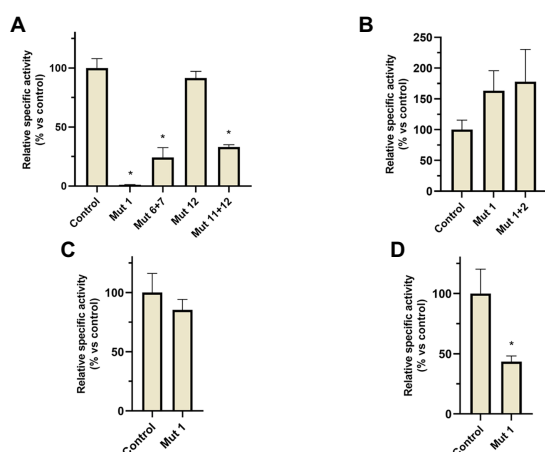


FIGURE 8

β -galactosidase relative specific activity measured in the different *H. mediterranei* R4 mutants analyzed in site-directed mutagenesis experiments (comparisons against non-mutated control strain). Cultures were grown under anaerobic conditions without an electron acceptor and measurements were taken after 24h of anoxic growth. (A) *nos* promoter mutants, (B) *nar* promoter mutants, (C) *nir* promoter mutants, and (D) *nor* promoter mutants. *Conditions with significant differences compared to control (value of $p < 0.05$).

to find if the activity of the promoter is affected by the mutations (Table 3). The results showed that Mut 1 (mutation in the first nucleotide of the semi-palindrome) presented the most drastic change in terms of specific activity in the *nosp* promoter (Figure 8A). Moreover, the other performed mutations in the promoter also have an effect, decreasing the activity significantly, except for mutation 12 (Figure 8A). Subsequently, the rest of the promoters were also analyzed mutating at least the first nucleotide of the palindrome. *Nir* and *nor* promoters displayed a decrease in the activity, but not as drastic as *nos* promoter, and only the *nor* (besides *nosp*) decrease was statistically significant (Figures 8C,D). On the contrary, *nar* showed different behavior as the mutations displayed a nonsignificant increase in the activity of the

promoter (Figure 8B). This difference could be due to the long distance of the consensus sequence to the first gene of the *nar* operon (319 base pairs upstream of the start codon) which indicates that, probably, this sequence is not controlling the expression of the *nar* operon. As well, the motif presents a high variation in the *nar*p (Figure 7).

In fact, in the previously observed data (Section 3.1), *nar*p showed a different activation under different conditions in comparison to the other promoters. This result may suggest that the semi-palindrome found in *nar*p is regulating another genetic element, or that the semi-palindrome has not a regulatory role. Therefore, it is unlikely that the mutation of the semi-palindrome in *nar*p influences promoter activation. *Nar*p probably has a different genetic regulation than the other denitrification promoters. However, *nirp*, *nor*p, and *nosp* promoters show more similarities regarding their activation pattern and in the location and functionality of this semi-palindromic motif (despite the *nirp* mutation was not statistically significant). Thus, these three promoters might share a common regulator.

4. Conclusion

Denitrification in haloarchaea is a poorly understood pathway compared to bacterial denitrification. Nevertheless, the last studies have pointed out that denitrification in hypersaline environments could be a source of nitrogenous gasses to the atmosphere (mainly if partial denitrification predominates vs. complete denitrification), and greater efforts should be made to deepen the knowledge of these extreme environments (Miralles-Robledillo et al., 2021). In this context, this work represents an advance in the understanding of the external signals that can modulate the activation of this alternative respiration. The presence of electron acceptors and low oxygen conditions have been found to be essential signals for the activation of denitrification. Moreover, these conditions must occur together, since individually they are not capable of producing a high activation of any of the promoters of the main enzymes involved in the process. Another important feature observed in the cultures with oxygen is that denitrification begins with oxygen present in the media ($\approx 100 \mu\text{M}$). This data supports the patterns proposed in previous work: induction of denitrification in *H. mediterranei* does not require full anaerobiosis and it is rather a parallel process that can occur at the same time as aerobic respiration if there is low oxygen availability in the environment (Chen and Strous, 2013; Torregrosa-Crespo et al., 2020a). Finally, it has been concluded that at least *nor*p, *nosp* and probably *nir*p could have common transcriptional regulators because of the presence of a well-conserved motif in their sequence and similar promoter activation patterns. The *nar*p promoter also carries a similar motif, but it is not as conserved or in the same position as in the other promoters. Therefore, these data, together with the promoter activity under different conditions, suggest that a distinct regulatory mechanism should control the activity of the *nar*p promoter. In addition, this motif has also been found in the promoter region of other proteins previously suggested as accessory proteins of the denitrification pathway in haloarchaea, indicating that this motif could be the operator of a denitrification regulon.

In summary, the genetic regulation driving denitrification in haloarchaea appears to be like the bacterial counterpart. Negative regulation of denitrification by O_2 and shared regulators for different genes acting as sensors for *N*-species are common in denitrifying bacteria (Spiro, 2016). However, in haloarchaea, these regulators

remain unknown. The next step in this research is to look for key transcriptional regulators that are under this regulation. Therefore, further research using different approaches is needed to find new potential regulators.

Data availability statement

The original contributions presented in the study are included in the article/[Supplementary material](#), further inquiries can be directed to the corresponding author.

Author contributions

CP and RM-E conceived the project and managed the funding. JM-R, CP, and RM-E conceived and designed the study, analyzed the results, and contributed equally to the writing and review of the original manuscript. JM-R conducted the experiments and data analysis. All authors contributed to the article and approved the submitted version.

Funding

This work was funded by MINECO, Spain (2019/00673/001), Generalitat Valenciana, Spain (PROMETEO/2021/055), and VIGROB-309 (University of Alicante).

References

- Antón, J., Meseguer, I., and Rodríguez-Valera, F. (1988). Production of an extracellular polysaccharide by *Haloferax mediterranei*. *Appl. Environ. Microbiol.* 54, 2381–2386. doi: 10.1128/AEM.54.10.2381-2386.1988
- Asher, E. C., Dacey, J. W. H., Jarníková, T., and Tortell, P. D. (2015). Measurement of DMS, DMSO, and DMSP in natural waters by automated sequential chemical analysis. *Limnol. Oceanogr. Methods* 13, 451–462. doi: 10.1002/LOM3.10039
- Asif, A., Mohsin, H., Tanvir, R., and Rehman, Y. (2017). Revisiting the mechanisms involved in calcium chloride induced bacterial transformation. *Front. Microbiol.* 8:2169. doi: 10.3389/FMICB.2017.02169/BIBTEX
- Bailey, T. L., Johnson, J., Grant, C. E., and Noble, W. S. (2015). The MEME suite. *Nucleic Acids Res.* 43, W39–W49. doi: 10.1093/nar/gkv416
- Boán, I. F., García-Quesada, J. C., Antón, J., Rodríguez-Valera, F., and Marcilla, A. (1998). Flow properties of haloarchaeal polysaccharides in aqueous solutions. *Polymer* 39, 6945–6950. doi: 10.1016/S0023-3861(98)00118-9
- Bonete, M. J., Martínez-Espínosa, R. M., Pire, C., Zafrilla, B., and Richardson, D. J. (2008). Nitrogen metabolism in haloarchaea. *Saline Syst.* 4, 9–12. doi: 10.1186/1746-1448-4-9
- Bradford, M. M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* 72, 248–254. doi: 10.1016/0003-2697(76)90527-3
- Cavigelli, M. A., del Grosso, S. J., Liebig, M. A., Snyder, C. S., Fixen, P. E., Venterea, R. T., et al. (2012). US agricultural nitrous oxide emissions: context, status, and trends. *Front. Ecol. Environ.* 10, 537–546. doi: 10.1890/120054
- Chen, J., Mitra, R., Xiang, H., and Han, J. (2020). Deletion of the *pps*-like gene activates the cryptic *phaC* genes in *Haloferax mediterranei*. *Appl. Microbiol. Biotechnol.* 104, 9759–9771. doi: 10.1007/S00253-020-10898-0
- Chen, J., and Strous, M. (2013). Denitrification and aerobic respiration, hybrid electron transport chains and co-evolution. *Biochim. Biophys. Acta Bioenerg.* 1827, 136–144. doi: 10.1016/j.bbabi.2012.10.002
- Conrad, R. (1996). Soil microorganisms as controllers of atmospheric trace gases (H₂, CO, CH₄, OCS, N₂O, and NO). *Microbiol. Mol. Biol. Rev.* 60, 609–640. doi: 10.1128/mr.60.4.609-640.1996
- DasSarma, S., Fomenkov, A., DasSarma, S. L., Vincze, T., DasSarma, P., Roberts, R. J., et al. (2019). Methylomes of two extremely halophilic archaea species, *Haloarcula marismortui* and *Haloferax mediterranei*. *Microbiol. Resour. Announc.* 8:e00577-19. doi: 10.1128/MRA.00577-19
- Delgado-Baquerizo, M., Maestre, F. T., Gallardo, A., Bowker, M. A., Wallenstein, M. D., Quero, J. L., et al. (2013). Decoupling of soil nutrient cycles as a function of aridity in global drylands. *Nature* 502, 672–676. doi: 10.1038/nature12670
- Durand, S., and Guillier, M. (2021). Transcriptional and post-transcriptional control of the nitrate respiration in bacteria. *Front. Mol. Biosci.* 8:667758. doi: 10.3389/FMOLB.2021.667758
- Dyall-Smith, M. L. (2009). *The Haloarchaea - Protocols for Haloarchaeal Genetics*. Available at: <http://www.haloarchaea.com/resources/haloarchaea/index.html>
- Ehhalt, D., Prather, M., Dentener, F., Derwent, R., Dlugokencky, E., Holland, E., et al. (2001). "Atmospheric Chemistry and Greenhouse Gases," in *Climate Change 2001: The Scientific Basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change*. eds. J. T. Houghton, Y. Ding, D. J. Griggs, M. Noguer, P. J. van der Linden, X. Dai (Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press), 241–287.
- El-Mezayen, M. M., Rueda-Roa, D. T., Essa, M. A., Muller-Karger, F. E., and Elghobashy, A. E. (2018). Water quality observations in the marine aquaculture complex of the Deeba triangle, Lake Manzala, Egyptian Mediterranean coast. *Environ. Monit. Assess.* 190, 1–12. doi: 10.1007/S10661-018-6800-6
- Esclapez, J., Bravo-Barrales, G., Bautista, V., Pire, C., Camacho, M., and Bonete, M. J. (2014). Effects of nitrogen sources on the nitrate assimilation in *Haloferax mediterranei*: growth kinetics and transcriptomic analysis. *FEMS Microbiol. Lett.* 350, 168–174. doi: 10.1111/1574-6968.12325
- Esclapez, J., Zafrilla, B., Martínez-Espínosa, R. M., and Bonete, M. J. (2013). Cu-NirK from *Haloferax mediterranei* as an example of metalloprotein maturation and exportation via tat system. *BBA-Proteins Proteomics* 1834, 1003–1009. doi: 10.1016/j.bbapap.2013.03.002
- Feng, J., Li, F., Zhou, X., Xu, C., Ji, L., Chen, Z., et al. (2018). Impact of agronomy practices on the effects of reduced tillage systems on CH₄ and N₂O emissions from agricultural fields: a global meta-analysis. *PLoS One* 13:e0196703. doi: 10.1371/journal.pone.0196703
- Firestone, M. K., Firestone, R. B., and Tiedje, J. M. (1980). Nitrous oxide from soil denitrification: factors controlling its biological production. *Science* 208, 749–751. doi: 10.1126/science.208.4445.749

Acknowledgments

The authors of the manuscript would like to thank María Jose Nueda (University of Alicante) for her support with the statistical analysis.

Conflict of interest

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.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2023.1109550/full#supplementary-material>

- Gaimster, H., Alston, M., Richardson, D. J., Gates, A. J., and Rowley, G. (2018). Transcriptional and environmental control of bacterial denitrification and N₂O emissions. *FEMS Microbiol. Lett.* 365, 1–8. doi: 10.1093/femsle/fnx277
- Gaimster, H., Hews, C. L., Griffiths, R., Soriano-Laguna, M. J., Alston, M., Richardson, D. J., et al. (2019). A central small RNA regulatory circuit controlling bacterial denitrification and N₂O emissions. *MBio* 10, 1–12. doi: 10.1128/mbio.01165-19
- Geng, M., and Duan, Z. (2010). Prediction of oxygen solubility in pure water and brines up to high temperatures and pressures. *Geochim. Cosmochim. Acta* 74, 5631–5640. doi: 10.1016/j.gca.2010.06.034
- Giani, M., Garbayo, I., Vilchez, C., and Martínez-Espinosa, R. M. (2019). Haloarchaeal carotenoids: healthy novel compounds from extreme environments. *Mar. Drugs* 17:524. doi: 10.3390/MD17090524
- Griess, P. (1879). Bemerkungen zu der Abhandlung der HH. Weselsky und Benedikt Ueber einige Azoverbindungen. *Ber. Dtsch. Chem. Ges.* 12, 426–428. doi: 10.1002/CBER.187901201117
- Hattori, T., Shiba, H., Ashiki, K. I., Araki, T., Nagashima, Y. K., Yoshimatsu, K., et al. (2016). Anaerobic growth of haloarchaeon *Haloferax volcanii* by denitrification is controlled by the transcription regulator NarO. *J. Bacteriol.* 198, 1077–1086. doi: 10.1128/JB.00833-15
- Holmes, M. L., and Dyall-Smith, M. L. (2000). Sequence and expression of a halobacterial β -galactosidase gene. *Mol. Microbiol.* 36, 114–122. doi: 10.1046/j.1365-2958.2000.01832.x
- Ji, B., Yang, K., Zhu, L., Jiang, Y., Wang, H., Zhou, J., et al. (2015). Aerobic denitrification: a review of important advances of the last 30 years. *Biotechnol. Bioprocess Eng.* 20, 643–651. doi: 10.1007/S12257-015-0009-0/METRICS
- Kool, D. M., Dolfing, J., Wrage, N., and van Groenigen, J. W. (2011). Nitrifier denitrification as a distinct and significant source of nitrous oxide from soil. *Soil Biol. Biochem.* 43, 174–178. doi: 10.1016/j.soilbio.2010.09.030
- Koyanagi, I., Dohra, H., and Fujiwara, T. (2021). Nitrate-responsive suppression of dimethyl sulfoxide respiration in a facultative anaerobic haloarchaeon, *Haloferax volcanii*. *J. Bacteriol.* 203:e0065520. doi: 10.1128/JB.00655-20
- Lledó, B., Martínez-Espinosa, R. M., Marhuenda-Egea, F. C., and Bonete, M. J. (2004). Respiratory nitrate reductase from haloarchaeon *Haloferax mediterranei*: biochemical and genetic analysis. *Biochim. Biophys. Acta Gen. Subj.* 1674, 50–59. doi: 10.1016/j.bbagen.2004.05.007
- Martínez-Espinosa, R. M., Cole, J. A., Richardson, D. J., and Watmough, N. J. (2011). Enzymology and ecology of the nitrogen cycle. *Biochem. Soc. Trans.* 39, 175–178. doi: 10.1042/BST0390175
- Martínez-Espinosa, R. M., Dridge, E. J., Bonete, M. J., Butt, J. N., Butler, C. S., Sargent, F., et al. (2007). Look on the positive side! The orientation, identification and bioenergetics of “archaeal” membrane-bound nitrate reductases. *FEMS Microbiol. Lett.* 276, 129–139. doi: 10.1111/j.1574-6968.2007.00887.x
- Martínez-Espinosa, R. M., Lledó, B., Marhuenda-Egea, F. C., and Bonete, M. J. (2007). The effect of ammonium on assimilatory nitrate reduction in the haloarchaeon *Haloferax mediterranei*. *Extremophiles* 11, 759–767. doi: 10.1007/s00792-007-0095-9
- Masson-Delmotte, V., Zhai, P., Pörtner, H.-O., Roberts, D., Skea, J., Calvo, E., et al. (2019). Climate change and land an IPCC special report on climate change, desertification, land degradation, sustainable land management, food security, and greenhouse gas fluxes in terrestrial ecosystems. Head of TSU (Operations) IT/Web Manager Senior Administrat. Available at: www.ipcc.ch
- Máthé, I., Borsodi, A. K., Tóth, E. M., Felföldi, T., Jurecska, L., Krett, G., et al. (2014). Vertical physico-chemical gradients with distinct microbial communities in the hypersaline and heliothermal Lake Ursu (Sovata, Romania). *Extremophiles* 18, 501–514. doi: 10.1007/S00792-014-0633-1
- Medhaug, I., Stolpe, M. B., Fischer, E. M., and Knutti, R. (2017). Reconciling controversies about the “global warming hiatus”. *Nature* 545, 41–47. doi: 10.1038/nature22315
- Miralles-Robledillo, J. M., Bernabeu, E., Giani, M., Martínez-Serna, E., Martínez-Espinosa, R. M., and Pire, C. (2021). Distribution of denitrification among haloarchaea: a comprehensive study. *Microorganisms* 9, 1–14. doi: 10.3390/microorganisms9081669
- Muntean, M., Guizzardi, D., Schaaf, E., Crippa, M., Solazzo, E., and Olivier, J. (2019). EDGAR-emissions Database for global atmospheric research. Emissions Database for Global Atmospheric Research. Available at: <https://edgar.jrc.ec.europa.eu/>
- Myhre, G., Shindell, D., Bréon, F., Collins, W., Fuglestad, J., Huang, J., et al. (2013). “Anthropogenic and natural radiative forcing,” in *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. eds. T. F., Stocker, D. Qin, G.-K. Plattner, M. Tignor, S. K. Allen, and J. Boschung (Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press), 659–740.
- Olivier, J. J. G., and Peters, J. A. H. W. (2019). *Trends in global CO2 and total greenhouse gas emissions: 2019 report*. The Hague.
- Oren, A. (2002). Molecular ecology of extremely halophilic archaea and bacteria. *FEMS Microbiol. Ecol.* 39, 1–7. doi: 10.1016/S0168-6496(01)00200-8
- Oren, A. (2016). Probing Saltern brines with an oxygen electrode: what can we learn about the community metabolism in hypersaline systems? *Life* 6:23. doi: 10.3390/LIFE602023
- Oren, A. (2017). Why isn't *Haloferax mediterranei* more “weed-like”? *FEMS Microbiol. Lett.* 364, 1–2. doi: 10.1093/femsle/fnx070
- Oren, A., and Trüper, H. G. (1990). Anaerobic growth of halophilic archaeobacteria by reduction of dimethylsulfoxide and trimethylamine N -oxide. *FEMS Microbiol. Lett.* 70, 33–36. doi: 10.1111/j.1574-6968.1990.tb03772.x
- Pavlova, P., Markova, K., Tanev, S., and Davis, J. S. (1998). Observations on a solar saltworks near Burgas, Bulgaria. *Int. J. Salt Lake Res.* 7, 357–368. doi: 10.1007/BF02442144
- Poole, R. K. (2005). Nitric oxide and nitrosative stress tolerance in bacteria. *Biochem. Soc. Trans.* 33, 176–180. doi: 10.1042/BST0330176
- Qi, Q., Ito, Y., Yoshimatsu, K., and Fujiwara, T. (2015). Transcriptional regulation of dimethyl sulfoxide respiration in a haloarchaeon, *Haloferax volcanii*. *Extremophiles* 20, 27–36. doi: 10.1007/S00792-015-0794-6
- Qu, Z., Bakken, L. R., Molstad, L., Frostegård, Å., and Bergaust, L. L. (2016). Transcriptional and metabolic regulation of denitrification in *Paracoccus denitrificans* allows low but significant activity of nitrous oxide reductase under oxic conditions. *Environ. Microbiol.* 18, 2951–2963. doi: 10.1111/1462-2920.13128
- 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
- Richardson, D., Felgate, H., Watmough, N., Thomson, A., and Baggs, E. (2009). Mitigating release of the potent greenhouse gas N₂O from the nitrogen cycle-could enzymic regulation hold the key? *Trends Biotechnol.* 27, 388–397. doi: 10.1016/j.tibtech.2009.03.009
- Rodríguez-Valera, F., Ruiz-Berraquero, F., and Ramos-Cormenzana, A. (1980). Behaviour of mixed populations of halophilic bacteria in continuous cultures. *Can. J. Microbiol.* 26, 1259–1263. doi: 10.1139/m80-210
- Rodríguez-Valera, F., Ventosa, A., Juez, G., and Imhoff, J. F. (1985). Variation of environmental features and microbial populations with salt concentrations in a multi-pond saltern. *Microb. Ecol.* 11, 107–115. doi: 10.1007/BF02010483
- Schuler, M. S., Cañedo-Argüelles, M., Hintz, W. D., Dyack, B., Birk, S., and Relyea, R. A. (2019). Regulations are needed to protect freshwater ecosystems from salinization. *Philos. Trans. R. Soc. Lond. B: Biol. Sci.* 374:20180019. doi: 10.1098/rstb.2018.0019
- Shcherbak, I., Millar, N., and Robertson, G. P. (2014). Global metaanalysis of the nonlinear response of soil nitrous oxide (N₂O) emissions to fertilizer nitrogen. *Proc. Natl. Acad. Sci. U. S. A.* 111, 9199–9204. doi: 10.1073/pnas.1322434111
- Snell, F. D., and Snell, C. T. (1944). *Colorimetric Methods of Analysis*. 2nd ed. New York: D. Van Nostrand Co.
- Spiro, S. (2016). “Regulation of denitrification,” in *RSC Metalobiology*. eds. I. Moura, J. J. G. Moura, S. R. Pauleta, and L. B. Maia (Cambridge, United Kingdom: Royal Society of Chemistry), 312–330.
- Torregrosa-Crespo, J., Bergaust, L., Pire, C., and Martínez-Espinosa, R. M. (2018). Denitrifying haloarchaea: sources and sinks of nitrogenous gases. *FEMS Microbiol. Lett.* 365, 1–6. doi: 10.1093/femsle/fnx270
- Torregrosa-Crespo, J., González-Torres, P., Bautista, V., Esclapez, J. M., Pire, C., Camacho, M., et al. (2017). Analysis of multiple haloarchaeal genomes suggests that the quinone-dependent respiratory nitric oxide reductase is an important source of nitrous oxide in hypersaline environments. *Environ. Microbiol. Rep.* 9, 788–796. doi: 10.1111/1758-2229.12596
- Torregrosa-Crespo, J., Martínez-Espinosa, R. M., Esclapez, J., Bautista, V., Pire, C., Camacho, M., et al. (2016). Anaerobic metabolism in *Haloferax* genus: denitrification as case of study. *Adv. Microb. Physiol.*, 68, 41–85. doi: 10.1016/bs.ampbs.2016.02.001
- Torregrosa-Crespo, J., Pire, C., Bergaust, L., and Martínez-Espinosa, R. M. (2020a). *Haloferax mediterranei*, an archaeal model for denitrification in saline systems, characterized through integrated physiological and transcriptional analyses. *Front. Microbiol.* 11, 1–13. doi: 10.3389/fmicb.2020.00768
- Torregrosa-Crespo, J., Pire, C., Martínez-Espinosa, R. M., and Bergaust, L. (2019). Denitrifying haloarchaea within the genus *Haloferax* display divergent respiratory phenotypes, with implications for their release of nitrogenous gases. *Environ. Microbiol.* 21, 427–436. doi: 10.1111/1462-2920.14474
- Torregrosa-Crespo, J., Pire, C., Richardson, D. J., and Martínez-Espinosa, R. M. (2020b). Exploring the molecular machinery of denitrification in *Haloferax mediterranei* through proteomics. *Front. Microbiol.* 11:605859. doi: 10.3389/fmicb.2020.605859
- Wang, J., Chadwick, D. R., Cheng, Y., and Yan, X. (2018). Global analysis of agricultural soil denitrification in response to fertilizer nitrogen. *Sci. Total Environ.* 616–617, 908–917. doi: 10.1016/j.scitotenv.2017.10.229
- Wood, N. J., Alizadeh, T., Bennett, S., Pearce, J., Ferguson, S. J., Richardson, D. J., et al. (2001). Maximal expression of membrane-bound nitrate reductase in *Paracoccus* is

induced by nitrate via a third FNR-like regulator named NarR. *J. Bacteriol.* 183, 3606–3613. doi: 10.1128/JB.183.12.3606-3613.2001

Yang, J., Feng, L., Pi, S., Cui, D., Ma, F., Zhao, H. P., et al. (2020). A critical review of aerobic denitrification: insights into the intracellular electron transfer. *Sci. Total Environ.* 731:139080. doi: 10.1016/j.SCITOTENV.2020.139080

Yoch, D. C. (2002). Dimethylsulfoniopropionate: its sources, role in the marine food web, and biological degradation to Dimethylsulfide. *Appl. Environ. Microbiol.* 68, 5804–5815. doi: 10.1128/AEM.68.12.5804-5815.2002

Zhang, Y., Mu, Y., Zhou, Y., Tian, D., Liu, J., and Zhang, C. (2016). NO and N₂O emissions from agricultural fields in the North China plain: origination and mitigation. *Sci. Total Environ.* 551-552, 197–204. doi: 10.1016/j.SCITOTENV.2016.01.209



OPEN ACCESS

EDITED BY

Marleen van Wolferen,
University of Freiburg,
Germany

REVIEWED BY

Changyi Zhang,
University of Illinois Urbana-Champaign,
United States
Jörg Soppa,
Goethe University Frankfurt,
Germany

*CORRESPONDENCE

Susanne Erdmann
✉ serdmann@mpi-bremen.de

SPECIALTY SECTION

This article was submitted to
Biology of Archaea,
a section of the journal
Frontiers in Microbiology

RECEIVED 11 November 2022

ACCEPTED 13 March 2023

PUBLISHED 31 March 2023

CITATION

Gebhard LJ, Duggin IG and Erdmann S (2023)
Improving the genetic system for *Halorubrum*
lacusprofundi to allow in-frame deletions.
Front. Microbiol. 14:1095621.
doi: 10.3389/fmicb.2023.1095621

COPYRIGHT

© 2023 Gebhard, Duggin and Erdmann. 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.

Improving the genetic system for *Halorubrum lacusprofundi* to allow in-frame deletions

L. Johanna Gebhard¹, Iain G. Duggin² and Susanne Erdmann^{1*}

¹Archaeal Virology, Max Planck Institute for Marine Microbiology, Bremen, Germany, ²The Australian Institute for Microbiology and Infection, University of Technology Sydney, Sydney, NSW, Australia

Halorubrum lacusprofundi is a cold-adapted halophilic archaeon isolated from Deep Lake, Antarctica. *Hrr. lacusprofundi* is commonly used to study adaptation to cold environments and thereby a potential source for biotechnological products. Additionally, in contrast to other haloarchaeal model organisms, *Hrr. lacusprofundi* is also susceptible to a range of different viruses and virus-like elements, making it a great model to study virus-host interactions in a cold-adapted organism. A genetic system has previously been reported for *Hrr. lacusprofundi*; however, it does not allow in-frame deletions and multiple gene knockouts. Here, we report the successful generation of uracil auxotrophic (*pyrE2*) mutants of two strains of *Hrr. lacusprofundi*. Subsequently, we attempted to generate knockout mutants using the auxotrophic marker for selection. However, surprisingly, only the combination of the auxotrophic marker and antibiotic selection allowed the timely and clean in-frame deletion of a target gene. Finally, we show that vectors established for the model organism *Haloferax volcanii* are deployable for genetic manipulation of *Hrr. lacusprofundi*, allowing the use of the portfolio of genetic tools available for *H. volcanii* in *Hrr. lacusprofundi*.

KEYWORDS

Halorubrum lacusprofundi, archaea, genetic system, cold adaptation, haloarchaea, auxotrophic mutant

Introduction

Halophilic archaea (haloarchaea) are established as commonly used model systems to study archaeal cell biology and virus-host interactions in archaea (Luk et al., 2014; van Wolferen et al., 2022). Additionally, they are of great interest for their potential use in biotechnological applications (Litchfield, 2011; Javier et al., 2017; Singh and Singh, 2017; Correa and Abreu, 2020; Li et al., 2022). They are easy to culture and genetic systems have already been developed for some members, with *Haloferax volcanii* currently being the most studied model organism (Hartman et al., 2010; Pohlschroder and Schulze, 2019). However, only one virus has been isolated for *H. volcanii* so far (Alarcón-Schumacher et al., 2022), while the majority of hosts for haloarchaeal viruses remain genetically inaccessible. In contrast to *Haloferax* species, that have been shown to be rather unsuspensible to virus infection (Aguirre Sourrouille et al., 2022), a great number of haloarchaeal viruses have been isolated on *Halorubrum* species (Atanasova et al., 2015; Pietilä et al., 2016; Liu et al., 2021). *Halorubrum lacusprofundi* ATCC 49239 (ACAM34) (Franzmann et al., 1988; McGenity and Grant, 1995), a model organism to study cold adaptation in archaea (DasSarma et al., 2013, 2017; Williams et al., 2017), has previously been shown to be genetically accessible (Liao et al., 2016). Several different viruses were shown

to infect *Hrr. lacusprofundi* (Nuttall and Dyall-Smith, 1993; Dyall-Smith, 2009d; Krupović et al., 2010; Atanasova et al., 2012; Porter et al., 2013; Li et al., 2014; Tschitschko et al., 2015; Alarcón-Schumacher et al., 2022; Mercier et al., 2022). Additionally, the formation of plasmid vesicles (PVs), proposed to be an evolutionary precursor of viruses, has only been reported for *Hrr. lacusprofundi* (Erdmann et al., 2017). This makes *Hrr. lacusprofundi* a great model for studying virus-host relationships in archaea, in particular under low temperature, and to investigate the origin and evolution of viruses. A stable genetic system would allow for more in depth studies.

A genetic system, based on an auxotrophic selection marker, that allows in-frame deletions using the pop-in pop-out method has been developed in *H. volcanii* on the basis of the orotate phosphoribosyltransferase gene (*pyrE2*) (Lam and Doolittle, 1992; Wendoloski et al., 2001; Bitan-Banin et al., 2003; Allers et al., 2004; Liao et al., 2021). The Δ *pyrE2* deletion in *H. volcanii* was generated with a non-replicative plasmid containing a non-functional gene construct, created by fusing upstream and downstream flanking regions of the wild-type Δ *pyrE2* gene, and selection with the halobacterial novobiocin resistance gene (*gyrB*) (Bitan-Banin et al., 2003). Pop-in refers to the homologous recombination of plasmid and host chromosome caused by selection pressure applied with novobiocin in the culture media. Excision of the plasmid (pop-out) can either result in the wild-type or the mutated gene remaining in the chromosome; therefore, a fraction of the resulting clones is expected to be mutants with a non-functional target gene. The pop-out can either be initiated by removing the original selection pressure (e.g., novobiocin) or with an additional selection pressure using 5-fluoroorotic acid (5-FOA) and uracil (Bitan-Banin et al., 2003). The practicality of 5-FOA as counter-selectable marker was first described in yeast (Boeke et al., 1984) but has since been widely used for both bacteria and archaea (Liu et al., 2011; Redder and Linder, 2012; Liao et al., 2021; Matsuda et al., 2022; Piatek et al., 2022). In prokaryotes, orotate phosphoribosyltransferase (*pyrE*) and orotidine 5-phosphate decarboxylase (*pyrF*) catalyze the final steps in the pyrimidine biosynthesis pathway. The two enzymes convert 5-FOA to the toxic 5-fluoro-UMP leading to cell death (Redder and Linder, 2012).

However, generating spontaneous *pyrE2* mutants by subjecting the cells to 5-FOA and uracil, has not been successful for *Hrr. lacusprofundi* so far (Liao et al., 2016). Liao et al. (2016) described a method for knocking out the acetamidase *amd3* in *Hrr. lacusprofundi* with a suicide plasmid containing *hmgA* (pJWID1), conferring resistance to pravastatin. The 3-hydroxy-3-methylglutaryl coenzyme A reductase gene (*hmgA*) is involved in the isoprenoid lipid synthesis pathway in archaea. Overexpression of *hmgA* leads to mevinolin, fluvastatin, simvastatin, and pravastatin resistance in haloarchaea (Blaseio and Pfeifer, 1990; Lam and Doolittle, 1992; Wendoloski et al., 2001). However, in practice, pravastatin has proven to be the most reliable selection marker for *Hrr. lacusprofundi* (Liao et al., 2016). Liao et al. (2016) were able to show that the *hmgA* gene was inserted between flanking regions of *amd3*. After double recombination, *hmgA* replaced the target gene on the genome resulting in a successful deletion of the target gene. However, the *hmgA* gene remains in the genome and can interfere with the expression of genes in the same operon. The pravastatin resistance conferred by the permanent insertion into the genome also makes it nearly impossible to use the strain for the knockout of additional genes except when using very high pravastatin concentrations, e.g., 150 µg/mL (Liao et al., 2016).

However, consecutive culturing of *Hrr. lacusprofundi* in pravastatin concentrations of 10 µg/mL and low concentrations of mevinolin and simvastatin led to multiple insertions and duplications of the *hmgA* gene into all three chromosomes of the strain (Hamm et al., 2019). Additionally, *Hrr. lacusprofundi* transposases also disrupted genes of the vector used, rendering them non-functional.

The aim of the study was to develop a stable system for the genetic manipulation of *Hrr. lacusprofundi* ACAM34 that uses the pop-in pop-out method and allows for in-frame deletions without leaving the antibiotic resistance gene in the genome. We successfully generated Δ *pyrE2* mutants for two *Hrr. lacusprofundi* ACAM34 strains. Subsequently, we demonstrate that the auxotrophic mutants can be used for in-frame deletions of genes using vectors that have been designed for *H. volcanii*, eliminating the necessity to design entirely new vectors.

Methods

Strains and cultivation conditions

Unless otherwise specified, strains that are referred to only by their strain name are all strains of *Halorubrum lacusprofundi* ACAM34 (ATCC 49239). *Hrr. lacusprofundi* ACAM34_DSMZ was obtained from the DSMZ (German Collection of Microorganisms and Cell Cultures). *Hrr. lacusprofundi* ACAM34_UNSW is a variant with a reduced genome, derived from an original ACAM34 strain, and has been described in Mercier et al. (2022). The standard medium for cultivation was DBCM2 (Dyall-Smith, 2009b), with slight modifications. Standard DBCM2 medium contained 180 g/L NaCl, 25 g/L MgCl, 29 g/L MgSO₄·7H₂O, 5.8 g/L KCl (all Merck, Germany) as well as 0.3 g/L peptone and 0.05 g/L yeast extract (both by Oxoid, Thermo Fisher Scientific, United States), dissolved in ddH₂O, and adjusted to pH 7.5 with 1 M Tris-HCl pH 8.0 prior to autoclaving. After autoclaving, supplements were added 6 mL/L 1 M CaCl₂, 2 mL/L K₂HPO₄ buffer (Dyall-Smith, 2009b), 4.4 mL/L 25% sodium pyruvate, 5 mL/L 1 M NH₄Cl, 1 mL/L SL10 trace elements solution (Dyall-Smith, 2009c), and 3 mL/L vitamin 10 solution (Dyall-Smith, 2009c). Initially, agar plates were prepared with ddH₂O and 18 g/L Oxoid™ agar. However, tap water was found to be best suited to support growth of ACAM34 on agar plates. Starting with the plating of pTA131_ *hmgA*_ Δ *trpA* transformants from liquid onto agar plates, the components for agar plates were dissolved in tap water and supplemented with 18 g/L Difco™ agar (BD, United States). We are aware that tap water is prone to contaminations and changes in composition, and indeed, we observed slight growth of auxotrophic mutants on selection plates in some seasons of the year. However, using tap water allowed for growth of ACAM34 colonies on plates in a reasonable time, while ddH₂O often did not result in colonies. For the rich medium DBCM2+, peptone and yeast extract concentrations were adjusted to 1 g/L and 0.5 g/L, respectively. Medium DBCM2 contained 5 g/L casamino acids (Oxoid) instead. For the generation of ACAM34_UNSW Δ *pyrE2*, the strain was plated on Hv-Cab solid agar as described in de Silva et al. (2021). Liquid cultures were inoculated at 28°C, because biofilm formation was often observed at 37°C (published optimal growth temperature) in liquid; however, plates were incubated at 37°C leading to a timely (7–10 days) formation of colonies. For light microscopy images, cells in mid-exponential

growth were fixed for 1 h at room temperature (RT) with 1% glutaraldehyde, immobilized on 4% agar slides, and imaged on a Zeiss AxioPhot microscope with AxioCam MRm.

Escherichia coli strains (DH5alpha and C2925) were grown at 37°C, either in liquid lysogeny broth (LB) medium (10 g/L tryptone, 5 g/L yeast extract, 10 g/L NaCl₂) (AppliChem, Germany and by Merck) in ddH₂O at pH 7.0 or on LB plates with 15 g/L Bacto™ agar (BD). Ampicillin was used as a selection agent at a final concentration of 100 µg/mL.

Molecular cloning procedures

DNA extractions were performed with the Bioline Isolate II Genomic DNA Kit (Meridian Bioscience, Cincinnati, United States). Plasmid extractions were performed using the Bioline Isolate II Plasmid Mini Kit, and DNA purification (from PCR or gel electrophoresis) was performed using the Bioline Isolate II PCR and Gel Clean-Up Kit. All extractions were performed following the manufacturer's instructions. Plasmids used in this study are summarized in [Supplementary Table 1](#). The primers are listed with their respective annealing temperatures and targets in [Supplementary Table 2](#). Amplification of archaeal genes was performed with Q5 Polymerase (NEB) as follows: Approximately 100 ng of template DNA was added to PCR master mix (per 100 µL; 1x reaction buffer (NEB, Ipswich, United States), 1x high GC enhancer (NEB), 40 µM dNTPS (NEB), 100 µM of forward and reverse primer each ([biomers.net](#), Ulm, Germany), and 1 µL Q5 polymerase (NEB) and PCR was performed with the following program: Initial denaturation at 96°C for 1 min, 30x cycles at denaturation 96°C for 30 s, annealing (temperature in [Supplementary Table 2](#)) for 30 s, elongation at 72°C (1 min/kb), a final elongation at 72°C for 10 min, and storage at 12°C. When generating PCR fragments for cloning (with restriction sites included), a two-step PCR was performed with 10 cycles using the annealing temperature for the part of the primer that binds the template and 30 cycles with the annealing temperature calculated for the entire length of the primer, including additional bases (e.g., the restriction site). Overlapping PCRs were performed with Phusion Polymerase (NEB) according to the manufacturer's instructions with 5% DMSO in the reaction in two steps: 24 cycles without primers and 35 cycles with fresh dNTPs, Phusion Polymerase, and the outermost primers added to the PCR mix (forward primer used to generate the upstream fragment and reverse primer used to generate the downstream fragment). Restriction digest was performed in 30 µL reactions with 1 µL enzyme (as specified for individual cloning reactions), 3 µL 10x reaction buffer, and 1–3 µg of DNA for 2 h at 37°C with an inactivation step of 20 min at 70°C. Ligation reactions were performed in 20 µL volume with 1 µL T4 Ligase (NEB), 2 µL 10x reaction buffer approximately 100 ng of vector and 300 ng of insert for 2 h at RT.

Ligated constructs were transformed into *E. coli* DH5alpha, *via* heat-shock for 1 min at 42°C and plated on LB medium agar plates amended with 100 µg/mL ampicillin, followed by transformation into *E. coli* C2925 (NEB) for demethylation ([Liao et al., 2021](#)). Both *E. coli* and *Hrr. lacusprofundi* colonies were screened *via* colony PCR after transformation. Scratched cell material from each colony was spread on a fresh agar plate and the remaining cell material on the sterile pipet tip was either directly swirled in Q5 PCR master mix aliquots

(for *E. coli*) or in 10 µL of sterile ddH₂O (for *Hrr. lacusprofundi*), from which 1 µL was used per colony as the PCR template as described above. Plasmid constructs were verified by Sanger sequencing in-house on the ABI 3130xl Sequence Analyzer (Applied Biosystems, Foster City, United States).

Plasmid construction

The PCR construct to attempt knockout of the *pyrE2* gene by homologous recombination in *Hrr. lacusprofundi* was generated by PCR amplification of the upstream fragment using Pyr_USF and Pyr_USR, the downstream fragment using Pyr_DSf and Pyr_DSR and subsequent fusion of the two fragments by overlapping PCR (as described previously, e.g., [Dattani et al., 2022](#); [Supplementary Table 2](#)). The PCR construct was either used directly in a transformation or cloned into the pTA131 vector ([Allers et al., 2004](#)) using *Hind*III and *Bam*HI restriction sites. For pTA131_Δ*pyrE2*_hmgA, the *hmgA* gene was amplified from pJWID1 ([Wendoloski et al., 2001](#); [Liao et al., 2016](#)) using prav_F and prav_R in a two-step PCR. The resulting product was then cloned into pTA131 using *Xba*I and *Not*I restriction sites.

For pTA131_hmgA_Δ*trpA*, the *hmgA* gene was cloned into pTA131 as described for pTA131_Δ*pyrE2*_hmgA. Subsequently, upstream (1273UF and 1273UR) and downstream (1273DF and 1273DR) flanking regions of the *trpA* gene Hlac_1273 were amplified separately and fused together in consecutive PCR reactions (as described above). The primers included restriction sites for *Hind*III and *Bam*HI, through which the PCR product was cloned into the pTA131_hmgA vector. The pTA131_hmgA_Δ*Hlac*_2746 and pTA132_Δ*Hlac*_2746 vectors were constructed the same way, *via* amplification and fusion of flanking regions of the putative GTPase gene Hlac_2746 (Hlac_2746UF, Hlac_2746UR, Hlac_2746DF, and Hlac_2746DR), followed by insertion into the plasmids (see [Supplementary Table 1](#)) with *Hind*III and *Bam*HI as described above.

Genetic manipulation of *Hrr. lacusprofundi*

All plasmids were transformed into *E. coli* C2925 (NEB) for demethylation before transformation into *Hrr. lacusprofundi* strains. Transformation of *Hrr. lacusprofundi* was performed as described earlier ([Liao et al., 2016](#)) following a PEG₆₀₀-based protocol with slight modifications. The incubation of the cells with DNA after the addition of PEG₆₀₀ lasted 70 min. Incubation in regeneration buffer was performed overnight at 30°C. For the transformation with pTA132_Δ*Hlac*_2746, the *Hrr. lacusprofundi* ACAM34_UNSWΔ*pyrE2*Δ*trpA* transformants were resuspended in regeneration solution containing 10x casamino acids replacing 10x YPC with trace elements and vitamins added in the ratio common for DBCM2 medium ([Dyall-Smith, 2009a](#)). Transformants were plated on selective media and liquid cultures were inoculated as well. For *pyrE2* deletions, both ACAM34_UNSW and ACAM34_DSMZ were directly plated on plates containing 150 µg/mL 5-fluoroorotic acid (5-FOA) and 50 µg/mL uracil. For *trpA* deletions, ACAM34_UNSW and ACAM34_UNSWΔ*pyrE2* were grown in DBCM2- media with pravastatin at concentrations indicated. Pop-out was performed on plates containing 150 µg/mL 5-FOA and 50 µg/mL uracil. ACAM34_UNSWΔ*pyrE2*Δ*trpA* mutants S1 and S2 were plated on

DBC2-media with uracil (50 µg/mL) selecting for pTA132_Δ*Hlac*_2746 which includes a functional *trpA* gene. The deletion mutant of *Hlac*_2746 in ACAM34_UNSWΔ*pyrE2* was subsequently generated by transformation with the pTA131_Δ*hmgA*_Δ*Hlac*_2746 plasmid and a two-step selection as described above for the *trpA* deletion mutants. We screened liquid cultures and single colonies for the correct insert by colony PCRs and Sanger sequencing as described above.

Genome sequencing and analysis of *trpA* mutants

Library preparation (FS DNA Library, NEBNext® Ultra™) and sequencing (Illumina HiSeq3000, 2 × 150 bp, 1 Gigabase per sample) was performed at the Max Planck-Genome-Centre Cologne (Cologne, Germany). For genome analysis, reads were mapped to ACAM34_UNSW genome using “geneious mapper” (Geneious Prime® 2022.2.1) with medium-low sensitivity and default settings. No reads could be recruited for the described gene deletion.

Results

Attempts to generate uracil auxotroph mutants by directed double crossover resulted in spontaneous mutants with unexpected deletions

Orotate *phosphoribosyltransferase* (*pyrE2*) is an ideal target to generate auxotrophic strains for genetic manipulation. Previous attempts to generate spontaneous Δ*pyrE2* mutants of *Hrr. lacusprofundi* with uracil (U) and 5-fluoroorotic acid (5-FOA) treatments (Bitan-Banin et al., 2003) have been unsuccessful (Liao et al., 2016). Therefore, we attempted to generate *pyrE2* deletions by transforming *Hrr. lacusprofundi* strain ACAM34_UNSW with a knockout (KO) PCR product composed of the fused up- and downstream flanks of the *pyrE2* gene of ACAM34, followed by selection on U/5-FOA plates (Figure 1C). In a successful double-crossover homologous recombination, the construct would be inserted into the genome and mutation of wild-type (WT) genes would be induced by 5-FOA. However, none of the obtained cultures tested showed the expected deletion in the *pyrE2* gene (Supplementary Figure 1), even though single-nucleotide polymorphisms (SNPs) and small insertions or deletions cannot be ruled out. Next, we cloned the KO PCR product into pTA131, a non-replicating vector designed for *H. volcanii* (Allers et al., 2004). We then transformed pTA131_Δ*pyrE2* and pTA131 as control into *Hrr. lacusprofundi* ACAM34_UNSW and ACAM34_DSMZ followed by selection on U/5-FOA plates. Similar growth was detected for both strains, and for both the control and pTA131_Δ*pyrE2*. The majority of colonies tested for ACAM34_UNSW revealed either wild-type size of the gene (approximately 60%) or an increased size of the gene, indicating the insertion of a transposase (approximately 40%) (Supplementary Figure 2) that was confirmed for one representative clone by sequencing. One out of 64 tested colonies revealed a PCR product indicative of a gene deletion; however, wild-type gene copies could still be detected. Haloarchaea are known for being polyploid

(Maurer et al., 2018; Ludt and Soppa, 2019) and the mixture of wild-type gene and mutant gene is likely the result of the heterozygous genome of this particular clone. After re-streaking, we obtained a clone (ACAM34_UNSWΔ*pyrE2* S1) with a clean PCR signal for a gene deletion without any apparent background levels of wild-type genes present (Figure 1A; Supplementary Figure 2B). The strain was unable to grow in absence of uracil, confirming the inactivation of the *pyrE2* gene. Surprisingly, sequencing of the genomic region revealed a deletion of 628 base pairs (bp) including the C-terminus of *pyrE2* [amino acids (aa) 73–191] and the C-terminus (last 84 aa) of a downstream gene, *Hlac*_0585, an ATPase associated with various cellular activities encoded on the opposite strand (Figure 1C). We conclude that this deletion is a spontaneous deletion, because it does not reflect the region we targeted with pTA131_Δ*pyrE2*.

In a third attempt to obtain a clean *pyrE2* deletion in ACAM34_UNSW, we cloned the *hmgA* gene from pJWID1, conferring resistance against pravastatin (Liao et al., 2016), into pTA131_Δ*pyrE2*. Pravastatin selection was used to force the insertion of the KO construct into the genome (pop-in) and then U/5-FOA to force the vector to excise again (pop-out). The construct was transformed into ACAM34_UNSW and two colonies that were positive in PCR screens for the *hmgA* gene were chosen for pop-out after two successive dilutions in 2.5 µg/mL pravastatin. As with previous attempts, a double-crossover homologous recombination could not be achieved (Supplementary Figure 3). Out of 32 screened colonies one clone with a deletion was identified (Supplementary Figure 2C), hereafter referred to as ACAM34_UNSWΔ*pyrE2*_S2 (Figure 1A). Sequencing of the genomic region also showed a deletion identical to ACAM34_UNSWΔ*pyrE2*_S1 that we conclude to be a spontaneous mutation. We are aware that this mutation also affects a second gene. However, we did not detect any major defects in neither growth (Figure 2C) nor morphology (Figure 2D and Supplementary Figure 4). The mutants are furthermore unable to grow on medium without uracil (Figure 1A). Since an in-frame deletion could not be achieved, we selected this auxotrophic mutant for further studies.

A similar phenomenon occurred in ACAM34_DSMZ. None of 40 tested colonies from the transformation with pTA131_Δ*pyrE2*, showed a change in size of the target region. However, out of eight colonies screened from a control transformation with water, two showed increased size and one a reduced size (ACAM34_DSMZΔ*pyrE2*) (Figure 1B; Supplementary Figure 5). Sequencing of the PCR product revealed a short deletion that expands over the promoter region of the gene and the first 16 aa (Figure 1C). After re-streaking of the strain, we were able to isolate a colony that had a clean PCR result for the gene deletion and was not able to grow without uracil (Figure 1B).

Successful in-frame deletion of the *trpA* gene in uracil auxotrophic strains required a double selection

To test whether the newly generated auxotrophic *pyrE2* mutant ACAM34_UNSWΔ*pyrE2*, is stable and suitable to be used for genetic manipulation, we targeted the alpha subunit of the tryptophan synthase (*trpA*, *Hlac*_1273) that is also commonly used in *H. volcanii* as a second auxotrophic marker (Allers et al., 2004).

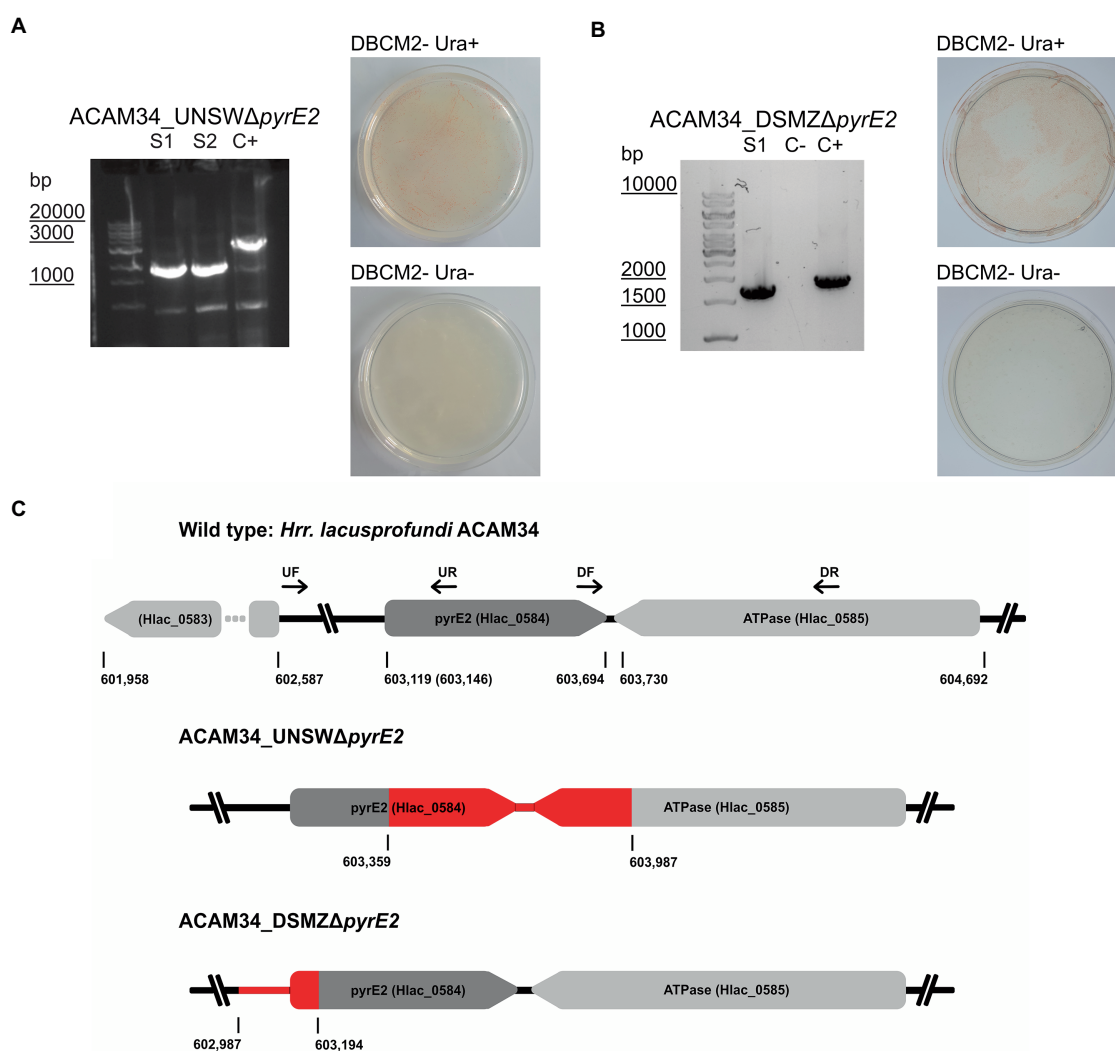


FIGURE 1

Successful *pyrE2* knockout in *Hrr. lacusprofundi* ACAM34 strains. **(A)** Characterization of ACAM34_UNSWΔpyrE2 mutants. Left: PCR on the *pyrE2* locus of the two knockout strains S1 and S2 (see also [Supplementary Figure 2](#)) and a wild-type ACAM34_UNSW control (C+). Right: Growth of ACAM34_UNSWΔpyrE2 on DBCM2- supplemented with uracil (top), and growth defect on DBCM2- without uracil (bottom). **(B)** Characterization of ACAM34_DSMZΔpyrE2. Left: PCR on the *pyrE2* locus of the knockout mutant, ACAM34_DSMZ wild-type DNA served as the positive control (+C) (see also [Supplementary Figure 5](#)). Right: Growth of ACAM34_DSMZΔpyrE2 on DBCM2- supplemented with uracil (top), and growth defect on DBCM2- without uracil (bottom). **(C)** Schematic representation of the genomes including the wild-type strain ACAM34_UNSW with the position (bp) of *pyrE2* and neighboring genes. ACAM34_DSMZ and ACAM34_UNSW deviate in the length of the annotated *pyrE2* gene, the different position is marked in brackets for ACAM34_DSMZ. The primer binding sites for upstream (UF, UR) and downstream (DF, DR) flanking regions are represented as arrows. The deletion in the ACAM34_UNSWΔpyrE2 S1 and S2 clones is identical and highlighted in red. The deletion in ACAM34_DSMZΔpyrE2 is also highlighted in red.

Both ACAM34_UNSW and ACAM34_UNSWΔpyrE2_S2 were transformed with pTA131_ *hmgA_ΔtrpA* containing a non-functional *trpA* construct ([Figure 3A](#)). While selection of ACAM34_UNSW transformants was only based on pravastatin resistance (2.5 μg/mL) conferred by *hmgA*, ACAM34_UNSWΔpyrE2 transformants were selected for the presence of the plasmid with both pravastatin (*hmgA*) and uracil depletion, forcing the uptake of the plasmid with the *H. volcanii pyrE2* gene (83% amino acid identity to ACAM34_UNSW). Growth of transformants was only achieved in liquid cultures and not on solid agar plates. However, while both ACAM34_UNSW and ACAM34_UNSWΔpyrE2 were growing in selective media, the presence of the plasmid could only be detected in

ACAM34_UNSWΔpyrE2 ([Supplementary Figure 6](#)). Since ACAM34_UNSW controls transformed with an empty vector (pTA131) or water also showed growth in liquid, we conclude that 2.5 μg/mL pravastatin was not sufficient for selection ([Supplementary Figure 6](#)). The slight growth of ACAM34_UNSWΔpyrE2 in liquid without uracil was only observed shortly after transformation and we assume that contaminations with uracil in the regeneration buffer could have supported the growth. The ACAM34_UNSWΔpyrE2 culture transformed with pTA131_ *hmgA_ΔtrpA* was plated on selective media (2.5 μg/mL pravastatin, DBCM2-) and single colonies were screened by PCR. While the presence of the *hmgA* gene was detectable in all clones, PCR with primers targeting the *trpA* gene on the genome

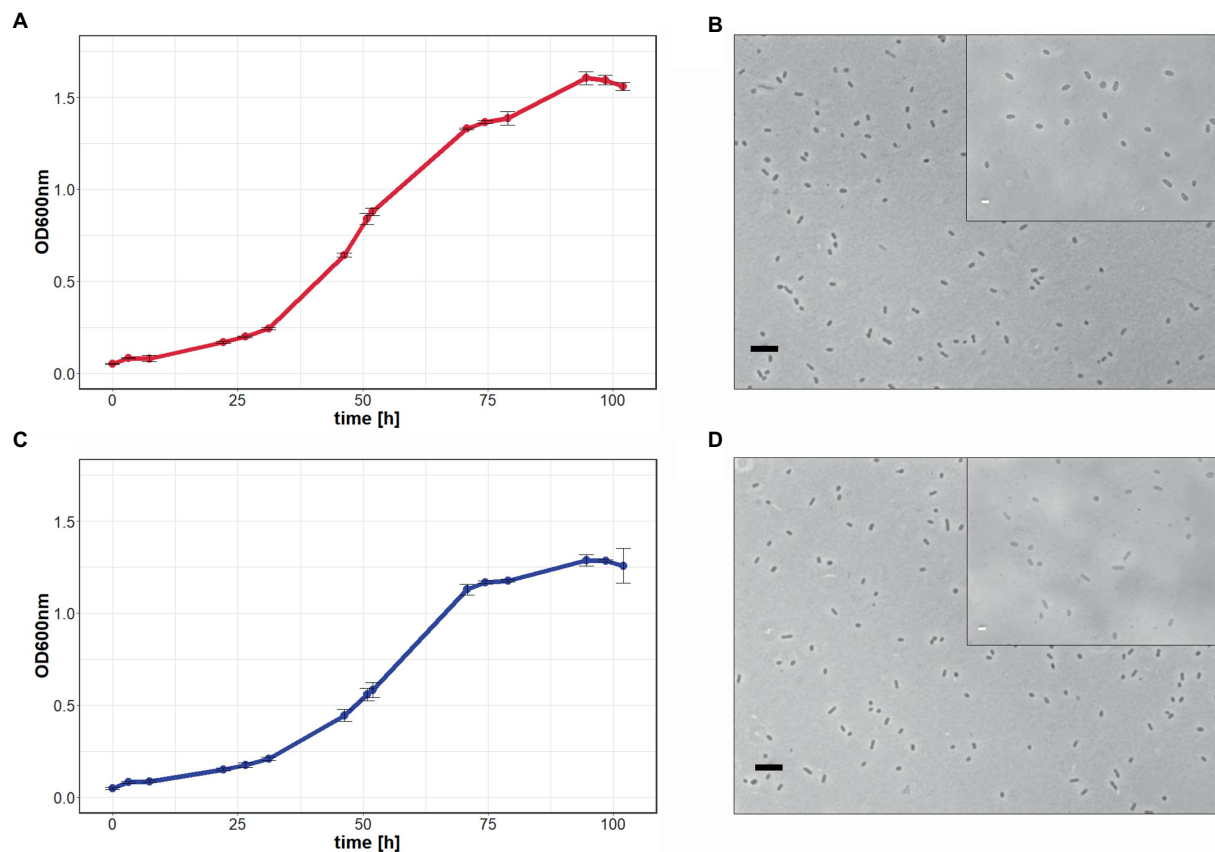


FIGURE 2

Growth comparison of *Hrr. lacusprofundi* ACAM34_UNSW and ACAM34_UNSWΔpyrE2. The growth of wild-type (A) and mutant cells (C) in rich medium was observed by optical density measurements at 600nm (OD_{600nm}) over 102h. Points represent the average values and each time point measured, including the standard deviation represented by error bars ($n=3$). Light microscopy images of wild-type (B) and mutant cells (D) represent one of the three biological replicates (Supplementary Figure 4). Larger image at 400x magnification, black scale bar indicates 10μm, inlet at 1,000x magnification, white scale bar indicates 2μm.

revealed mixed signal of both WT and mutated *trpA* genes in all samples (Supplementary Figure 7). We concluded that, similar to other haloarchaea (Maurer et al., 2018), *Hrr. lacusprofundi* exhibits several copies of its genome within one cell. The PCR signals from the targeted genomic region suggested that some chromosome copies had already excluded the plasmid, resulting in either a mutated or a wild-type *trpA*, and some wild-type copies might have not taken up the plasmid. To force the cell to exclude chromosome copies without the inserted plasmid before pop-out, we selected three colonies for culturing in increased (7.5 μg/mL) pravastatin concentrations in liquid culture; however, after four generations (G4), no changes were detected by PCR. The following generations were grown in increasing pravastatin concentrations of 10, 15, 20, and 40 μg/mL (G8) and both G4 and G8 were selected for pop-out on U/5-FOA plates. Colonies of G4 and G8 on pop-out plates were screened with genomic primers for the *trpA* gene. G4 colonies, with a PCR signal, all showed the wild-type gene with one exception that resulted in a mixture of wild-type and mutant *trpA* (Supplementary Figure 8). The majority of G8 pop-out clones showed a mixture of wild-type and mutant *trpA*, and five clones with a clean *trpA* deletion were identified (Supplementary Figure 9). These five clones were transferred into both full media (containing tryptophan and uracil) and selective media (with uracil only) to test for tryptophan auxotrophy. Four clones

reverted back to WT and were able to grow in selective media, but one clone referred to as ACAM34_UNSWΔpyrE2Δ*trpA* was unable to grow without tryptophan (Figure 3D) and PCR confirmed a corrupted *trpA* gene (Figure 3B; Supplementary Figure 10).

After replating, we selected two clones ACAM34_UNSWΔpyrE2Δ*trpA* (S1 and S2) for genome sequencing. For both ACAM34_UNSWΔpyrE2Δ*trpA* S1 and S2, no reads could be detected for the target region, between the upstream and downstream flanking regions of the non-functional *trpA* construct in pTA131_hmgA_Δ*trpA* (Figure 3C). The gap extends from aa63 to aa265, creating a clean 202aa deletion in both deletion strains without affecting any other genes or non-coding regions (Figure 3C).

A vector designed for *trpA* selection in *H. volcanii* can be used for selection in the tryptophan auxotrophic mutant ACAM34_UNSWΔpyrE2Δ*trpA*

All the plasmids created in this study were derived from plasmids originally created for the haloarchaeal model organism *H. volcanii*. For convenience in future experiments, we decided to test whether other plasmids designed for *H. volcanii* could be effectively used in

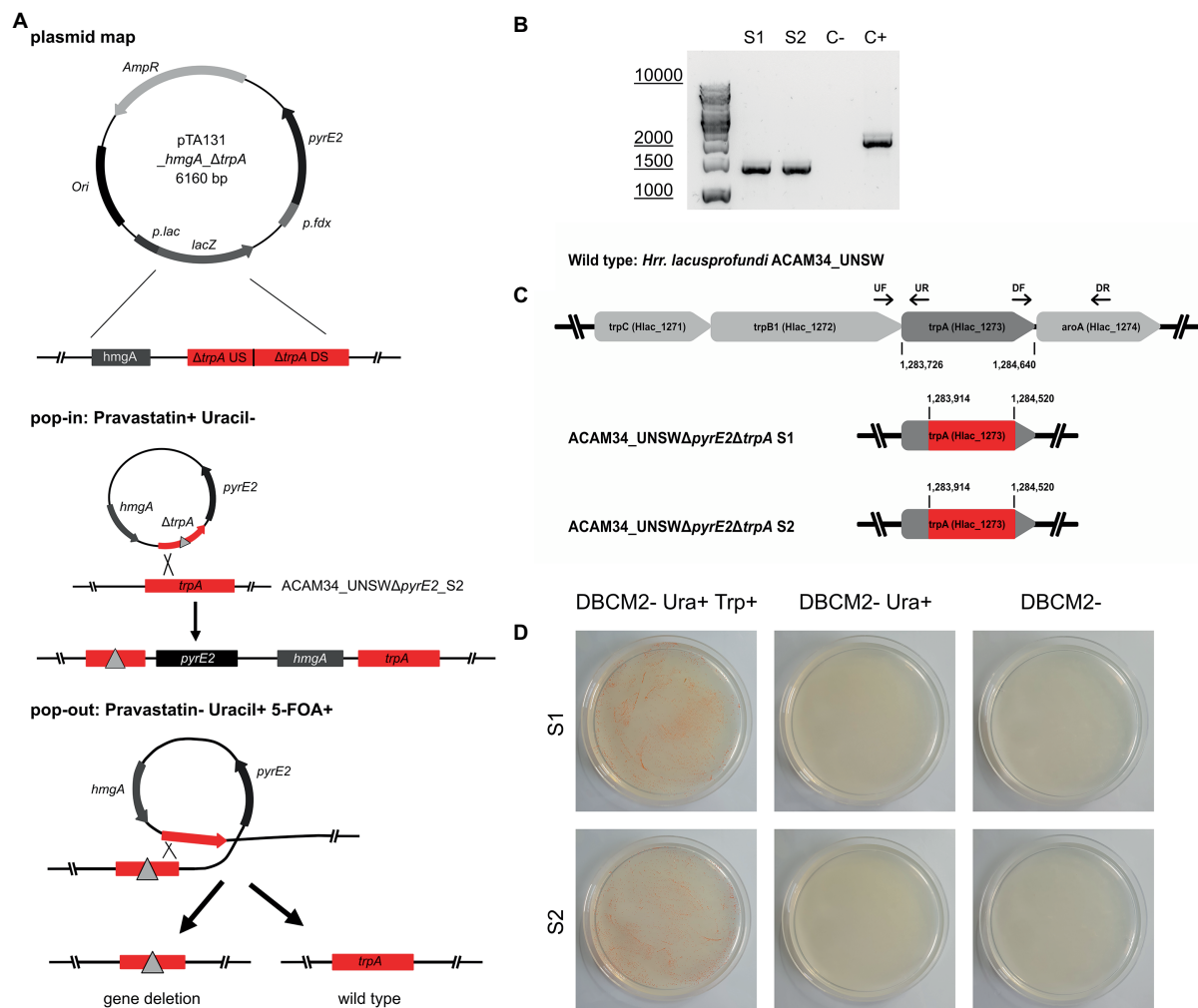


FIGURE 3

Successful *trpA* knockout in *Hrr. lacusprofundi* ACAM34_UNSWΔpyrE2. (A) Recombination strategy used to generate ACAM34_UNSWΔpyrE2Δ*trpA* via double selection. A vector based on pTA131 (Allers et al., 2004) (with *pyrE2* for uracil synthesis), was used to insert a pravastatin resistance gene (*hmgA*), as well as a non-functional copy of the *trpA* gene created by the fusion of upstream (US) and downstream (DS) fragments of the gene. Both the non-functional gene (represented by a triangle) and the functional wild-type *trpA* gene are marked in red. Crossover at the *trpA* genes leads to the integration of the plasmid into the genome (pop-in) using uracil depletion and the addition of pravastatin as selection pressure. Excision of the plasmid (pop-out) is enforced by the addition of 5-FOA and uracil, through which either the non-functional construct (left) or the wild-type gene (right) remains in the genome. (B) PCR on the *trpA* locus of the two knockout strains S1 and S2, with a wild-type ACAM34_UNSW control (C+) and a negative control (C-). (C) Operon encoding genes for tryptophan synthesis in the wild-type strain ACAM34_UNSW including the position (bp) of *trpA* and the primer binding sites for upstream (UF, UR) and downstream (DF, DR) flanking regions amplified for the deletion construct. Deletion in ACAM34_UNSWΔpyrE2Δ*trpA* S1 and in ACAM34_UNSWΔpyrE2Δ*trpA* S2 highlighted in red. (D) Growth of ACAM34_UNSWΔpyrE2Δ*trpA* S1 (top) and S2 (bottom) on DBCM2 supplemented with uracil and tryptophan (left), growth defect on DBCM2 with uracil but without tryptophan (middle), growth defect on DBCM2 (right).

Hrr. lacusprofundi without modifications. For this purpose, we selected a putative SAR1 family GTPase (Hlac_2746, Erdmann et al., 2017) as a target gene for a gene knockout. We chose pTA132 (Allers et al., 2004), a non-replicating vector including a functional *H. volcanii* *trpA* gene. In this case, plasmid uptake would be solely dependent on the *trpA* selection, allowing us to assess the efficiency of *trpA* selection in the newly generated ACAM34_UNSWΔpyrE2Δ*trpA* strain.

We transformed both ACAM34_UNSWΔpyrE2Δ*trpA* strains (S1 and S2) with pTA132_Δ*Hlac_2746* containing a non-functional Hlac_2746 construct, followed by selection with media lacking tryptophan [DBCM2- and uracil (50 μg/mL)]. Growth was only detected in liquid cultures of S2 transformants that went through regeneration in 10x casamino acids derived medium instead of 10x

YPC as in previous transformations (see Methods). No growth was detected for transformants of S1 neither on plates nor in liquid culture.

Without an additional antibiotic selection with pravastatin, we expected a lower selection efficiency and cultivated the transformants over successive generations before pop-out in selective media. After four generations (G4) we screened 34 colonies for plasmid insertion with primers binding in the region outside of Hlac_2746 and a long elongation time (6 min) to detect inserted plasmids (Supplementary Figure 11). Most colonies showed a mixture of both wild-type gene and inserted plasmid (approximately 95%), including two liquid cultures that represented the fifth generation (G5). We concluded that several copies of the genome are present in these cells some with an insertion and some without, as we observed

earlier. We conclude that pTA132 designed for *trpA*-mediated selection in *H. volcanii* was also effective for *trpA*-mediated selection in *Hrr. lacusprofundi*. However, *trpA* selection alone does not seem to be sufficient to force the cells to include the plasmid into all chromosome copies.

We plated G5 on rich medium (DBCM2+ 50 µg/mL uracil and tryptophan) to allow pop-out. After three successive passages (G6–G8) in rich medium, colonies ($n = 22$) screened in G9 either still retained the plasmid (41%), showed a mixture of wild-type Hlac_2746 and mutated Hlac_2746 (18%), a mixture of plasmid insertion and wild type (9%), a mixture of plasmid insertion, wild-type and mutated Hlac_2746 (23%), or no signal at all (9%) (Supplementary Figure 12). Since the insertion of the plasmid into all chromosome copies was already inefficient, it is not surprising that we could not identify any deletion mutants without background signal.

Subsequently, we attempted the knockout of Hlac_2746 in ACAM34_UNSWΔ*pyrE2* by double selection, using uracil depletion and pravastatin selection as described above. After transformation with pTA131_Δ*hmgA*_ΔHlac_2746, colonies that had successfully taken up the plasmid were subjected to increasing pravastatin concentrations over 8 generations in liquid culture (DBCM2-) followed by pop-out on U/5-FOA plates. Out of 62 colonies, 6 showed a PCR product smaller than wild-type Hlac_2746 (Supplementary Figure 13), and three of them were confirmed to contain the target deletion by sequencing.

Discussion

The previously reported genetic system for *Hrr. lacusprofundi* that is based on antibiotic selection, inserting the *hmgA* gene into the target gene (Liao et al., 2016), does not exclude effects on gene expression downstream of the deletion. Additionally, it was shown later that *hmgA* can be incorporated from a plasmid into the genome of *Hrr. lacusprofundi*, making it a volatile selection marker (Hamm et al., 2019). In this study, we aimed to generate a uracil auxotrophic mutant for robust selection and for the development of a genetic system allowing us to perform in-frame deletions.

Since generation of spontaneous *pyrE2* mutants on U/5-FOA plates was unsuccessful in a previous study (Liao et al., 2016), we attempted to generate *pyrE2* mutants by double crossover. All three attempts, using either a pure PCR product of the deletion construct or a vector including the deletion construct, with and without antibiotic selection, failed to produce the target deletion. However, we obtained spontaneous *pyrE2* mutants from these experiments for two strains of *Hrr. lacusprofundi* (ACAM34_UNSW and ACAM34_DSMZ). Two auxotrophic mutants of ACAM34_UNSW experienced the exact same deletion in two independent experiments (Figure 1B), suggesting that this deletion, including the 5'-part of a downstream gene, is not random. The strain does not show significant growth defects (Figure 2); therefore, ACAM34_UNSWΔ*pyrE2* was chosen as suitable candidate for further genetic manipulation.

To attempt the generation of an in-frame gene deletion, we chose the *trpA* gene as a target. We tested both wild-type ACAM34_UNSW and the uracil auxotroph ACAM34_UNSWΔ*pyrE2*. Antibiotics selection alone was not sufficient to achieve the uptake of the plasmid in ACAM34_UNSW wild type. However, double selection of the

auxotrophic mutant with antibiotic and uracil depletion in ACAM34_UNSWΔ*pyrE2* resulted in a successful uptake. A clean deletion mutant was only isolated from an ACAM34_UNSWΔ*pyrE2* transformant that was subjected to increasing concentrations of pravastatin (up to 40 µg/mL) before pop-out.

We subsequently tested the newly generated ACAM34_UNSWΔ*pyrE2*Δ*trpA* auxotrophic mutant for the efficiency of *trpA*-mediated selection without the additional selection markers *hmgA* and *pyrE2*. The plasmid, including a functional *trpA* gene derived from *H. volcanii*, was successfully taken up under tryptophan depleted conditions. However, an insertion into all chromosome copies could not be achieved without additional selection. Subsequently, we were not able to obtain clean deletion mutants. Haloarchaea are known for polyploidy (Maurer et al., 2018; Ludt and Soppa, 2019). Therefore, successive cultivation under selective conditions is essential to remove any remaining wild-type copies of the gene and selection for auxotrophic markers alone was not efficient in *Hrr. lacusprofundi* over four generations. Extending the selection over more generations might be sufficient, however, is also very time consuming. Alternatively, phosphate starvation, to reduce the number of genome copies (Maurer et al., 2018), could be used to add additional selection pressure. Nevertheless, the double selection with pravastatin and U/5-FOA yielded several clean deletion mutants that were not obtained with *trpA*-mediated selection when targeting the same gene. Both the selection with antibiotics alone and the auxotrophic marker alone proved to be unsuccessful. We conclude that the combination of an auxotrophic selection marker together with antibiotics selection is the most time efficient strategy to achieve clean deletion mutants in *Hrr. lacusprofundi*. Indeed, we and other collaborating laboratories have successfully generated several knockout mutants in *Hrr. lacusprofundi* ACAM34_UNSW using this strategy (unpublished results).

We successfully used plasmids originally designed for the haloarchaeal model organism *H. volcanii* for the genetic manipulation of *Hrr. lacusprofundi*. This opens up the possibility to use the entire genetic tool box available for *H. volcanii* (Dattani et al., 2022; Harrison and Allers, 2022) in *Hrr. lacusprofundi*. In addition to derivatives of the pTA131/132 used for gene knockouts (Liao et al., 2021), vectors for protein overexpression (Allers, 2010), recombinant luciferase (Davis et al., 2020), and fluorescent proteins (Reuter and Maupin-Furlow, 2004) are available to visually track gene expression (Duggin et al., 2015), protein interactions (Winter et al., 2018), and protein location inside of cells (Lestini et al., 2015; Liao et al., 2021; Nußbaum et al., 2021).

In conclusion, we successfully generated uracil and tryptophan auxotrophic mutants of *Hrr. lacusprofundi* ACAM34 and demonstrate that uracil auxotrophic mutants can be genetically manipulated using well-established plasmids designed for the model organism *H. volcanii*.

Data availability statement

The datasets presented in this study can be found in online repositories. Raw data from genome sequencing are available in the NCBI repository (<https://www.ncbi.nlm.nih.gov>) under BioProjectID: PRJNA894865.

Author contributions

LJG performed the majority of the experimental work. SE performed some of the experiments, conceived, and led the study. LJG and SE performed the primary writing of the manuscript. ID led the initial phase of the study. All authors participated in the analysis and interpretation of the data and contributed to the writing of the manuscript.

Funding

Funding was provided by the Volkswagen Foundation (reference 98 190), Germany, and the Max Planck Society (Munich, Germany). Part of this work was supported by the Australian Research Council Future fellowship to ID (FT160100010).

Acknowledgments

We thank the Max Planck-Genome-Centre Cologne (Cologne, Germany) for assistance with DNA sequencing. We thank Daniela Thies (MPI for Marine Microbiology, Bremen, Germany) for assistance with the experiments. We thank Thorsten Allers and Anita Marchfelder for providing *H. volcanii* plasmids. Finally, we want to

thank the Max-Planck-Institute for Marine Microbiology and the Max-Planck-Society for continuous support.

Conflict of interest

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.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2023.1095621/full#supplementary-material>

References

- Aguirre Sourrouille, Z., Schwarzer, S., Lequime, S., Oksanen, H. M., and Quax, T. E. F. (2022). The viral susceptibility of the *Haloferax* species. *Viruses* 14:1344. doi: 10.3390/v14061344
- Alarcón-Schumacher, T., Naor, A., Gophna, U., and Erdmann, S. (2022). Isolation of a virus causing a chronic infection in the archaeal model organism *Haloferax volcanii* reveals antiviral activities of a provirus. *Proc. Natl. Acad. Sci. U. S. A.* 119:e2205037119. doi: 10.1073/pnas.2205037119
- Allers, T. (2010). Overexpression and purification of halophilic proteins in *Haloferax volcanii*. *Bioeng. Bugs* 1, 290–292. doi: 10.4161/bbug.1.4.11794
- Allers, T., Ngo, H.-P., Mevarech, M., and Lloyd, R. G. (2004). Development of additional selectable markers for the halophilic archaeon *Haloferax volcanii* based on the *leuB* and *trpA* genes. *Appl. Environ. Microbiol.* 70, 943–953. doi: 10.1128/aem.70.2.943-953.2004
- Atanasova, N. S., Demina, T. A., Buivydas, A., Bamford, D. H., and Oksanen, H. M. (2015). Archaeal viruses multiply: Temporal screening in a solar saltern. *Viruses* 7, 1902–1926. doi: 10.3390/v7041902
- Atanasova, N. S., Roine, E., Oren, A., Bamford, D. H., and Oksanen, H. M. (2012). Global network of specific virus–host interactions in hypersaline environments. *Environ. Microbiol.* 14, 426–440. doi: 10.1111/j.1462-2920.2011.02603.x
- Bitan-Banin, G., Ortenberg, R., and Mevarech, M. (2003). Development of a gene knockout system for the halophilic archaeon *Haloferax volcanii* by use of the *pyrE* gene. *J. Bacteriol.* 185, 772–778. doi: 10.1128/jb.185.3.772-778.2003
- Blaseio, U., and Pfeifer, F. (1990). Transformation of *Halobacterium halobium*: Development of vectors and investigation of gas vesicle synthesis. *Proc. Natl. Acad. Sci. U. S. A.* 87, 6772–6776. doi: 10.1073/pnas.87.17.6772
- Boeke, J. D., LaCrute, F., and Fink, G. R. (1984). A positive selection for mutants lacking orotidine-5'-phosphate decarboxylase activity in yeast: 5-fluoro-orotic acid resistance. *Mol. Gen. Genet.* 197, 345–346. doi: 10.1007/bf00330984
- Correa, T., and Abreu, F. (2020). "Chapter 20 – Antarctic microorganisms as sources of biotechnological products" in *Physiological and biotechnological aspects of extremophiles*. eds. R. Salwan and V. Sharma (Cambridge: Academic Press), 269–284.
- DasSarma, S., Capes, M. D., Karan, R., and DasSarma, P. (2013). Amino acid substitutions in cold-adapted proteins from *Halorubrum lacusprofundi*, an extremely halophilic microbe from Antarctica. *PLoS One* 8:e58587. doi: 10.1371/journal.pone.0058587
- DasSarma, P., Laye, V. J., Harvey, J., Reid, C., Shultz, J., Yarborough, A., et al. (2017). Survival of halophilic archaea in Earth's cold stratosphere. *Int. J. Astrobiol.* 16, 321–327. doi: 10.1017/S1473550416000410
- Dattani, A., Harrison, C., and Allers, T. (2022). "Genetic manipulation of *Haloferax* species" in *Archaea: Methods and protocols*. ed. S. Ferreira-Cerca (New York, NY: Springer US), 33–56.
- Davis, C. R., Johnson, C. H., and Robertson, J. B. (2020). A bioluminescent reporter for the halophilic archaeon *Haloferax volcanii*. *Extremophiles* 24, 773–785. doi: 10.1007/s00792-020-01193-x
- de Silva, R. T., Abdul-Halim, M. F., Pittrich, D. A., Brown, H. J., Pohlschroder, M., and Duggin, I. G. (2021). Improved growth and morphological plasticity of *Haloferax volcanii*. *Microbiology* 167:001012. doi: 10.1099/mic.0.001012
- Duggin, I. G., Aylett, C. H., Walsh, J. C., Michie, K. A., Wang, Q., Turnbull, L., et al. (2015). CetZ tubulin-like proteins control archaeal cell shape. *Nature* 519, 362–365. doi: 10.1038/nature13983
- Dyall-Smith, M. (2009a). "Media recipes from the Thorsten Allers lab" in *The halohandbook: Protocols for halobacterial genetics*. ed. M. Dyall-Smith Available at: https://haloarchaea.com/wp-content/uploads/2018/10/Halohandbook_2009_v7.3mds.pdf
- Dyall-Smith, M. (2009b). "Medium DBCM2, for square haloarchaea of Walsby (*Haloquadratum walsbyi*)" in *The halohandbook: Protocols for halobacterial genetics*. ed. M. Dyall-Smith Available at: https://haloarchaea.com/wp-content/uploads/2018/10/Halohandbook_2009_v7.3mds.pdf
- Dyall-Smith, M. (2009c). "Trace element solution SL10" in *The halohandbook: Protocols for halobacterial genetics*. ed. M. Dyall-Smith Available at: https://haloarchaea.com/wp-content/uploads/2018/10/Halohandbook_2009_v7.3mds.pdf
- Dyall-Smith, M. (2009d). "HF1 and HF2 - examples of lytic, head-tail haloviruses" in *The halohandbook: Protocols for halobacterial genetics*. ed. M. Dyall-Smith Available at: https://haloarchaea.com/wp-content/uploads/2018/10/Halohandbook_2009_v7.3mds.pdf
- Erdmann, S., Tschitschko, B., Zhong, L., Raftery, M. J., and Cavicchioli, R. (2017). A plasmid from an Antarctic haloarchaeon uses specialized membrane vesicles to disseminate and infect plasmid-free cells. *Nat. Microbiol.* 2, 1446–1455. doi: 10.1038/s41564-017-0009-2
- Franzmann, P., Stackebrandt, E., Sanderson, K., Volkman, J., Cameron, D., Stevenson, P., et al. (1988). *Halobacterium lacusprofundi* sp. nov., a halophilic bacterium isolated from deep Lake, Antarctica. *Syst. Appl. Microbiol.* 11, 20–27. doi: 10.1016/S0723-2020(88)80044-4
- Hamm, J. N., Erdmann, S., Eloë-Fadros, E. A., Angeloni, A., Zhong, L., Brownlee, C., et al. (2019). Unexpected host dependency of Antarctic nanohaloarchaeota. *Proc. Natl. Acad. Sci. U. S. A.* 116, 14661–14670. doi: 10.1073/pnas.1905179116

- Harrison, C., and Allers, T. (2022). "Progress and challenges in archaeal genetic manipulation" in *Archaea: Methods and protocols*. ed. S. Ferreira-Cerca (New York, NY: Springer US), 25–31.
- Hartman, A. L., Norais, C., Badger, J. H., Delmas, S., Haldenby, S., Madupu, R., et al. (2010). The complete genome sequence of *Haloferax volcanii* DS2, a model archaeon. *PLoS One* 5:e9605. doi: 10.1371/journal.pone.0009605
- Javier, T.-C., Carmen Pire, G., and Rosa María, M.-E. (2017). "Biocompounds from haloarchaea and their uses in biotechnology" in *Archaea, Ch. 4*. eds. S. Haitham, N. Afef and G. Kais (Rijeka: IntechOpen)
- Krupović, M., Forterre, P., and Bamford, D. H. (2010). Comparative analysis of the mosaic genomes of tailed archaeal viruses and proviruses suggests common themes for virion architecture and assembly with tailed viruses of bacteria. *J. Mol. Biol.* 397, 144–160. doi: 10.1016/j.jmb.2010.01.037
- Lam, W. L., and Doolittle, W. F. (1992). Mevinolin-resistant mutations identify a promoter and the gene for a eukaryote-like 3-hydroxy-3-methylglutaryl-coenzyme A reductase in the archaeobacterium *Haloferax volcanii*. *J. Biol. Chem.* 267, 5829–5834. doi: 10.1016/S0021-9258(18)42628-2
- Lestini, R., Delpech, F., and Myllykallio, H. (2015). DNA replication restart and cellular dynamics of Hef helicase/nuclease protein in *Haloferax volcanii*. *Biochimie* 118, 254–263. doi: 10.1016/j.biochi.2015.07.022
- Li, J., Gao, Y., Dong, H., and Sheng, G.-P. (2022). Haloarchaea, excellent candidates for removing pollutants from hypersaline wastewater. *Trends Biotechnol.* 40, 226–239. doi: 10.1016/j.tibtech.2021.06.006
- Li, M., Wang, R., Zhao, D., and Xiang, H. (2014). Adaptation of the *Haloarcula hispanica* CRISPR-Cas system to a purified virus strictly requires a priming process. *Nucleic Acids Res.* 42, 2483–2492. doi: 10.1093/nar/gkt1154
- Liao, Y., Ithurbide, S., Evenhuis, C., Löwe, J., and Duggin, I. G. (2021). Cell division in the archaeon *Haloferax volcanii* relies on two FtsZ proteins with distinct functions in division ring assembly and constriction. *Nat. Microbiol.* 6, 594–605. doi: 10.1038/s41564-021-00894-z
- Liao, Y., Williams, T. J., Walsh, J. C., Ji, M., Poljak, A., Curmi, P. M. G., et al. (2016). Developing a genetic manipulation system for the Antarctic archaeon, *Halorubrum lacusprofundi*: Investigating acetamidase gene function. *Sci. Rep.* 6:34639. doi: 10.1038/srep34639
- Litchfield, C. D. (2011). Potential for industrial products from the halophilic archaea. *J. Ind. Microbiol. Biotechnol.* 38, 1635–1647. doi: 10.1007/s10295-011-1021-9
- Liu, Y., Demina, T. A., Roux, S., Aiewsakun, P., Kazlauskas, D., Simmonds, P., et al. (2021). Diversity, taxonomy, and evolution of archaeal viruses of the class Caudoviricetes. *PLoS Biol.* 19:e3001442. doi: 10.1371/journal.pbio.3001442
- Liu, H., Han, J., Liu, X., Zhou, J., and Xiang, H. (2011). Development of *pyrF*-based gene knockout systems for genome-wide manipulation of the archaea *Haloferax mediterranei* and *Haloarcula hispanica*. *J. Genet. Genomics* 38, 261–269. doi: 10.1016/j.jgg.2011.05.003
- Ludt, K., and Soppa, J. (2019). Polyploidy in halophilic archaea: Regulation, evolutionary advantages, and gene conversion. *Biochem. Soc. Trans.* 47, 933–944. doi: 10.1042/bst20190256
- Luk, A. W., Williams, T. J., Erdmann, S., Papke, R. T., and Cavicchioli, R. (2014). Viruses of haloarchaea. *Life* 4, 681–715. doi: 10.3390/life4040681
- Matsuda, R., Suzuki, S., and Kurosawa, N. (2022). Genetic study of four candidate holliday junction processing proteins in the thermophilic crenarchaeon *Sulfolobus acidocaldarius*. *Int. J. Mol. Sci.* 23:707. doi: 10.3390/ijms23020707
- Maurer, S., Ludt, K., and Soppa, J. (2018). Characterization of copy number control of two *Haloferax volcanii* replication origins using deletion mutants and haloarchaeal artificial chromosomes. *J. Bacteriol.* 200. doi: 10.1128/jb.00517-17
- McGenity, T. J., and Grant, W. D. (1995). Transfer of *Halobacterium saccharovorum*, *Halobacterium sodomense*, *Halobacterium trapanicum* NRC 34021 and *Halobacterium lacusprofundi* to the genus *Halorubrum* gen. Nov., as *Halorubrum saccharovorum* comb. nov., *Halorubrum sodomense* comb. nov., *Halorubrum trapanicum* comb. nov., and *Halorubrum lacusprofundi* comb. nov. *Syst. Appl. Microbiol.* 18, 237–243. doi: 10.1016/S0723-2020(11)80394-2
- Mercier, C., Thies, D., Zhong, L., Raftery, M. J., Cavicchioli, R., and Erdmann, S. (2022). In depth characterization of an archaeal virus-host system reveals numerous virus exclusion mechanisms. *bioRxiv*. doi: 10.1101/2022.10.18.512658
- Nußbaum, P., Gerstner, M., Dingethal, M., Erb, C., and Albers, S.-V. (2021). The archaeal protein SepF is essential for cell division in *Haloferax volcanii*. *Nat. Commun.* 12:3469. doi: 10.1038/s41467-021-23686-9
- Nuttall, S. D., and Dyal-Smith, M. L. (1993). HF1 and HF2: Novel bacteriophages of halophilic archaea. *Virology* 197, 678–684. doi: 10.1006/viro.1993.1643
- Piatek, P., Humphreys, C., Raut, M. P., Wright, P. C., Simpson, S., Köpke, M., et al. (2022). Agr Quorum Sensing influences the Wood-Ljungdahl pathway in *Clostridium autoethanogenum*. *Sci. Rep.* 12:411. doi: 10.1038/s41598-021-03999-x
- Pietilä, M. K., Roine, E., Sencilo, A., Bamford, D. H., and Oksanen, H. M. (2016). *Pleolipoviridae*, a newly proposed family comprising archaeal pleomorphic viruses with single-stranded or double-stranded DNA genomes. *Arch. Virol.* 161, 249–256. doi: 10.1007/s00705-015-2613-x
- Pohlschroder, M., and Schulze, S. (2019). *Haloferax volcanii*. *Trends Microbiol.* 27, 86–87. doi: 10.1016/j.tim.2018.10.004
- Porter, K., Tang, S.-L., Chen, C.-P., Chiang, P.-W., Hong, M.-J., and Dyal-Smith, M. (2013). PH1: An archaeovirus of *Haloarcula hispanica* related to SH1 and HHIV-2. *Archaea* 2013:456318. doi: 10.1155/2013/456318
- Redder, P., and Linder, P. (2012). New range of vectors with a stringent 5-fluoroorotic acid-based counterselection system for generating mutants by allelic replacement in *Staphylococcus aureus*. *Appl. Environ. Microbiol.* 78, 3846–3854. doi: 10.1128/AEM.00202-12
- Reuter, C. J., and Maupin-Furrow, J. A. (2004). Analysis of proteasome-dependent proteolysis in *Haloferax volcanii* cells, using short-lived green fluorescent proteins. *Appl. Environ. Microbiol.* 70, 7530–7538. doi: 10.1128/AEM.70.12.7530-7538.2004
- Singh, A., and Singh, A. K. (2017). Haloarchaea: Worth exploring for their biotechnological potential. *Biotechnol. Lett.* 39, 1793–1800. doi: 10.1007/s10529-017-2434-y
- Tschitschko, B., Williams, T. J., Allen, M. A., Páez-Espino, D., Kyrpides, N., Zhong, L., et al. (2015). Antarctic archaea–virus interactions: Metaproteome-led analysis of invasion, evasion and adaptation. *ISME J.* 9, 2094–2107. doi: 10.1038/ismej.2015.110
- van Wolferen, M., Pulschen, A. A., Baum, B., Gribaldo, S., and Albers, S. V. (2022). The cell biology of archaea. *Nat. Microbiol.* 7, 1744–1755. doi: 10.1038/s41564-022-01215-8
- Wendoloski, D., Ferrer, C., and Dyal-Smith, M. (2001). A new simvastatin (mevinolin)-resistance marker from *Haloarcula hispanica* and a new *Haloferax volcanii* strain cured of plasmid pHV2. The GenBank accession number for the sequence reported in this paper is AF123438. *Microbiology* 147, 959–964. doi: 10.1099/00221287-147-4-959
- Williams, T. J., Liao, Y., Ye, J., Kuchel, R. P., Poljak, A., Raftery, M. J., et al. (2017). Cold adaptation of the antarctic haloarchaea *Halohostia litchfieldiae* and *Halorubrum lacusprofundi*. *Environ. Microbiol.* 19, 2210–2227. doi: 10.1111/1462-2920.13705
- Winter, K., Born, J., and Pfeifer, F. (2018). Interaction of haloarchaeal gas vesicle proteins determined by Split-GFP. *Front. Microbiol.* 9:1897. doi: 10.3389/fmicb.2018.01897



OPEN ACCESS

EDITED BY

Solenne Ithurbide,
University of Freiburg, Germany

REVIEWED BY

Barry Whitman,
University of Georgia, United States
Mohea Couturier,
Vrije Universiteit Brussel, Belgium

*CORRESPONDENCE

Alexey Fomenkov
✉ fomenkov@neb.com
Richard J. Roberts
✉ roberts@neb.com

SPECIALTY SECTION

This article was submitted to
Biology of Archaea,
a section of the journal
Frontiers in Microbiology

RECEIVED 30 November 2022

ACCEPTED 21 March 2023

PUBLISHED 05 April 2023

CITATION

Fomenkov A, Weigele P, McClung C,
Madinger C and Roberts RJ (2023) Complete
genome assembly and methylome dissection
of *Methanococcus aeolicus* PL15/H^P.
Front. Microbiol. 14:1112734.
doi: 10.3389/fmicb.2023.1112734

COPYRIGHT

© 2023 Fomenkov, Weigele, McClung,
Madinger and Roberts. 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.

Complete genome assembly and methylome dissection of *Methanococcus aeolicus* PL15/H^P

Alexey Fomenkov*, Peter Weigele, Colleen McClung,
Casey Madinger and Richard J. Roberts*

New England Biolabs Inc., Ipswich, MA, United States

Although restriction-modification systems are found in both Eubacterial and Archaeal kingdoms, comparatively less is known about patterns of DNA methylation and genome defense systems in archaea. Here we report the complete closed genome sequence and methylome analysis of *Methanococcus aeolicus* PL15/H^P, a strain of the CO₂-reducing methanogenic archaeon and a commercial source for Mael, MaeII, and MaeIII restriction endonucleases. The *M. aeolicus* PL15/H^P genome consists of a 1.68 megabase circular chromosome predicted to contain 1,615 protein coding genes and 38 tRNAs. A combination of methylome sequencing, homology-based genome annotation, and recombinant gene expression identified five restriction-modification systems encoded by this organism, including the methyltransferase and site-specific endonuclease of MaeIII. The MaeIII restriction endonuclease was recombinantly expressed, purified and shown to have site-specific DNA cleavage activity *in vitro*.

KEYWORDS

complete genome, methylome, *Methanococcus aeolicus* PL15/H^P, restriction, modification, systems

Introduction

The role of DNA methylation in the restriction-modification (RM) mechanisms of bacteria and archaea was discovered more than 50 years ago (Arber, 1974; Loenen et al., 2014). Bacterial and archaeal DNA methyltransferases (MTases) create an epigenetic landscape in DNA by modifying specific motifs and, in concert with their cognate restriction endonucleases, determine the boundaries between bacterial species and protect against the entry of bacteriophages, conjugative plasmids and other mobile DNA elements. In addition to their roles in specifying self/non-self during genome defense, DNA methyltransferases appear as standalone genes; i.e., not associated with restriction endonucleases. Such “orphan” methyltransferases take on epigenetic functions such as regulation of genes for pathogenesis (Hernday et al., 2003) and ensuring fidelity of DNA repair (Schlagman et al., 1986).

Compared to their bacterial counterparts, the RM systems of archaea are relatively understudied. For example, of 54,450 genomes currently known to harbor RM systems, only 613 are from archaea (Roberts et al., 2022). Three restriction endonucleases, MaeI (C↓TAG), MaeII (A↓CGT), and MaeIII (↓GTNAC) have previously been characterized from *Methanococcus aeolicus* PL15/H^P (Schmid et al., 1984). *M. aeolicus* PL15/H^P, originally isolated from marine sediments of Lipari Island near Sicily (Kendall et al., 2006), is a strain of CO₂-reducing methanogenic Euryarchaeota. Because the genome sequence of this strain

has not been reported, we decided to perform SMRT sequencing of this archaeon to identify the Restriction-Modification (RM) system genes. Next generation sequencing platforms such as Single Molecule Real Time (SMRT) (Clark et al., 2012), Rapid Identification Methylase Specificity (RIMS) (Baum et al., 2021), and Oxford Nanopore Technology (ONT) (Tourancheau et al., 2021) allow not only the sequence and complete assembly of bacterial and archaeal genomes, but also enable the determination of methylation patterns and the confirmation of the MTase genes responsible for the modified motifs.

Herein, we report the genome sequence and methylome of *Methanococcus aeolicus* PL15/H^P, results which enabled the identification of genes encoding five RM-systems. Of these, methyltransferase activities of M.MaeIV and M.MaeV, as well as the restriction endonuclease activity of MaeIII, were further verified and characterized experimentally. These findings reveal the genetic basis of commercially relevant restriction endonucleases, connect methyltransferase genes to the epigenetic state of *M. aeolicus*, and extend what is known about RM systems in the Methanococcales branch of the Euryarchaeal lineage.

Materials and methods

Strains, plasmids, and reagents

Cells of *Methanococcus aeolicus* PL15/H^P were a gift from Prof. Dina Grohmann (Universität Regensburg), grown as previously described (Kendall et al., 2006). All laboratory *Escherichia coli* strains used in this study are listed in **Supplementary Table 1**. Cultures were grown in Luria-Bertani (LB) broth or agar on appropriate antibiotics. All plasmids used in this study are listed in **Supplementary Table 2**. Restriction enzymes, except MaeIII (Roche Diagnostic, GmbH, Mannheim, Germany), DNA substrates, DNA and protein markers were from New England Biolabs (NEB, MA). Q5 “Hot Start” DNA polymerase (M0543, NEB, MA) was used for PCR amplification of Methyltransferase and Restriction endonuclease genes for cloning. All PCR primers were synthesized by IDT, IA and are listed in **Supplementary Table 3**. The cloning of PCR amplified genes in appropriate vectors were performed using NEBuilder HiFi DNA assembly Master mix (E2621, NEB, MA). The plasmid DNA was purified using a Monarch Plasmid Miniprep kit (T1010, NEB, MA).

SMRT sequencing and genome assembly

Genomic DNA from *Methanococcus aeolicus* PL15/H^P was purified using a Monarch Genome Purification kit (T3010, NEB, Ipswich, MA, USA) and the DNA sample was sheared to an average size of ~ 10 kb using the G-tube protocol (Covaris, Woburn, MA, USA). DNA libraries were prepared using a SMRTbell express template prep kit 2.0 (100–938–900, Pacific Biosciences, Menlo Park, CA, USA) and ligated with hairpin barcoded adapter lbc1-lbc1. Incompletely formed SMRTbell templates were removed by digestion with a combination of exonuclease III and exonuclease VII (NEB, Ipswich, MA, USA). The qualification and quantification of the SMRTbell libraries were made on a Qubit fluorimeter

(Invitrogen, OR) and a 2,100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). SMRT sequencing was performed using an SQ1 (Pacific Biosciences, Menlo Park, CA, USA) based on the multiplex protocol for 10 kb SMRTbell library inserts. Sequencing reads were collected and *de novo* assembled using the Microbial Assembly version 10.1.0.1119588 program with default quality and read length parameters. In addition to genome assembly (Chin et al., 2013), the SMRT Analysis pipeline from Pacific Biosciences¹ enables the determination of the epigenetic status of sequenced DNA by identifying the m6A and m4C modified motifs (Flusberg et al., 2010; Clark et al., 2012; Korlach and Turner, 2012).

Computational and experimental identification of *M. aeolicus* PL15/H^P methyltransferases

Following an approach nicknamed the “Hungarian trick” (Szomolányi et al., 1980), 5 µg of *M. aeolicus* PL15/H^P gDNA was digested with BamHI, BglII and partially digested with Sau3AI restriction enzymes. DNA fragments were cloned into the compatible dephosphorylated BamHI sites of pUC19, pBR322 and pACYC184 vectors yielding 9 shot gun libraries. The foreign restriction and modification genes are usually very toxic for *E. coli*, therefore in order to clone both we need high level of methylation. These three vectors have different copy numbers pUC19 (high), pBR322 (medium) and pACYC184 (low). Moreover, they utilized different promoters P_{lac} and P_{tet}, since native *Methanococcus* promoters may or may not work in *E. coli*. Each purified plasmid library (1 µg) was challenged with 10 units of MaeIII restriction endonuclease and the digestion mixtures were transformed into *E. coli* strain ER2683. Plasmids were recovered from surviving transformants and analyzed by Sanger sequencing and restriction digest.

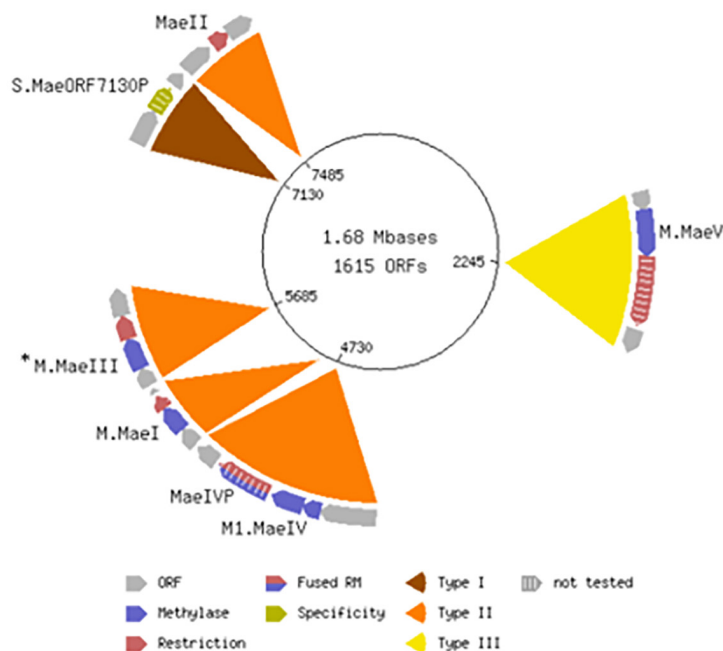
Homology-based searches of the *M. aeolicus* genome for putative methyltransferase genes were performed using the Seqware program (Roberts et al., 2022). Additional searches for MTase genes were performed using HMMer (HMMER 3.3.2)² to annotate each predicted protein coding sequence feature (CDS) of the *Methanococcus aeolicus* PL15/H^P genome. Genes matching MTase sequence profiles were further examined by structure prediction using the ColabFold implementation of AlphaFold2 (Jumper et al., 2021; Mirdita et al., 2022), followed by structure similarity search using predicted MTase models as query inputs to DALI (Holm, 2022).

Gene loci for M.MaeV and M1-M2.MaeIV were PCR amplified with P1-P2 primers and P3-P4 primers, respectively, and assembled into low copy pACYCΔtet vectors under control of a constitutive P_{tet} promoter and expressed in *E. coli* strain ER2683 (**Supplementary Table 1**). The correct clones were verified by Sanger sequencing of the inserts using an ABI DNA sequencer with pTetF and pTetR sequencing primers and transformed into the methylation deficient *E. coli* strain ER2796 (**Supplementary Table 1**).

1 <http://www.pacbiodevnet.com/SMRT-Analysis/Software/SMRT-Pipe>

2 <http://hmmer.org/>

A

Methanococcus aeolicus PL15/HP

B

motifString	centerPos	modificationType	fraction	nDetected	nGenome	meanScore	meanIpdRatio	meanCoverage	objectiveScore	methylase
GCAC	3	m6A	0.9979006	2852	2858	311.3103	3.822906	304.17252	886179.25	M1-M2.MaeIV
CGAG	3	m6A	0.9723698	1830	1882	320.3574	4.1349277	310.8328	571655.2	M.MaeV
GTAC	4	m6A	0.99863577	1464	1466	345.01843	4.044474	316.59357	504486.78	M.MaeIII
CTAG	1	m4C	0.9314961	1183	1270	246.58327	3.0786908	299.04733	273659.97	M.MaeI
ACGT	2	m5C								M.MaeII

FIGURE 1

(A) Schematic presentation of the complete closed circular genome of *Methanococcus aeolicus* PL15/HP showing the location of the RM systems. The *M.MaeIII* gene was not directly identified by the Seqware program (*). (B) Detection of modified DNA motifs identified by SMRT sequencing in *M. aeolicus* PL15/HP genome. Note that 5mC methylation by the *M.MaeII* gene was not directly identified by SMRT sequencing.

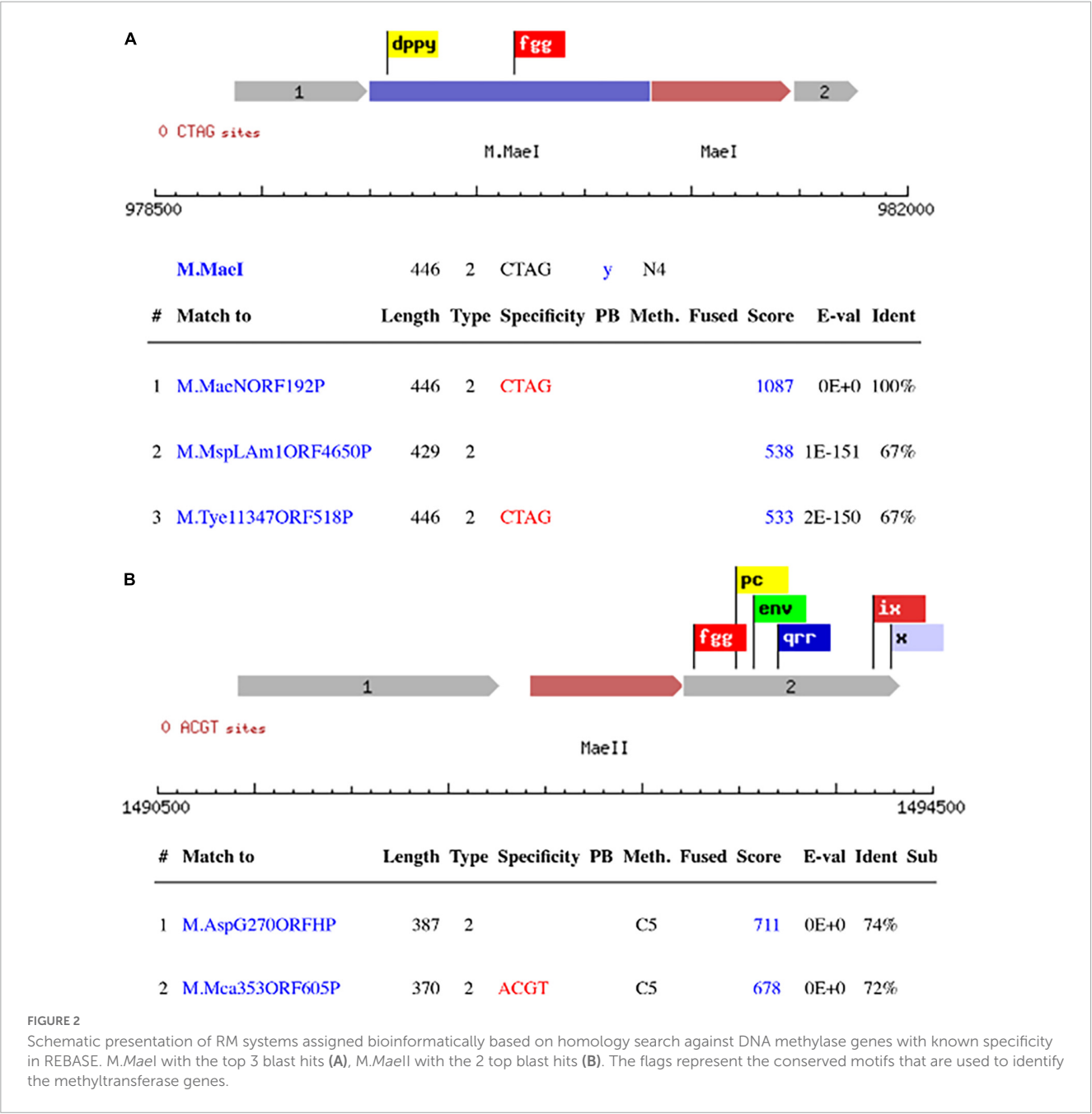
Liquid chromatography-tandem mass spectrometry analysis of functionally purified MaeIII endonuclease

Purified MaeIII endonuclease was subjected to 10–20% Tris-Glycine SDS-PAGE and stained with SimplyBlue SafeStain (Invitrogen, OR). The major band at approximately 35 kDa was excised for in-gel digestion (Figure 4A). The gel band was sliced into 1 mm³ pieces and the pieces were divided into two aliquots for multiple enzymatic digestions (Wiśniewski et al., 2019). The gel pieces were destained, and the protein was reduced with DTT, alkylated with iodoacetamide, and then digested with either Trypsin-ultra, Mass Spectrometry Grade (P8101; NEB, Ipswich, MA, USA) or subtilisin (Sigma-Aldrich, St. Louis, MO, USA) at a protease concentration of 2 ng/μL. Digestions were conducted for 1 h (50°C for trypsin, 37°C for subtilisin) and were quenched with trifluoroacetic acid at a final concentration of 0.5% for downstream analysis.

The resulting peptides were loaded via Proxeon Easy-nLC 1,200 (ThermoFisher, Waltham, MA, USA) onto a reversed phase analytical column 25 cm × 100 μm ID, Reprosil 3 μm C18 (New Objective, Littleton, MA, USA) and eluted at a flow rate

of 0.3 μL/min over a 35 min gradient of organic solvent (mobile phase A: 0.1% formic acid in water; mobile phase B: 90% acetonitrile/0.1% formic acid) as follows: 1-min equilibration at 10% B, 10–40% B in 25 min, 40–50% B in 4 min, and 50–78% B in 5 min. Peptides were electro-sprayed into a LTQ Orbitrap XL with a Nanospray Flex Ion Source (Thermo Fisher, Waltham, MA, USA) and fragmentation spectra were generated by collision-induced dissociation (CID). Full scans between 400–1,600 m/z were acquired with 30 k resolution. MS/MS spectra were generated in a data-dependent manner, with a normalized collision energy of 35.0 and dynamic exclusion of 30 s. The five most intense ions (including + 1 charge states for subtilisin-generated peptides) were selected for fragmentation.

Acquired MS/MS spectra were analyzed using PEAKS Studio Xpro (Bioinformatics Solutions, Inc., Waterloo, ON, USA). The data were searched against *Methanococcus aeolicus* PL15/HP translated ORFs from the genome sequence, allowing for a fixed carbamidomethyl modification on cysteines, variable oxidation of methionine, and two missed cleavages (tryptic peptides). The precursor and fragment mass tolerances were set to 20 ppm and 0.6 Da, respectively. Identified proteins were filtered by a 1% false discovery rate and required at least two unique peptides per protein.



Protein purification and endonuclease assay

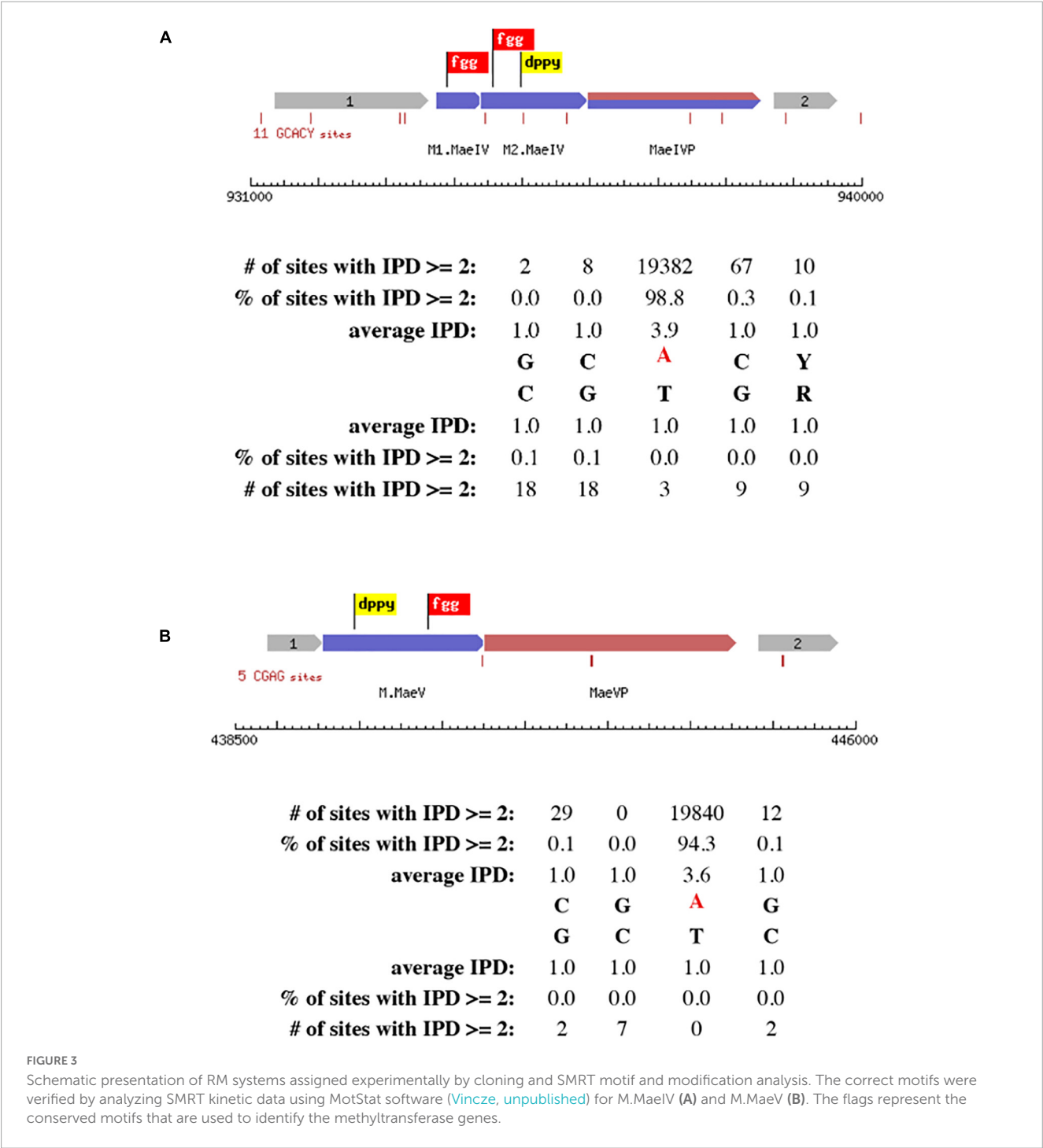
The recombinant *MaeIII* restriction enzyme was partially purified on Ni-NTA magnetic beads (S1423L, NEB, Ipswich, MA, USA) according to the manufacturer's protocol. DNA digestion patterns obtained by titration of restriction endonuclease activity from 250 mM imidazole Ni-NTA column elution fraction from *E. coli* clone 3081[pM.*MaeIII* + pET21b*MaeIII*] expressing *MaeIII* enzyme compared to *MaeIII* restriction enzyme from Roche on the same DNA substrate on λ DNA substrate. 1 μg of λ DNA incubated with Ni-NTA imidazole elution fraction of 6×His:*MaeIII* serial 3× dilution (from 9 to 0.1 μl) and serial 3× dilution of commercial *MaeIII* enzyme (from 9 to 0.1 units) for 1 h at 55°C in 1×

MaeIII commercial buffer (Roche Diagnostic, GmbH, Mannheim, Germany). 2× Buffer contains 40 mM Tris-HCl, 550 mM NaCl, 12 mM MgCl₂, 14 mM 2-mercaptoethanol, pH 8.2 (+55°C). The reaction mixtures were separated on 0.8% agarose gel in TBE buffer and visualized by staining with ethidium bromide.

Results

DNA-sequencing and genome assembly

Monarch purified gDNA (2 μg) from a *Methanococcus aeolicus* PL15/H^p culture was used to make one barcoded (lbc1-lbc1) 10 kb SMRT bell library sequenced on an SQ1 instrument



for 20 h. 418,400 long continuous sequencing subreads with 2,583 bp mean subread lengths were converted to 12,577 HiFi subreads with 5,328 bp mean subread length, yielding 43.7 Mb of high-quality sequencing data. The polished assembly generated one closed circular genomic element of 1,677,738 bp with 26-fold genome coverage and 29.93% G + C content for the chromosome (Figure 1A). The assembled sequence was annotated using the National Center for Biotechnology Information (NCBI) Prokaryotic Genome Annotation Pipeline (PGAP) (Tatusova et al., 2016; Haft et al., 2018) which identified 1,615 protein coding features and 38 tRNA genes. The annotated genome sequence of

M. aeolicus PL15/H^P is available from NCBI via the Bioproject PRJNA622823; BioSample: SAMN30825808, SRA: SRS15238210, and genome sequence: NZ_CP104873.

SMRT motif and modification analysis of *Methanococcus aeolicus* PL15/H^P genome

One advantage of the SMRT sequencing platform is its ability to directly detect the epigenetic state of sequenced DNA. As the

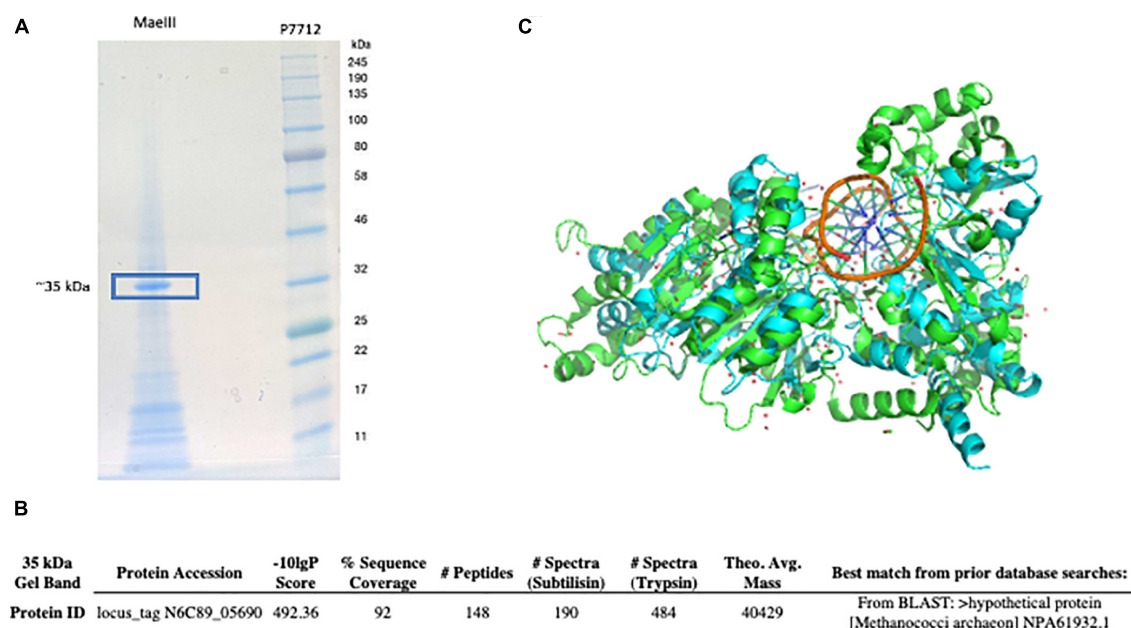


FIGURE 4

MaeIII restriction endonuclease separated on 10–20% Tris-Glycine SDS-PAGE and stained with SimplyBlue SafeStain. The protein band at ~35 kD was excised and used for MS/MS analysis (A). The sequence matched locus_tag N6C89_05690 from the *M. aeolicus* PL15/HP genome was detected by MS/MS analysis as putative MaeIII restriction enzyme (B). *M. aeolicus* PL15/HP assembled genome was scanned for additional DNA methyltransferase genes using an automated implementation of HMMer, which produced another potential hit corresponding to locus_tag N6C89_5685. The predicted structure of the M.MaeIII protein through AlphaFold2 monomeric modeling (C). The best scoring match to an AlphaFold2 predicted model of N6C89_5685 locus was 7lt5, the PDB structure of CamA, an adenine methyltransferase of *Clostridioides difficile* (Zhou et al., 2021).

DNA sequencing polymerase traverses methylated nucleobases in the sequencing template, changes in the amount of time between the nucleotide incorporation events (also known as the Interpulse Duration, or IPD) are recorded and analyzed by the SMRT-Link software yielding m6A and m4C modification motifs (Flusberg et al., 2010; Clark et al., 2012). The complete motif and modification analysis of assembled *Methanococcus aeolicus* PL15/HP using these kinetic signatures detected three m6A and one m4C modified motifs (Figure 1B). Two of the motifs with m4C CTAG and m6A CTNAG were recognized as signatures of the previously discovered MaeI and MaeIII RM systems (Schmid et al., 1984). The m5C motif of MaeII system ACGT cannot be directly detected by SMRT sequencing due to the weak IPD ratio signal generated by DNA polymerase on m5C modified base. Two additional m6A motifs with single strand DNA modification GCACY and CGAG were also detected by SMRT motif and modification analysis.

Matching putative DNA MTase genes to appropriate motifs

The primary scanning of the *M. aeolicus* PL15/HP assembled genome using Seqware identified 4 putative methyltransferase genes. Two of them for M.MaeI and M.MaeII where assigned bioinformatically based on homology search against methyltransferase genes in REBASE with known specificities (Figures 2A, B). The two other putative methyltransferase genes were assigned experimentally. Gene loci for M.MaeV and M1-M2.MaeIV were cloned into low copy pACYCΔtet vectors under

the control of a constitutive P_{tet} promoter and their expression induced in *E. coli* strain ER2683 (Supplementary Table 1). SMRT sequencing and methylation analysis of *E. coli* gDNA recovered from pM1-2.MaeIV and pM.MaeV expressing clones revealed that M.MaeV recognized the m6A modified motif CGAG. M1.MaeIV and M2.MaeIV belong to a family of so called split methylase/RM genes similar to DrdVI (Fomenkov et al., 2019) and recognize the m6A GCACY modified motif (Figures 3A, B). However, at that point, we were not able to assign the m6A GTNAC motif, corresponding to M.MaeIII, as there was no candidate gene identified by the PGAP annotation.

Searching for the MaeIII restriction-modification system

To find the missing gene for M.MaeIII we decided to apply an old methylation selection method known as the “Hungarian trick” (Szomolányi et al., 1980). Basically, this method is based on selection of clone(s) expressing the methyltransferase by treatment of the library with the cognate restriction endonuclease to destroy unmodified plasmid DNAs, leaving only plasmids that contain an active methyltransferase gene. In addition, we made a fosmid library in the pCC1Fos vector (Epicenter, WI). The plasmid and fosmid DNAs from survival clones were tested using a methylation protection assay for MaeIII resistance to digestion. Unfortunately, after selection and screening we were not able to find any MaeIII resistant clones (data not shown). Therefore, we decided to use a different approach to find the missing genes.

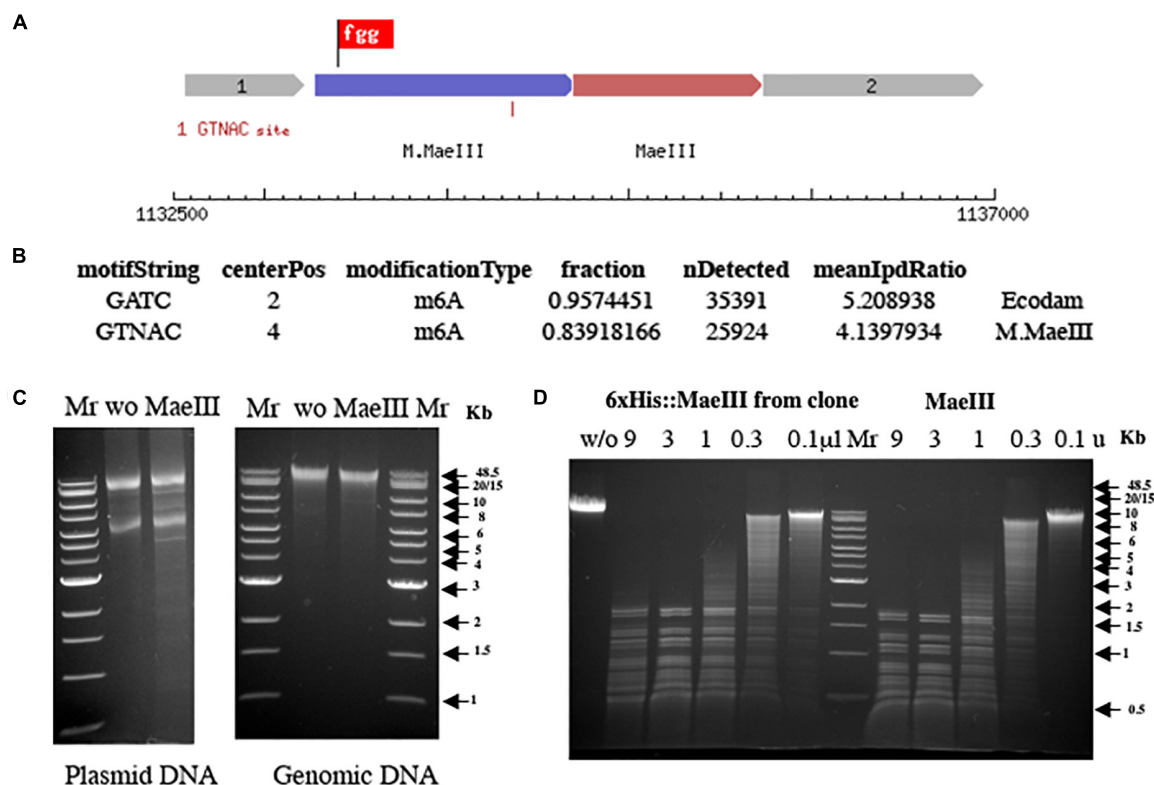


FIGURE 5

Schematic presentation of the MaeIII restriction-modification system (A). Motif and modification analysis by SMRT sequencing of gDNA from *E. coli* strain ER3081 [pM.MaeIII] (B). Note that this strain contains a Dam MTase in addition to M.MaeIII. MaeIII methylation protection assay on both plasmid and gDNA from *E. coli* 3081[pM.MaeIII] strain (C). Plasmid and gDNA from *E. coli* 3081[pM.MaeIII] were purified and 1 μg of each DNA was incubated without (w/o) as a negative control or with 1 unit of MaeIII restriction enzyme at 55°C for 1 h in 1× MaeIII commercial buffer (Roche Diagnostic, GmbH, Mannheim, Germany). The results demonstrated that both plasmid, and gDNA are resistant to MaeIII restriction endonuclease digestion due to modification of the DNA *in vivo* by M.MaeIII methyltransferase at GTNAC sites. λ DNA digestion patterns obtained by titration of restriction endonuclease activity from 250 mM imidazole Ni-NTA column elution fraction from *E. coli* clone 3081[pM.MaeIII] + pET21b 6xHis:MaeIII expressing MaeIII enzyme compared to MaeIII restriction enzyme from Roche on the same DNA substrate (D). 1 μg of λ DNA incubated with Ni-NTA imidazole elution fraction of 6xHis:MaeIII serial 3× dilution (from 9 to 0.1 μl) and serial 3× dilution of commercial MaeIII enzyme (from 9 to 0.1 units) for 1 h at 55°C in 1× MaeIII commercial buffer (Roche Diagnostic, GmbH, Mannheim, Germany). 2× Buffer contains 40 mM Tris-HCl, 550 mM NaCl, 12 mM MgCl₂, 14 mM 2-mercaptoethanol, pH 8.2 (+55°C). The reaction mixtures were separated on 0.8% agarose gel in TBE buffer and visualized by staining with ethidium bromide.

We used an MS/MS method to identify the protein sequence in functionally purified MaeIII from Roche (Figure 4A). This yielded a top target hit corresponding to locus_tag N6C89_05690 from *M. aeolicus* PL15/H^P (Figure 4B). In addition, the *M. aeolicus* PL15/H^P assembled genome was scanned for additional DNA methyltransferase genes using an automated implementation of HMMer, which produced another potential hit corresponding to locus_tag N6C89_5685. The best scoring match to an AlphaFold2 predicted model of N6C89_5685 was 7lt5, the PDB structure of CamA, an adenine methyltransferase of *Clostridioides difficile* (Figure 4C; Zhou et al., 2021). Thus, the close proximity of the two identified target genes strongly suggested that they corresponded to the MaeIII RM system.

Cloning and expression of the MaeIII restriction-modification system

The genes corresponding to locus_tags 5,685 and 5,690 (Figure 5A) were cloned into the low copy pACYCΔtet vector

under control of a constitutive P_{tet} promoter for expression in *E. coli* T7 expression strain, ER3081. The plasmid and gDNA from this strain were purified and challenged with 10 units of the MaeIII restriction enzyme to test for methylation protection (Figure 5C). In addition, the DNA obtained from *E. coli* expressing this putative M.MaeIII methyltransferase was sequenced using the SMRT analysis platform, identifying the sequence GTNAC as modified (Figure 5B), consistent with the M.MaeIII modification motif previously reported (Schmid et al., 1984). These results clearly indicated that locus_tag N6C89_5685 did indeed code for M.MaeIII. Therefore the strain was designated as ER3081[pM.MaeIII].

A PCR amplicon spanning the presumptive endonuclease gene, locus_tag 5690, was cloned in-frame with the N-terminal 6× His tag and linker sequence of the pET21b expression vector under the control of the pT7 promoter and transformed into an *E. coli* strain previously transformed with a compatible M.MaeIII expressing plasmid. The resulting two-plasmid strain encoding M.MaeIII and the putative MaeIII endonuclease gene was induced and the his-tagged putative endonuclease purified using Ni-NTA

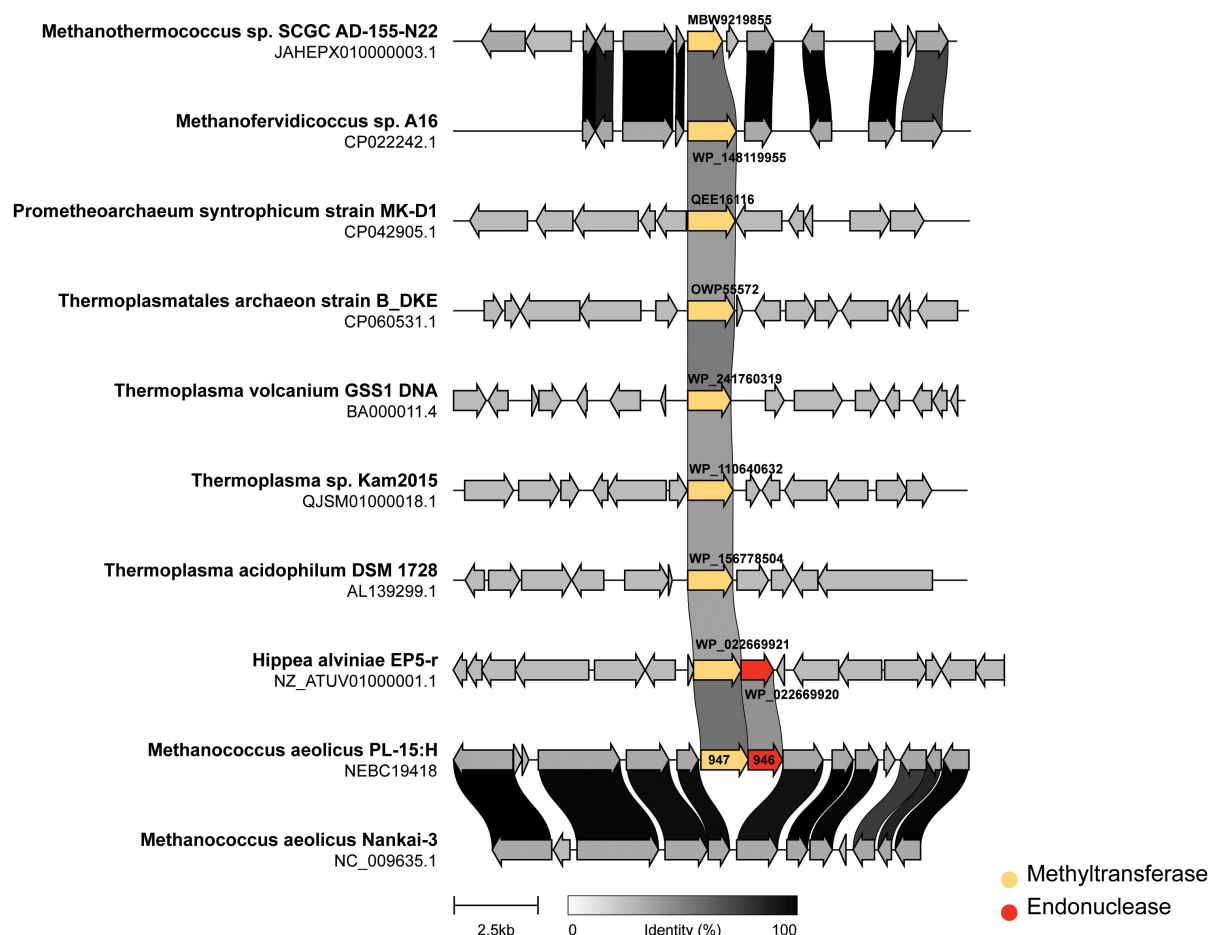


FIGURE 6

Genomic neighborhoods of *M.MaeIII* homologs across representative archaea. The m6A DNA methylase (*M.MaeIII*) of *M. aeolicus* PL-15:H was used as a query sequence for a BLAST search taxonomically limited to archaea (taxid:2,157) at NCBI. Eight genomes from cultured archaea and one bacterium containing matching sequences to *M.MaeIII* with an *e*-value of less than 10^{-30} were selected for genome neighborhood analysis. Sub genomic sequences surrounding the *M.MaeIII* homologs were manually obtained by computationally extracting 7.5 kb of sequence, plus annotations, upstream and downstream of the methyltransferase encoding sequence. The resulting gene clusters were aligned and rendered using “clinker” gene diagramming software (Gilchrist and Chooi, 2020) with the methyltransferase colored yellow and endonuclease colored in red. The degree of amino acid identity among coding sequences are indicated by grayscale links connecting homologous genes.

beads as described. *MaeIII* restriction endonuclease activity was subsequently detected in a 250 mM imidazole elution fraction (Figure 5D).

Discussion

In this paper we report the complete closed genome of *Methanococcus aeolicus* PL15/H^p. The sequencing data are publicly available from NCBI. In addition, the complete methylome of this strain and the identified RM systems may help other investigators to develop novel genetic systems for *M. aeolicus* and create efficient transformation protocols for this archaeal species.

In addition we performed genomic neighborhood analysis of *M.MaeIII* homologs across representative archaea. Figure 6 shows the methyltransferases (indicated in yellow) either in an apparent “orphan” configuration (i.e., no neighboring restriction endonuclease) or associated with an endonuclease sequence (indicated in red). *M. aeolicus* PL-15:H and *M. aeolicus* Nankai-3

are highly syntenic at this locus, however, *M. aeolicus* Nankai-3 completely lacks the *MaeIII* RM system. Like the Nankai-3 strain, *M. jannaschii* lacks a *M.MaeIII* homolog. An *MaeIII*-like RM system is found in the thermophilic bacterium *Hippea alviniae* EP5-r. *M.MaeIII* homologs are found in a variety of unrelated contexts across representatives from Euryarchaeota and Lokiarchaeota. One conserved configuration of an *M.MaeIII*-like gene occurs between *Methanothermococcus* sp., SCGC AD-144-N22 and *Methanofervidicoccus* sp., A16 where apparent orphan methylases are found in syntenic loci sharing a high degree of amino-acid identity. The function of these orphan methyltransferases is currently unknown.

Taken together, the data presented herein extend what is known about the occurrence and function of RM systems in the Methanococcales branch of the Euryarchaeota lineage. The synteny/neighborhood analysis presented in Figure 5 are consistent with previously reported findings in the Haloarchaea (Fullmer et al., 2019); i.e., the *MaeIII* RM system does not seem to be conserved between closely related *Methanococcus* species and yet

is found intact in a eubacterial extremophile. These observations are suggestive of a mobile and dynamic distribution not only of the intact RM system, but also homologous methyltransferases found in “orphan” configuration. It remains to be seen what, if any, would be the impacts of the RM systems and orphan methyltransferases of *Methanococcus* on mating (Makkay et al., 2020), transformability (Zatopek et al., 2021), and DNA repair.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: <https://www.ncbi.nlm.nih.gov/genbank/>, CP104873.

Author contributions

AF performed SMRT sequencing, genome assembly, and writing—original manuscript. PW performed bioinformatic analysis. CMc and CMA performed MS/MS analysis. RR performed methylome analysis and editing final manuscript. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by New England Biolabs Inc.

References

- Arber, W. (1974). DNA modification and restriction. *Prog. Nucleic Acid Res. Mol. Biol.* 14, 1–37.
- Baum, C., Lin, Y. C., Fomenkov, A., Anton, B. P., Chen, L., Yan, B., et al. (2021). Rapid identification of methylase specificity (RIMS-seq) jointly identifies methylated motifs and generates shotgun sequencing of bacterial genomes. *Nucleic Acids Res.* 49:e113. doi: 10.1093/nar/gkab705
- Chin, C.-H., Alexander, D. H., Marks, P., Klammer, A. A., Drake, J., Heiner, C., et al. (2013). Nonhybrid, finished microbial genome assemblies from long-read SMRTsequencing data. *Nat. Methods* 10, 563–569. doi: 10.1038/nmeth.2474
- Clark, T. A., Murray, I. A., Morgan, R. D., Kislyuk, A. O., Spittle, K. E., Boitano, M., et al. (2012). Characterization of DNA methyltransferase specificities using single-molecule, real-time DNA sequencing. *Nucleic Acids Res.* 40, 11450–11462.
- Flusberg, B. A., Webster, D. R., Lee, J. H., Travers, K. J., Olivares, E. C., Clark, T. A., et al. (2010). Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nat. Methods* 7, 461–465.
- Fomenkov, A., Luyten, Y., Vincze, T., Anton, B. P., Roberts, R. J., and Morgan, R. D. (2019). Complete Genome Sequence and Methylome Analysis of *Deinococcus wulumuqiensis* 479. *Microbiol Resour Announc.* 12:e00066-19. doi: 10.1128/MRA.00066-19
- Fullmer, M. S., Ouellette, M., Louyakis, A. S., Papke, R. T., and Gogarten, J. P. (2019). The Patchy Distribution of Restriction-Modification System Genes and the Conservation of Orphan Methyltransferases in Halobacteria. *Genes* 10:233. doi: 10.3390/genes10030233
- Gilchrist, C. L. M., and Chooi, Y.-H. (2020). Clinker & clustermap.js: Automatic generation of gene cluster comparison figures. *Bioinformatics* 37, 2473–2475. doi: 10.1093/bioinformatics/btab007
- Haft, D. H., DiCuccio, M., Badretdin, A., Brover, V., Chetvernin, V., O'Neill, K., et al. (2018). RefSeq: an update on prokaryotic genome annotation and curation. *Nucleic Acids Res.* 46, D851–D860. doi: 10.1093/nar/gkx1068.6
- Hernday, A., Bruce, A. B., and Low, D. (2003). The mechanism by which DNA adenine methylase and PapI activate the pap epigenetic switch. *Mol. Cell* 12, 947–957.
- Holm, L. (2022). Dali server: Structural unification of protein families. *Nucleic Acids Res.* 50, W210–W215. doi: 10.1093/nar/gkac387
- Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature* 2021, 583–589. doi: 10.1038/s41586-021-03819-2
- Kendall, M. M., Liu, Y., Sieprawska-Lupa, M., Stetter, K. O., Whitman, W. B., and Boone, D. R. (2006). *Methanococcus aeolicus* sp. nov., a mesophilic, methanogenic archaeon from shallow and deep marine sediments. *Int. J. Syst. Evol. Microbiol.* 56, 1525–1529. doi: 10.1099/ijs.0.642160
- Korlach, J., and Turner, S. W. (2012). Going beyond five bases in DNA sequencing. *Curr. Opin. Struct. Biol.* 22, 251–261. doi: 10.1016/j.sbi.2012.04.002
- Loenen, W. A., Dryden, D. T., Raleigh, E. A., Wilson, G. G., and Murray, N. E. (2014). Highlights of the DNA cutters: A short history of the restriction enzymes. *Nucleic Acids Res.* 42, 3–19. doi: 10.1093/nar/gkt990
- Makkay, A. M., Louyakis, A. S., Ram-Mohan, N., Gophna, U., Gogarten, J. P., and Papke, R. T. (2020). Insights into gene expression changes under conditions that facilitate horizontal gene transfer (mating) of a model archaeon. *Sci. Rep.* 10:22297. doi: 10.1038/s41598-020-79296-w
- Mirdita, M., Schütze, K., Moriawaki, Y., Heo, L., Ovchinnikov, S., and Steinegger, M. (2022). ColabFold: Making protein folding accessible to all. *Nat. Methods* 19, 679–682. doi: 10.1038/s41592-022-01488-1
- Roberts, R. J., Vincze, T., Posfai, J., and Macelis, D. (2022). REBASE—a database for DNA restriction and modification: Enzymes, genes and genomes. *Nucleic Acids Res.* [Epub ahead of print]. doi: 10.1093/nar/gkac975
- Schlagman, S., Hattman, S., and Marinus, M. (1986). Direct role of the *Escherichia coli* Dam DNA methyltransferase in methylation-directed mismatch repair. *J. Bacteriol.* 165, 896–900. doi: 10.1128/jb.165.3.896-900.1986

Acknowledgments

We thank Yi Sun and Sean Johnson with their help on bioinformatic analysis and Léa Chuzel for help with the construction of the fosmid library.

Conflict of interest

The authors work for New England Biolabs Inc., a company that sells research reagents, including restriction enzymes and DNA methyltransferases, to the scientific community.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2023.1112734/full#supplementary-material>

- Schmid, K., Thomm, M., Laminet, A., Laue, F. G., Kessler, C., Stetter, K. O., et al. (1984). Three new restriction endonucleases MaeI, MaeII and MaeIII from *Methanococcus aeolicus*. *Nucleic Acids Res.* 6, 2619–2628. doi: 10.1093/nar/12.6.2619
- Szomolányi, E., Kiss, A., and Venetianer, P. (1980). Cloning the modification methylase gene of *Bacillus sphaericus* R in *Escherichia coli*. *Gene* 3, 219–225. doi: 10.1016/0378-1119(80)90051-7
- Tatusova, T., DiCuccio, M., Badretdin, A., Chetvernin, V., Nawrocki, E. P., Zaslavsky, L., et al. (2016). NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Res.* 44, 6614–6624.
- Tourancheau, A., Mead, E. A., Zhang, X. S., and Fang, G. (2021). Discovering multiple types of DNA methylation from bacteria and microbiome using nanopore sequencing. *Nat. Methods* 5, 491–498. doi: 10.1038/s41592-021-01109-3
- Vincze, T. (unpublished). *MotStat software for SMRT motif and modification analysis*.
- Wiśniewski, J. R., Wegler, C., and Artursson, P. (2019). Multiple-Enzyme-Digestion Strategy Improves Accuracy and Sensitivity of Label- and Standard-Free Absolute Quantification to a Level That Is Achievable by Analysis with Stable Isotope-Labeled Standard Spiking. *J. Proteome Res.* 18, 217–224. doi: 10.1021/acs.jproteome.8b00549
- Zatopek, K. M., Burkhart, B. W., Morgan, R. D., Gehring, A. M., Scott, K. A., Santangelo, T. J., et al. (2021). The Hyperthermophilic Restriction-Modification Systems of *Thermococcus kodakarensis* Protect Genome Integrity. *Front. Microbiol.* 12:657356. doi: 10.3389/fmicb.2021.657356
- Zhou, J., Horton, J. R., Blumenthal, R. M., Zhang, X., and Cheng, X. (2021). *Clostridioides difficile* specific DNA adenine methyltransferase CamA squeezes and flips adenine out of DNA helix. *Nat. Commun.* 12:3436. doi: 10.1038/s41467-021-23693-w



OPEN ACCESS

EDITED BY

Michel Geovanni Santiago-Martínez,
University of Connecticut, United States

REVIEWED BY

Jose Munoz,
Northumbria University, United Kingdom
Sergio Romero Romero,
University of Bayreuth, Germany

*CORRESPONDENCE

Olivier N. Lemaire
✉ olemaire@mpi-bremen.de
Tristan Wagner
✉ twagner@mpi-bremen.de

RECEIVED 03 March 2023

ACCEPTED 03 April 2023

PUBLISHED 11 May 2023

CITATION

Lemaire ON, Belhamri M and Wagner T (2023)
Structural and biochemical elucidation of class
I hybrid cluster protein natively extracted from
a marine methanogenic archaeon.
Front. Microbiol. 14:1179204.
doi: 10.3389/fmicb.2023.1179204

COPYRIGHT

© 2023 Lemaire, Belhamri and Wagner. 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.

Structural and biochemical elucidation of class I hybrid cluster protein natively extracted from a marine methanogenic archaeon

Olivier N. Lemaire*, Mélissa Belhamri and Tristan Wagner*

Max Planck Institute for Marine Microbiology, Bremen, Germany

Whilst widespread in the microbial world, the hybrid cluster protein (HCP) has been paradoxically a long-time riddle for microbiologists. During three decades, numerous studies on a few model organisms unravelled its structure and dissected its metal-containing catalyst, but the physiological function of the enzyme remained elusive. Recent studies on bacteria point towards a nitric oxide reductase activity involved in resistance during nitrate and nitrite reduction as well as host infection. In this study, we isolated and characterised a naturally highly produced HCP class I from a marine methanogenic archaeon grown on ammonia. The crystal structures of the enzyme in a reduced and partially oxidised state, obtained at a resolution of 1.45 and 1.36-Å, respectively, offered a precise picture of the archaeal enzyme intimacy. There are striking similarities with the well-studied enzymes from *Desulfovibrio* species regarding sequence, kinetic parameters, structure, catalyst conformations, and internal channelling systems. The close phylogenetic relationship between the enzymes from *Methanococcales* and many *Bacteria* corroborates this similarity. Indeed, *Methanococcales* HCPs are closer to these bacterial homologues than to any other archaeal enzymes. The relatively high constitutive production of HCP in *M. thermolithotrophicus*, in the absence of a notable nitric oxide source, questions the physiological function of the enzyme in these ancient anaerobes.

KEYWORDS

methanogenic archaea (MA), hybrid cluster protein, anaerobic biochemistry, structural biology, metalloenzyme active site, internal channel, nitric oxide

1. Introduction

Hybrid clusters are metallo-cofactors constituted of a mixture of Fe atoms bridged with sulphur and oxygen atoms. The most notorious enzymes harbouring them as a prosthetic group are the hybrid cluster proteins (HCPs, also historically named prismane protein) (Aragão et al., 2008; Hagen, 2022), which are present in strict or facultative anaerobic archaea and bacteria [e.g., the bacterium *Desulfovibrio vulgaris* (Cooper et al., 2000) and the archaeon *Pyrococcus furiosus* (Overeijnder et al., 2009)], and even a few unicellular eukaryotes [e.g., the microalga *Chlamydomonas reinhardtii* (van Lis et al., 2020)]. Since the first publication in 1989 (Hagen et al., 1989), HCPs have been extensively studied by metal content determination (Moura et al., 1992; Pierik et al., 1992), different types of spectroscopy (Hagen et al., 1989; Moura et al., 1992; Pierik et al., 1992; Stokkermans et al., 1992; van den Berg et al., 1994; Marritt et al., 1995; Hagen et al., 1998; Kröckel et al., 1998; Tavares et al., 1998; Pereira et al., 1999; van den Berg et al., 2000; Macedo et al., 2002), redox titration (Pierik et al., 1992; Stokkermans et al., 1992; van den Berg et al., 1994,

2000), and X-ray crystallography (Cooper et al., 2000; Macedo et al., 2002, 2003; Aragão et al., 2003, 2008; Fujishiro et al., 2021), characterising the hybrid cluster in its different redox states. The cluster undergoes a complex chemistry with four oxidation states (reduced, semi-oxidised, oxidised, and super-oxidised, see Hagen, 2022) with distinct spectroscopic properties and structures also recently observed in the carbon monoxide dehydrogenase (CODH) type V. The CODH type V is homologous to HCP and shares the cluster chemical composition and a similar overall fold whilst increasing in complexity (Jeoung et al., 2022). The most studied HCPs are those found in *Desulfovibrio vulgaris* and *Desulfovibrio desulfuricans* (named below DvHCP and DdHCP, respectively), followed by the isoform found in *Escherichia coli* (EcHCP). The data gathered on enzymes from other organisms remain scarce.

Despite the undertaken efforts, the physiological function of HCPs is still elusive 30 years after their discovery. Various studies showed that the protein is produced in the presence of nitrate/nitrite (van den Berg et al., 2000; Kim et al., 2003; Haveman et al., 2004; Beliaev et al., 2005; Filenko et al., 2005; He et al., 2006; Cadby et al., 2017), sulphate (Aragão et al., 2003), oxidative stress (Briolat and Reysset, 2002; Almeida et al., 2006), or during pathogenic infection (Okinaka et al., 2002; Kim et al., 2003; Roos and Klemm, 2006). HCP was initially shown to reduce hydroxylamine (a side product of nitrate/nitrite reduction) to ammonia, which would allow its assimilation (Wolfe et al., 2002). Hydroxylamine reduction was confirmed in different HCPs (Wolfe et al., 2002; Aragão et al., 2003; Cabello et al., 2004; Overeijnder et al., 2009), and the protein was shown to enhance cellular resistance to nitrite and hydroxylamine (Cabello et al., 2004; Yurkiw et al., 2012), allowing nitrogen acquisition (Cabello et al., 2004). Nevertheless, the kinetic parameters of the enzyme (K_M in the millimolar range, low activity compared to other enzymes, especially at physiological pH) are incompatible with a physiological hydroxylamine reductase function (Almeida et al., 2006; Cole, 2021; Hagen, 2022). The activity is considered a side reaction, as it has been shown for other enzymes, including the Ni-dependent CODH homologous to HCPs (Mager, 1960; Faeder et al., 1974; Ostrowski et al., 1989; Heo et al., 2002; Hagen, 2022; Jeoung et al., 2022). A peroxidase function was later demonstrated (Almeida et al., 2006), but the gene expression was not systematically correlated during oxidative stress (Zhang et al., 2006; Wang et al., 2022), and the kinetic parameters (catalytic turnover ranging from 0.05 to 0.17 s⁻¹, concomitant production of more efficient peroxidases in the culture conditions) again appeared incoherent with a physiological peroxidase function (Cole, 2021; Hagen, 2022). More recently, the HCP was shown to act as high-affinity nitric oxide (NO)-reductase, detoxifying NO into nitrous oxide N₂O at physiological concentrations, and its overexpression protects cells against nitrosative stress (Kim et al., 2003; Wang et al., 2016; Hagen, 2019). Accordingly, HCP production depends on a NO-inducible promoter in some organisms, in agreement with a possible role in the nitrate/nitrite reduction pathway as the nitrite reduction is a source of NO (Filenko et al., 2007; Spiro, 2007; Wang et al., 2016; Bulot et al., 2019; Gao et al., 2019; Yu et al., 2020). The similar effects/co-occurrence of reactive oxygen species, NO/nitrosylative stress, and the promiscuity of NO for oxidative stress sensors would explain the production of HCP during oxidative stress (Filenko et al., 2007). The NO reductase activity was later amended by a role in protein transnitrosylation (Seth et al., 2018) in which

HCP generates R-S-nitrosothiol from NO. If the exact process of HCP-dependent protein nitrosylation has not been fully unveiled yet [and is still debated (Cole, 2021; Hagen, 2022)], spectroscopic analysis confirmed the existence of a dinitrosyl form of the hybrid cluster, thought to be an intermediate during NO reduction to N₂O, that could represent a nitrosyl donor for nitrosylation (Hagen, 2019). Finally, an HCP study on a photosynthetic microalga speculates a nitrate-reducing S-nitrosylase function that has to be experimentally proven (van Lis et al., 2020). The HCP physiological function appears to be on the path to being solved. Yet, alternative functions might exist depending on the organism, the growth conditions, or the HCP class/isoform.

Three HCP classes were historically sorted based on their sequences (van den Berg et al., 2000; Hagen, 2022). Classes I and II differ by a 4-residue insertion in the cubane [4Fe-4S] cluster binding motif, whilst class III exhibits a reduced N-terminal domain. Recent phylogenetic analysis pointed out that these classes do not form phylogenetic groups: if class III forms a branch distinct from both other classes, class II appears to be a subgroup of class I (van Lis et al., 2020). The same study suggested that class III even precedes the separation of the life domains and would be present in the last universal common ancestor (LUCA) (van Lis et al., 2020; Hagen, 2022; Jeoung et al., 2022). The gene coding for class II HCP, such as EcHCP, is often found in the genome next to a gene coding for the HCP reductase, a NAD(P)H oxidising flavoenzyme involved in oxidoreduction of the HCP and whom equivalent in class I and III is not known (van den Berg et al., 2000; Hagen, 2022). Another noticeable difference between HCP classes is that HCP class III appears to be dimeric, whilst all characterised classes I and II were shown to be monomeric [with the exception of one report in the case of nitrosylated enzyme (Seth et al., 2018)].

Many archaeal genomes harbour a gene coding for a putative HCP belonging exclusively to class I or III (van Lis et al., 2020). The sole archaeal HCP biochemically studied is from *Pyrococcus furiosus* and belongs to class III (Overeijnder et al., 2009). Noteworthy, the only available structure of an archaeal HCP is from the thermophilic methanogen *Methanothermobacter marburgensis*. According to our knowledge, this class III representative [available in the PDB database under ID 7E0L (Fujishiro and Takaoka, 2022)] has yet to be described in any publication. This structure will not be used for more than sequences and comparisons to respect the authors' study.

The presence of HCP in methanogens is unexpected as most of these ancient organisms are not known to use nitrate, nitrite, or NO as these oxidants are toxic to them (Klüber and Conrad, 1998). Recent transcriptomics studies on the nitrogen metabolism from the thermophilic archaeon *Methanothermococcus thermolithotrophicus* have been investigated by our group and showed a constitutively rather high expression of an HCP class I at a level comparable to that of genes coding for enzymes involved in methanogenesis (Maslać et al., 2022). In this study, we purified the HCP from *M. thermolithotrophicus* grown on ammonia as the only nitrogen source and obtained its structure in reduced and semi-oxidised states. The conservation in architecture, active site, reactions, and internal tunnelling systems is discussed. Phylogenetic analyses show that the HCPs from *Methanococcales* species are closer to bacterial HCPs than to any archaeal enzymes, which further questions their physiological functions in methanogens.

2. Materials and methods

2.1. Strains and growth conditions

Methanothermococcus thermolithotrophicus strain DSM 2095 was obtained from the Deutsche Sammlung von Mikroorganismen und Zellkulturen (DSMZ, Braunschweig, Germany). The archaeon was grown on ammonia and sulphate in a fermenter at 65°C, as described in [Jespersen et al. \(2023\)](#).

2.2. Protein purification

The purification of the native *MtHCP* was originally performed based on its colour and the natural abundance of the protein. It was purified several times with a similar, reproducible protocol. The optimal one is described below. The heterologous production of the enzyme was not attempted.

Cell lysis was performed under an N₂/CO₂ (90:10%) atmosphere. About 19 g (wet weight) of frozen cells were thawed and diluted with four volumes of lysis buffer (50 mM tricine/NaOH pH 8.0, 2 mM dithiothreitol (DTT)), sonicated (3 × 10 s at 70% power, probe KE76 Bandelin SONOPULS Berlin, Germany), and centrifuged for 1 h at 45,000 × g at 18°C. The following overall procedure was performed under yellow light and at 20°C. The supernatant was transferred to an anaerobic Coy tent containing an N₂/H₂ (97:3%) atmosphere, filtered through a 0.2 µm filter (Sartorius, Göttingen, Germany), and loaded on a 20 ml HiTrap DEAE-Sepharose fast flow column (GE Healthcare, Munich, Germany). Elution was performed with a 100–600 mM NaCl linear gradient for 15 column volumes (CVs) at a flow rate of 2.5 ml/min. The sample of interest eluted between 125 and 310 mM NaCl. The resulting pool was diluted with four volumes of lysis buffer, filtered, and loaded on a 15 ml HiTrap Q-Sepharose High-Performance column (GE Healthcare, Munich, Germany). Elution was performed with a linear gradient ranging from 0 to 500 mM NaCl for 13.33 CV at a flow rate of 2 ml/min. The fractions of interest was eluted between 248 and 320 mM NaCl. The resulting pooled fraction was diluted with four volumes of 25 mM Tris/HCl pH 7.6, 2 M (NH₄)₂SO₄, and 2 mM DTT, filtered through a 0.2 µm filter (Sartorius, Göttingen, Germany), and injected on a 5 ml HiTrap Phenyl Sepharose High-Performance column (GE Healthcare, Munich, Germany). The sample was eluted with a 1.8 to 0 M linear gradient of (NH₄)₂SO₄ for 24 CV at a flow rate of 1.5 ml/min. The protein was eluted between 0.41 and 0.25 M (NH₄)₂SO₄. After concentration up to 400 µl with a 10-kDa cutoff centrifugation filter (Merck, Darmstadt, Germany), contaminants were separated by size-exclusion chromatography on a Superose 6 Increase 10/300 GL (GE Healthcare, Munich, Germany) in 25 mM Tris/HCl pH 7.6, 2 mM DTT, and 10% (v/v) glycerol at a flow rate of 0.4 ml/min. The final pool was concentrated with a 10-kDa cutoff centrifugation filter (Merck, Darmstadt, Germany), and protein concentration was estimated by the Bradford method (Bio-Rad Laboratories, Munich, Germany). Each purification step was systematically controlled by denaturing sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE). The protein was directly used for crystallisation or aliquoted and stored frozen at –80°C.

2.3. Protein crystallisation

Crystallisation was performed anaerobically by initial screening at 20°C using the sitting drop method on 96-Well MRC 2-Drop polystyrene Crystallisation Plates (SWISSCI) in a Coy tent containing an N₂/H₂ (97:3%) atmosphere. The reservoir chamber was filled with 90 µl of crystallisation condition (JBScreen Wizard crystallisation screen, Jena Bioscience), and the crystallisation drop was formed by spotting 0.55 µl of purified protein with 0.55 µl of precipitant. Both structures were obtained from the same drop, opened several months apart. The protein was crystallised at 9.9 mg/ml in a solution containing 20% (w/v) PEG 3,350 and 200 mM magnesium formate. Densities in the electron density map suggest contamination from another crystallisation condition spatially close and containing 30% (v/v) 2-methyl-2,4-pentanediol, 20 mM calcium chloride, and 100 mM sodium acetate, pH 4.6.

For the reduced state, the crystals were soaked in the crystallisation solution supplemented with 20% (v/v) glycerol for a few seconds before freezing in liquid nitrogen. For the reduced/oxidised state, the crystals were soaked in the crystallisation solution supplemented by 100 mM hydroxylamine/HCl for 5.7 min and then transferred in the crystallisation solution supplemented with 20% (v/v) ethylene glycol for a few seconds before freezing in liquid nitrogen.

2.4. Data collection and structural analysis

The diffraction experiments were performed at 100 K on the beamline P11 from DESY and PXIII (X06DA) from the Swiss Light Source (SLS). The data were processed and scaled with *autoPROC* ([Vonrhein et al., 2011](#)). The data of the structure of *MtHCP* in the reduced/oxidised state presented anisotropy (along the following axes: $a = 1.451 \text{ \AA}$, $b = 1.343 \text{ \AA}$, and $c = 1.748 \text{ \AA}$) and was further processed with *STARANISO* correction integrated with the *autoPROC* pipeline ([Tickle et al., 2018](#)) (*STARANISO*. Cambridge, United Kingdom: Global Phasing Ltd.).

The structure of the reduced *MtHCP* was solved by performing a single anomalous dispersion experiment at the Fe K-edge ([Supplementary Table S1](#)) using the *SHELX* package ([Usón and Sheldrick, 2018](#)). The first model was then refined by extending the resolution to 1.45 Å by a second dataset. The reduced state was used as a template to solve the reduced/oxidised state structure by molecular replacement with *PHASER* from the *PHENIX* package ([Liebschner et al., 2019](#)). The resolution limit was determined based on the obtained electron density map quality and R_pim as major criteria. All models were manually built *via* COOT ([Emsley et al., 2010](#)) and refined with *PHENIX* (version 1.20.1-4487). The last refinement steps were performed by refining with translation libration screw (TLS) for the reduced state and considering all atoms anisotropic for the reduced/oxidised state and were validated by the MolProbity server ([Chen et al., 2010](#))¹. Both models were refined with hydrogens in the riding position. Hydrogens were omitted in the final deposited models. The PDB ID codes of the structures are 8CNR and 8CNS for the *MtHCP*_{red} and *MtHCP*_{mix} structures, respectively. Data collection

¹ <http://molprobity.biochem.duke.edu>

and refinement statistics for the deposited models are listed in [Supplementary Table S1](#). All figures were generated and rendered with PyMOL (Version 2.2.0, Schrödinger, LLC, New York, NY, United States). The structural superpositions in [Supplementary Table S2](#) and [Supplementary Figure S3](#) were performed with PyMOL or the secondary-structure matching (SSM) tool from the CCP4 suite (Krissinel and Henrick, 2004; Winn et al., 2011). For the elaboration of [Supplementary Figure S3](#), the B-factor column of 8CNR was replaced by the residues deviation obtained from SSM by the PyMOL script `data2bfactor.py` written by Robert L. Campbell. The protein was then coloured by residues deviation. Non-aligned residues, residues with aberrant deviation, and fragments of three residues or less were omitted for clarity. Internal tunnel predictions were performed by the CAVER tool (Kozlikova et al., 2014) by applying a probe radius of 1.0 Å and starting from the C β of the cysteine at position 455 (*MtHCP*) or the equivalent cysteine.

2.5. High-resolution clear native PAGE (hrCN PAGE) and size-exclusion chromatography

The hrCN PAGE protocol was adapted from [Lemaire et al. \(2018\)](#). Glycerol was added to the sample at a final amount of 20% (v/v). Ponceau S at a final concentration of 0.001% (w/v) served as a marker to follow the migration. The buffer composition for the electrophoresis cathode was the following: 50 mM tricine, 15 mM Bis-Tris/HCl, pH 7.0, 0.05% (w/v) sodium deoxycholate, 2 mM DTT, and 0.01% (w/v) dodecyl maltoside, whilst the anode buffer contained 50 mM Bis-Tris/HCl, pH 7.0, 2 mM DTT. An 8–15% linear polyacrylamide gradient gel was used, and electrophoresis was run under a N₂/CO₂ (90:10%) atmosphere with a constant 40 mA current (PowerPac™ Basic Power Supply, Bio-Rad). After electrophoresis, protein bands were visualised with Ready Blue™ Protein Gel stain (Sigma Aldrich, Hamburg, Germany). The native protein ladder used is NativeMark™ Unstained Protein Standard (Thermo Fischer Scientific, Driesch, Germany).

The determination of the oligomeric state by gel filtration was performed in triplicate on a Superose 6 Increase 10/300 GL (GE Healthcare, Munich, Germany) in 25 mM Tris/HCl pH 7.6, 2 mM DTT, 10% (v/v) glycerol at a flow rate of 0.4 ml/min, and in an anaerobic Coy tent containing an N₂/H₂ (97:3%) atmosphere. High molecular weight range gel filtration calibration kit (GE Healthcare, Munich, Germany) was used as the protein standard.

2.6. Absorbance spectra

The protein absorbance spectra were monitored aerobically and anaerobically. The aerobic oxidised spectrum was monitored using a Cary 60 UV-Vis spectrophotometer (Agilent Technologies). Absorbance spectra were measured by scanning the wavelength from 250 to 600 nm, with a measurement every 0.5 nm, in a 1 mm-path length TrayCell (Hellma Analytics) at room temperature. The protein concentration was 26.4 mg/ml, and the baseline absorbance spectrum of the buffer was subtracted from the measured protein spectrum. The anaerobic spectra were monitored on a BMG Labtech FLUOstar Omega Microplate reader

in an anaerobic chamber filled with an N₂ (100%) atmosphere at room temperature. The protein concentration was 3.4 mg/ml, and the baseline absorbance spectrum of the buffer was subtracted from the measured protein spectrum. The spectrum was monitored on a 384 wells transparent plate using an 18 μ l sample volume. The protein was reduced or oxidised by the addition of 200 μ M of sodium dithionite or potassium ferricyanide, respectively.

2.7. Enzymatic assays

The hydroxylamine reductase assay was performed according to the protocol set up by [Wolfe et al. \(2002\)](#). The activity was monitored by the enzyme-dependent decrease in the absorbance of methyl viologen (MV) at 600 nm on a BMG Labtech FLUOstar Omega Microplate reader in an anaerobic chamber filled with an N₂ (100%) atmosphere. For practical reasons (physical limits of the equipment and limitation of the evaporation), activities were monitored at 50°C. Measurements were acquired in 100 mM CHES/NaOH pH 9.0 or in 50 mM HEPES/HCl pH 7.35 with the addition of 10 μ M EDTA, 10 mM MV, and 75 μ M sodium dithionite. The final protein concentration was 26.4 μ g/ml at pH 7.35 and 2.64 μ g/ml at pH 9.0 (200 μ l final volume), and the tested hydroxylamine concentration ranged from 0 to 105 mM final. At pH 9.0, activities with a hydroxylamine concentration above 15 mM could not be reliably measured as it appears to induce reproducible drops in enzyme activity, probably due to the combined high temperature and pH. Enzyme-independent viologen oxidation after hydroxylamine addition, proportional to hydroxylamine concentration, had to be subtracted for every measurement. The molar extinction coefficient of MV used for calculation was determined experimentally in our working conditions (12,791 M⁻¹. cm⁻¹ and 13,489 M⁻¹. cm⁻¹ at pH 7.35 and 9.0, respectively). The activity measurements at both pHs can be fitted with a similar coefficient of determination using a Michaelis–Menten or a Hill equation, yet at pH 7.35, the Hill equation appears to be a better fit ($R^2 = 0.963$ vs. 0.932 for the Michaelis–Menten equation). No cooperativity mechanism has been observed or proposed for other HCPs, and there is no experimental evidence for a reaction mechanism in HCP that would deviate from a Michaelis–Menten equation. The difference in R^2 observed in the fits of our experiments is insufficient to question the proposed reaction mechanism in HCP, and the kinetic parameters used in the text were determined assuming a Michaelis–Menten equation. All measurements have been done in triplicates using a protein obtained from a unique purification and are presented in μ mol of reduced hydroxylamine per minute per mg of protein, considering a mole of hydroxylamine reduced for two moles of viologen oxidised.

The peroxidase assay was performed aerobically according to the protocol set up by [Almeida et al. \(2006\)](#). The activity was monitored by following the enzyme-dependent decrease in the absorbance of sodium ascorbate at 290 nm ($\epsilon_{290\text{nm}}$ of 2,800 M⁻¹. cm⁻¹) due to the oxidation of ascorbate by hydrogen peroxidase reduction, using quartz cuvettes and a Cary 60 UV-Vis spectrophotometer (Agilent Technologies). The reaction mixture (400 μ l) contained 50 mM potassium phosphate buffer, pH 7.0, and 0.2 mM sodium ascorbate. The experiment was assessed at 30, 40, 50, and 60°C. The used hydrogen peroxide concentration varied from 1 mM to 10 mM, and the *MtHCP* concentration varied from 3.3 μ g/ml to 132 μ g/ml.

Appropriate controls were performed for each assay and showed that no enzyme-dependent significant decrease in absorbance could be detected in any reproducible assay.

2.8. Sequence alignments and phylogenetic analysis

The protein sequence of *MtHCP* was used as a query for BLAST (Altschul et al., 1997) in Protein Data Bank (PDB) and in the RefSeq database. The RefSeq database was selected to limit species redundancy and because it usually contains genomes of non-questionable quality. The search in the PDB extracted the sequences of structurally characterised HCPs (Cooper et al., 2000; Macedo et al., 2002; Aragão et al., 2003; Macedo et al., 2003; Fujishiro et al., 2021; Fujishiro and Takaoka, 2022). The maximal E-value (9×10^{-122}) was with *MmHCP* (PDB 7E0L). The searches in the RefSeq were set up to extract 500 sequences without additional parameters for the first research (maximal E-value 0.0) and limited to the *Archaea* kingdom for the second (maximal E-value for HCP, 1×10^{-108} , E-value for CODH 1×10^{-13}). The Supplementary Figure S2 was constructed using the sequence of *MtHCP* and the sequences extracted from the BLAST search in the PDB. It was designed with ESPript 3 (Robert and Gouet, 2014) by using an alignment constructed with Clustal Omega (Sievers et al., 2011).

The phylogenetic trees were constructed using the maximum likelihood method and were generated with the MEGA programme (Kumar et al., 2018) by using an alignment constructed with MUSCLE (Edgar, 2004). A total of 200 replicates were used to calculate each node score. The first presented tree was constructed with the sequences from the non-limited search in the RefSeq databank and the results from the PDB. The second tree was constructed with the sequences from the search in the RefSeq databank limited to *Archaea*, in which the sequences corresponding to CODH were manually deleted except one (CODH α from *Methanospirillum hungatei*, WP_011448292.1) and the results from the PDB. The tree shown in Supplementary Figure S11 was constructed with the sequences obtained from the three BLAST results (in the PDB and in the RefSeq database, non-restricted, and restricted to *Archaea* in which CODH sequences were deleted).

3. Results

3.1. *MtHCP* overall architecture

The native HCP from the thermophilic methanogen *M. thermolithotrophicus* (*MtHCP*) was purified anaerobically by multiple chromatography steps and immediately crystallised without freezing. The purified brown protein fraction presents a major band of approximately 60 kDa (Da) on SDS-PAGE (Figure 1A, *MtHCP* expected molecular weight of 59,665.17 Da). The native electrophoresis and gel filtration mostly fit a monomeric organisation of the enzyme (Figures 1A,B). The spectrophotometric absorbance spectrum is typical of the [4Fe-4S] cluster-containing proteins and similar to previously published spectra of HCPs (Moura et al., 1992; Pierik et al., 1992; Stokkermans et al., 1992; Pereira et al., 1999; van den Berg et al., 2000; Figure 1C). Even if purified anaerobically and in the presence of dithiothreitol in the storage buffer,

the spectrum of the frozen purified enzyme suggests an oxidised state that might be due to the long storage at -80°C (Figure 1C). The crystal structure was solved by single-wavelength anomalous dispersion at the Fe K-edge (SAD) and refined to a 1.45-Å resolution (see Materials and methods, Figure 1D, Supplementary Figures S1A,B, and Supplementary Table S1). *MtHCP* is a monomeric 548-amino acid peptide organised in three domains (first domain residues 1–222, second domain 223–368, and third domain 369–548) harbouring a cubane [4Fe-4S] cluster and the hybrid cluster (Figure 1D). Based on the nomenclature, *MtHCP* would belong to class I, coherently with cluster binding motif and its monomeric state confirmed in the crystal structure (Supplementary Table S2 and Supplementary Figure S2). Five extra residues, not expected in the available sequence (due to the incorrect prediction of the initiating methionine, WP_026182881.1), span from the N-terminal. The overall enzyme architecture is highly similar to the available HCP structures from bacteria (Supplementary Figure S3 and Supplementary Table S2). The major differences in structures are located on peripheral loops and helices. The well-defined electron density shows no sign of nitrosylation in the active site surroundings.

Differences in surface charges distribution were noticed between HCP class I for *Desulfovibrio* species and class II for *E. coli* (Fujishiro et al., 2021). An inspection of the surface charges in *MtHCP* nuances the contrasts between classes I and II (Figure 2). *MtHCP* exhibits a large negatively charged area on its surface surrounding a ~ 17 -Å invagination towards the [4Fe-4S] cluster (Figure 2), also conserved in the bacterial structures. The cavity is polar, filled with water molecules in all structures, and exhibits a positively charged depth close to the cluster (Figure 2). It is interrupted in the structure of *EcHCP* by the tryptophan at position 193 (Figure 2D) in a loop that is displaced in class I HCP structures. However, considering the loose electron density for this residue, the negatively charged cavity might be open to the solvent (modelled in Figure 2E). Furthermore, the residue is not conserved in all class II HCPs.

3.2. Structure of the active site in reduced and oxidised states

The structure harbours a hybrid cluster in a [4Fe-3S] composition, matching the bacterial homologues in their reduced state (Figure 3 and Supplementary Figure S1). The active sites of *MtHCP* and the 1.25-Å resolution structure of the reduced *DdHCP* are superimposable. Notably, residues surrounding the hybrid cluster are remarkably conserved between the sequences from *Desulfovibrio* species and *M. thermolithotrophicus* (Figure 3 and Supplementary Figure S2). The main difference is a 1.0-Å shift of the “Y-atom” at the hybrid cluster vicinity, which has also been attributed to an oxygen atom in *MtHCP* based on the electron density (Figure 3). Because of the unequivocal chemical composition [4Fe 2 μ 3-S 2 μ 2-S (O)] of the hybrid cluster, we propose that the as-isolated *MtHCP* represents the reduced state (Supplementary Figure S1C, the nomenclature *MtHCP*_{red} will be used below). The difference between the reduced state in the structure and the oxidised state observed in spectrophotometry could come from the prolonged storage freezing at -80°C that might have led to the oxidation of the protein (Figure 1C). Alternatively, the crystallisation may have conserved or triggered the cluster reduction.

In order to obtain a different state of the active site or an intermediate of reaction, crystals were soaked in a solution of hydroxylamine (non-physiological substrate, 100 mM final) for a few minutes under

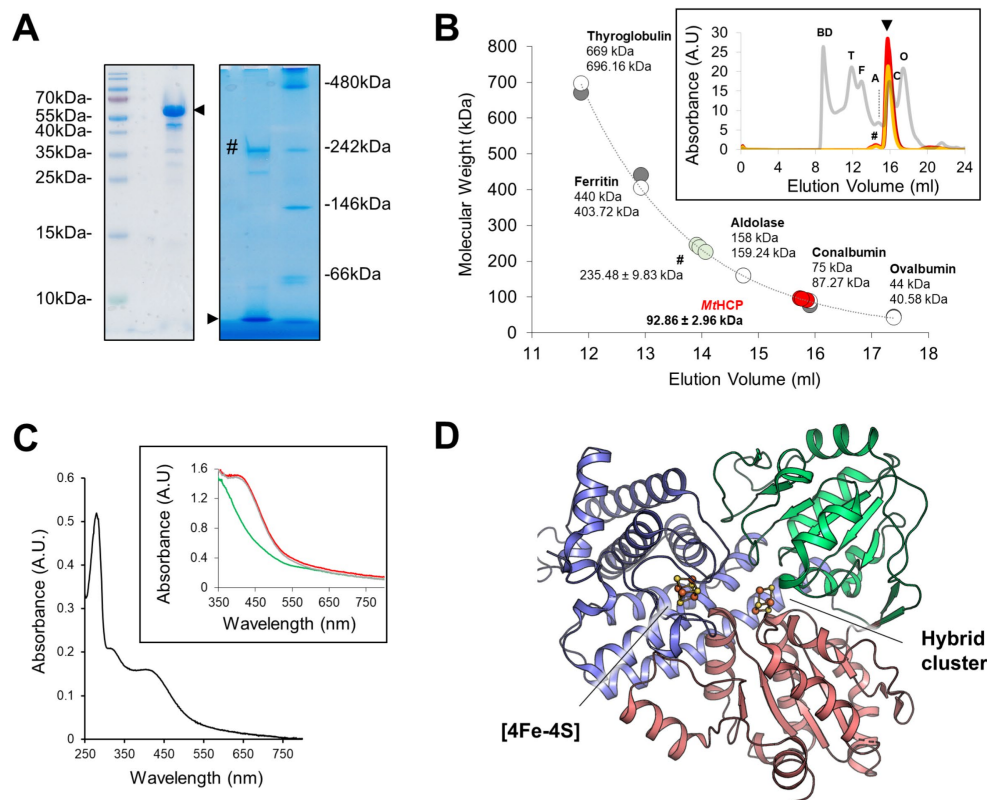


FIGURE 1

Oligomeric state, absorbance spectrum, and MthCP overall structure. (A) SDS-PAGE (left) and high-resolution Clear Native PAGE (right) profiles of the enriched MthCP fraction. About $3\mu\text{g}$ of the sample kept frozen were loaded on gels. A black arrow indicates the position of MthCP, and a '#' symbol marks the main contaminant of MthCP. (B) Molecular weight estimation by size-exclusion chromatography. Protein standards (GE Healthcare, grey line in the insert) and three samples of MthCP (deep red, red, and orange lines in the insert) are overlaid. The measured elution volume of the standard is plotted (grey dots, upper sizes), and the extracted exponential curve ($y=310753e^{-0.514x}$, $R^2=0.9927$) was used for molecular weight estimation. The theoretical size of the standard proteins (white dots, smaller sizes) is given for reference. BD stands for blue dextran, T for thyroglobulin, F for ferritin, A for aldolase, C for conalbumin, and O for ovalbumin. (C) Absorbance spectra of MthCP after freezing and storage at -80°C . The aerobic spectrum is coloured black in the main graph. Anaerobic measurements as isolated (red line), dithionite-reduced (green line), and ferricyanide-oxidised (grey line) are shown in the insert. (D) Overall structure of the MthCP. The protein is presented in cartoon coloured by domains (N-terminal, central, and C-terminal domains being coloured blue, green, and red, respectively) with the clusters presented as ball and stick. Sulphur and iron are coloured yellow and orange, respectively.

anaerobic conditions before freezing in liquid nitrogen. The crystal structure was refined to $1.36\text{-}\text{\AA}$ resolution (Supplementary Table S1), and the protein backbone was superimposed closely to the MthCP_{red} structure (Supplementary Table S2). The hybrid cluster appeared to have both reduced and oxidised/semi-oxidised states. Therefore, the structure is referred to as MthCP_{mix} (Figure 4 and Supplementary Figure S4). A similar case of redox state mixture was reported for the HCP of *D. desulfuricans* (Macedo et al., 2003) (PDB code 1UPX). Both states of the hybrid cluster can be appropriately modelled due to the previous structural characterisation of HCPs. The reduced state is similar to the MthCP_{red} structure, with a noticeable difference in the glutamate 490 side chain position (Figures 4B,C). The carboxy group of the glutamate rotates 60° and disengages from the Fe atom (Fe number 8), leaving the Fe with three covalent bonds (Supplementary Figure S5). The oxidised/semi-oxidised state that contains a $[4\text{Fe } \mu_3\text{-S } \mu_2\text{-S, a persulphido and } 3\mu_2\text{-O}]$ hybrid cluster fits very well with the ones observed in bacterial homologues. Here, the Fe atom 8 (according to the numbering from Hagen, 2022) is coordinated by two oxygen atoms, the glutamate 490 and the persulphido-cysteine 402 (Figures 4D,E). This "extracted" state will be referred to as MthCP_{ox} below.

The partial oxidation of the active site gives an opportunity to describe the conformational changes between reduced and oxidised states in MthCP (Figure 5). As for bacterial enzymes, the major changes are the position of Fe8 and S7 that undergo large movement (1.8 and 3.9 \AA , respectively). Assuming a unique sulphur atom, the S7 position between Fe5 and Fe7 is replaced by oxygen and forms a persulphido group on the cysteine 402. The Fe8 loses the bonding with cysteine 402 and an oxygen bond the Fe6 and Fe8.

3.3. MthCP enzymatic activities

Because of their structural similarity, we aimed to compare the enzymatic properties of MthCP with that of previously characterised HCPs. The hydroxylamine reductase activity was assessed at 50°C , at a pH of 7.35, close to the physiological pH (Jarrell and Sprott, 1981), but also at pH 9.0 as tested in other studies, enabling a higher reaction turnover (Wolfe et al., 2002; Aragão et al., 2003; Cabello et al., 2004; Overijnder et al., 2009). At pH 7.35, we determined an apparent K_M of $2.2 \pm 0.5\text{ mM}$, an apparent maximal velocity approximately

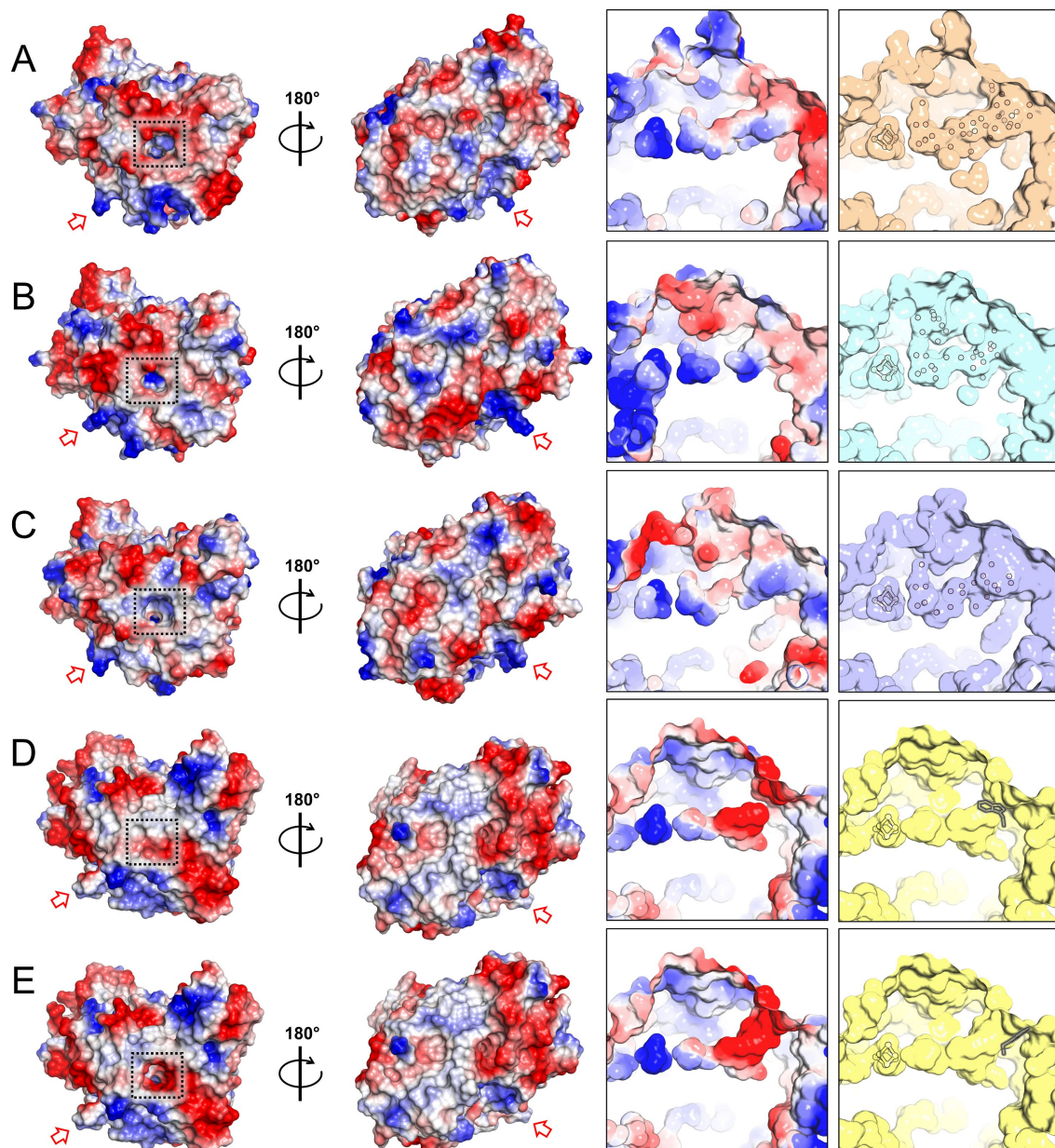


FIGURE 2

Charge distribution and access to the [4Fe-4S] cluster in HCP structures. (A–E) Charge distribution on the protein surface (left) and on the polar cavity near the cluster (right). Left: the proteins are shown in two orientations. A dashed black square highlights the entry of the polar cavities and the red arrow points to the loop proposed to interact with the partner in EcHCP and its equivalent in other structures. Right panel: surface of the polar cavity and filled water molecules. The water molecules, [4Fe-4S] cluster, and tryptophan 193 (framed in grey) of EcHCP are shown as balls and sticks. The structures presented are *MthHCP_{red}* (PDB 8CNR, A), *DdHCP_{red}* (PDB 1OA0, B), *DvHCP_{red}* (PDB 1OA1, C), *EcHCP* (PDB 7DE4, D), and the *EcHCP* structure with a different conformation of tryptophan 193 (E).

$0.072 \pm 0.004 \mu\text{mol}$ of hydroxylamine reduced/min/mg of purified HCP, and a turnover of 0.072 s^{-1} (Supplementary Figure S6). At pH 9.0, an apparent K_M of $3.9 \pm 0.9 \text{ mM}$, V_{max} of $1.660 \pm 0.157 \mu\text{mol}$ of hydroxylamine reduced/min/mg of purified HCP, and a turnover of 1.667 s^{-1} were obtained. Therefore, the enzyme exhibits an activity approximately 20–25-fold higher at pH 9.0, similar to the enzyme of *E. coli* (Wolfe et al., 2002) and in the range of other HCPs (Cabello et al., 2004; Overeijnder et al., 2009). However, purified *MthHCP* exhibits a comparable K_M , in the millimolar range, at both pHs, unlike the enzyme from *E. coli*. The kinetic parameters of the enzyme appear

to be unsuitable with a physiological reduction of hydroxylamine, as for other HCPs (Hagen, 2022).

The peroxidase activity of *MthHCP* was also tested but could not be monitored with the protocol used by Almeida and colleagues (Almeida et al., 2006). No significant enzyme-dependent ascorbate reduction could be recorded even at an enzyme concentration of $20 \mu\text{M}$ final. A significant enzyme-independent decrease in ascorbate absorbance occurred, especially at higher temperatures. As the specific peroxidase activities previously measured using *EcHCP* and *DdHCP* were relatively low [0.17 and $0.05 \mu\text{mol} \cdot \text{min}^{-1} \cdot \text{mg}^{-1}$, respectively

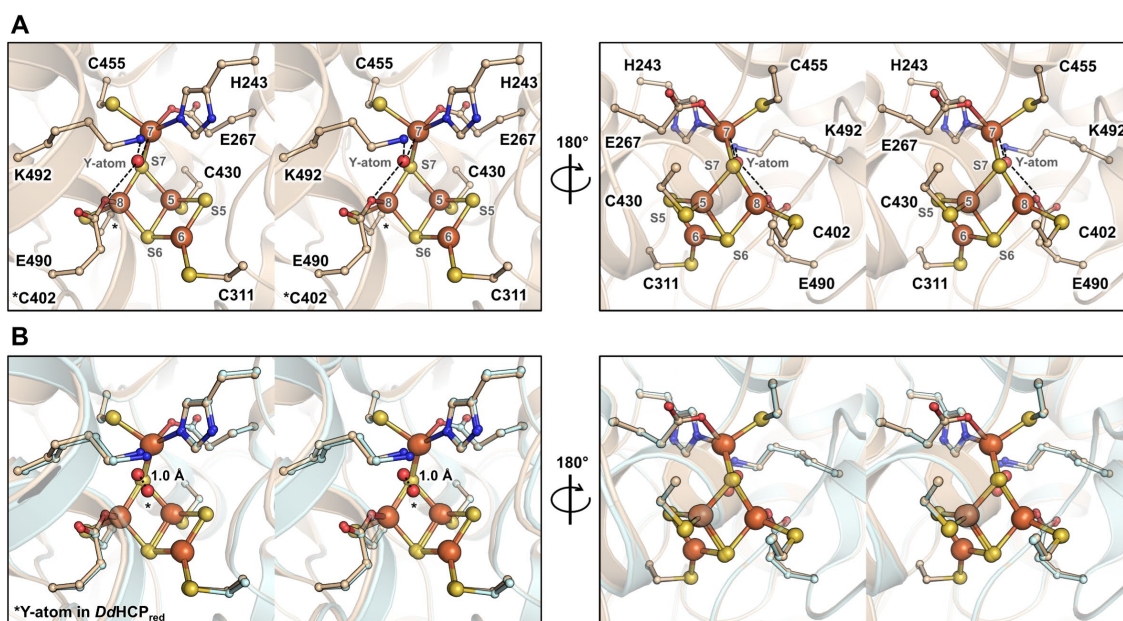


FIGURE 3

Overall organisation and active site of the isolated MtHCP. (A,B) Stereo representation of the active site in two different orientations in the structures of MtHCP_{red} (A) and superimposition of the structures of MtHCP_{red} and DdHCP_{red} (PDB 10A0, B). The residues coordinating or interacting with the cluster are shown as balls and sticks. The structures are shown as transparent cartoons with the MtHCP structure coloured wheat and the DvHCP_{red} coloured light cyan. Oxygen, nitrogen, sulphur, and iron are coloured red, blue, yellow, and orange, respectively. The atom of the cluster is labelled according to previous studies (Hagen, 2022). The distance between the position of the Y-atom in MtHCP_{red} and DdHCP_{red} (modelled as oxygen in both structures) is shown as a dashed line and indicated in Ångström.

(Almeida et al., 2006)], a weak peroxidase activity of MtHCP, possibly overlooked in our experiments, cannot be ruled out.

3.4. Hydrophobic tunnels

The study of the first X-ray structure of an HCP already highlighted the presence of hydrophobic cavities within the enzyme structure (Cooper et al., 2000). Similarly to CODH (Lemaire and Wagner, 2021; Biester et al., 2022; Jeoung et al., 2022). A cavity analysis by the CAVER programme (Kozlikova et al., 2014) showed that the hydrophobic channels are also present in the MtHCP_{red} structure (Figure 6A). Five channels devoid of any water molecules were detected, spanning an overall distance of approximately 63-Å (tunnel length from the extremities of tunnels 1 and 5) within the enzyme from an edge to another. The channelling system connects the active site to several openings on the protein surface, and most of the side chains structuring the channels are hydrophobic (57/62, see Supplementary Figures S2, S7), probably for repelling charged or polar molecules that would disturb the reaction. This is in agreement with the absence of water molecules in these tunnels. Depending on the tunnel, the bottleneck exhibits a radius ranging from 1.24 to 1.32 Å (Figure 5A). Analogous tunnels are predicted in the MtHCP_{mix} structure, coherent with the structural conservation in both structures (Figures 6B–D and Supplementary Table S2). Noteworthy, tunnel 2 is obstructed by a molecule that we modelled as 2-methyl-pentane-2,4-diol based on the electron density (Supplementary Figure S8). As it is probably a non-physiological artefact of crystallisation, it was ignored

for cavity prediction. Virtually identical tunnels are predicted in the structures of HCP homologues, with various additional extensions of tunnel 5 depending on the model used for the analysis (Figure 6 and Supplementary Figure S9). Most of these tunnels are conserved in the class II EcHCP, but the slight difference in the position of the peptide chain (Supplementary Figure S3 and Supplementary Table S2) offsets the predicted channels (Figure 6).

As previously seen in CODH type V (Jeoung et al., 2022), the redox state of the hybrid cluster affects its access (Supplementary Figures S10A,B). Once in the oxidised state, the subtle changes in the hybrid cluster generate a so-called pinball effect to provoke the expulsion of the product, restraining the access to the cluster. MtHCP obeys the same rule, narrowing the access to the hybrid cluster due to the formation of the persulphido-cysteine 402 (Supplementary Figures S10A,B). A similar movement occurs in DdHCP and DvHCP (Supplementary Figures S10C,D).

3.5. Phylogenetic position of MtHCP

According to van Lis et al. (2020) study, most of the archaeal class I HCPs cluster in a single group containing bacterial sequences including DvHCP and DdHCP. This common branch suggests that this class of HCP was initially present in one of the domains and was acquired by the second via horizontal gene transfer (van Lis et al., 2020).

A phylogenetic tree was constructed with the available sequences of HCPs from the Protein Data Bank and the 500 closest sequences to

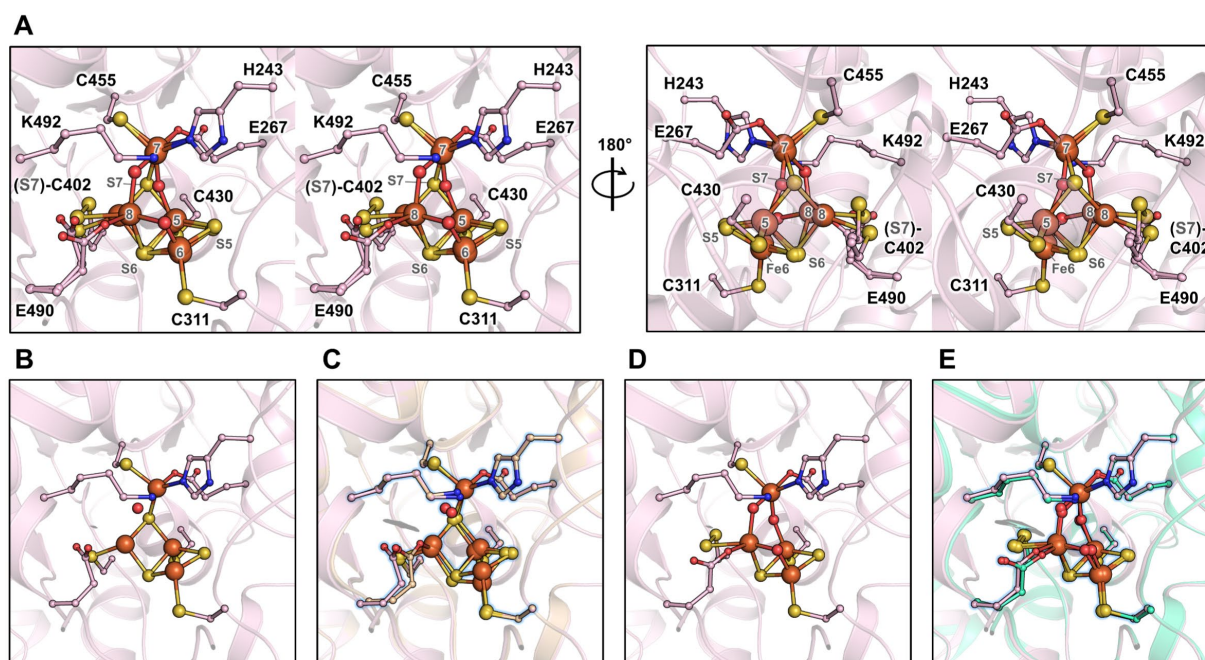


FIGURE 4

Active site of the MthCP in mixed redox state and the isolated states. (A) Stereo representation of the active site of the MthCP_{mix} containing a mixture of reduced and oxidised/semi-oxidised states. (B) Reduced state of the cluster extracted from the MthCP_{mix} structure. (C) Superimposition of the extracted structure of the reduced state and MthCP_{red}. (D) Oxidised/semi-oxidised state (MthCP_{ox}) of the cluster extracted from the MthCP_{mix} structure. (E) Superimposition of the MthCP_{ox} structure and the structure of the oxidised state of DdHCP (PDB 1GNL). (A–E) The structures are shown as transparent cartoons, with the hybrid cluster represented as spheres and the coordinating/interacting residues as balls and sticks. MthCP_{mix} and its extracted structures are coloured light pink, the MthCP_{red} structure is coloured wheat, and the DdHCP_{ox} structure is coloured green cyan. Oxygen, nitrogen, sulphur, and iron are coloured red, blue, yellow, and orange, respectively. (C,E) The extracted states of MthCP_{mix} are highlighted by a blue glow for clarity.

MthCP obtained by a BLAST analysis in the RefSeq database (Figure 7A). Over 500 obtained sequences, only 13 were from archaea, all from species belonging to the *Methanococcales* order. The remaining sequences were from bacterial class I enzymes belonging to different phyla, with a clear dominance of *Bacillota* (formerly *Firmicutes*). The archaeal sequences form a monophyletic group amongst the bacterial tree, and the sequences of the class II EchCP and the class III MmHCP branch as outgroups of the generated class I HCPs tree. If the precise position of the archaeal HCPs amongst the bacterial sequences cannot be extracted from this analysis, it appears that MthCP and the enzymes from other *Methanococcales* belong to class I, as suggested by their sequences. In order to verify the suspicious position of the enzymes from *Methanococcales*, closer to bacterial HCPs than other archaeal species, another phylogenetic tree was constructed. The sequences used to construct the second tree were extracted from a similar BLAST analysis but restricted to the *Archaea* domain. The extracted sequences consisted of 184 sequences of archaeal HCPs, the remaining being annotated as CO-dehydrogenases. A preliminary tree showed that the CODH sequences were forming a completely separated branch and were discarded, a single sequence being kept as an outgroup (Figure 7B). This second tree gathers every archaeal HCP sequence from the RefSeq database. The class III HCPs, including MmHCP and the enzyme from *P. furiosus*, form a separate branch, whilst the distinction between class I and II is unclear in this tree. This was expected because of the position of class II within the class I tree and because of the several ramifications of the class I phylogenetic tree (van Lis et al., 2020). The different branches mostly group organisms

belonging to the same archaeal class, and some signs of horizontal gene transfer can be observed. For instance, a sequence of a class I HCP from a species belonging to the *Thermoplasmata* phylum groups in a branch gathering class I HCPs from *Methanomicrobia* species (Figure 7B). Similarly, the unique class III HCP found in the *Methanomicrobia* phylum groups in branch gathering enzymes from *Methanobacteria* (Figure 7B). The branch corresponding to *Methanococcales* is intriguingly close to that containing the class I HCPs from *Halobacteria*, yet *Methanococcales* and *Halobacteria* belong to a different clade (Adam et al., 2017). Moreover, the *Methanococcales* branch also forms a relatively clear group (node score of 77) with the sequences of bacterial class I HCP from *D. vulgaris* and *D. desulfuricans* (Figure 7B). It confirms the results of the previous BLAST analysis suggesting that MthCP and the other enzyme from *Methanococcales* are closer to bacterial enzymes than to any other archaeal HCP. This tendency is the same when both trees are gathered (Supplementary Figure S11).

4. Discussion

4.1. The HCP from *Methanococcales* as a sister group of characterised bacterial class I

The hybrid cluster protein is encoded in the genome of many anaerobic bacteria, archaea, and unicellular eukaryotes, including common laboratory models. Yet its function is still unclear. The

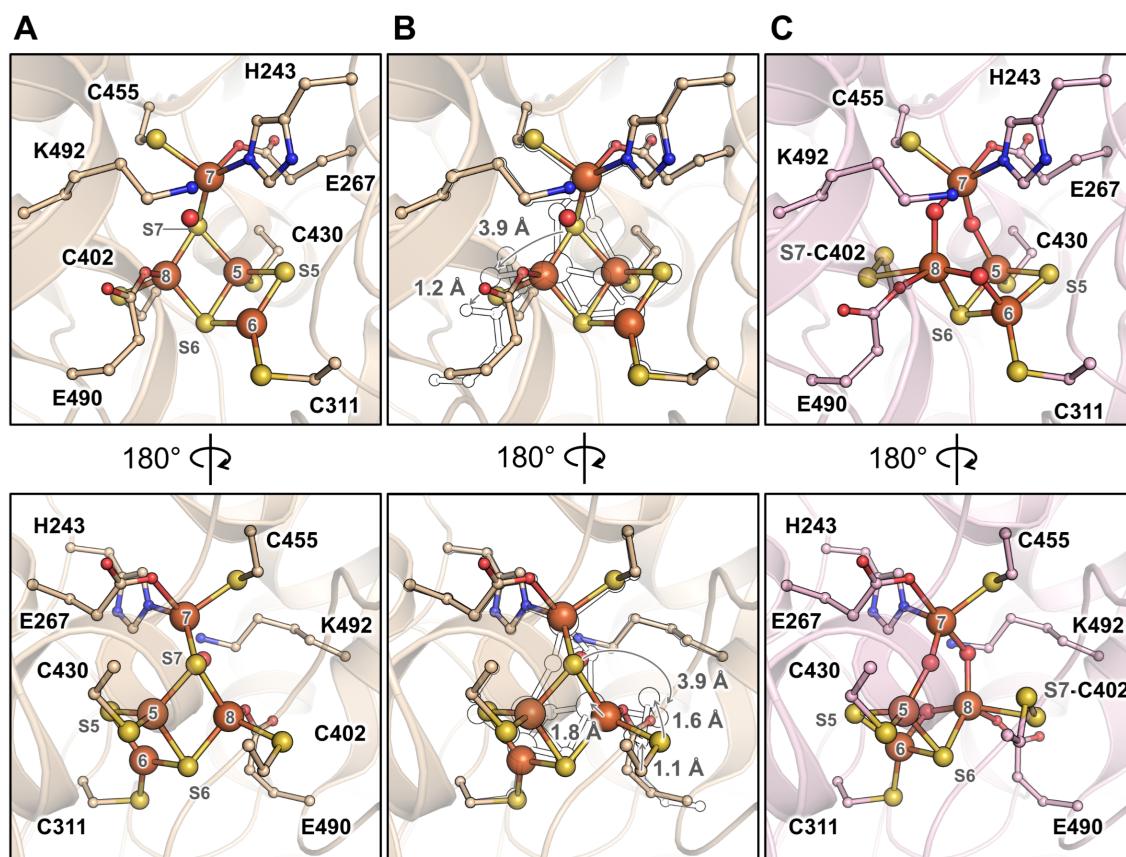


FIGURE 5

Transition of the hybrid cluster from reduced to oxidised/semi-oxidised state. (A) The reduced state structure *MthCP_{red}*. (B) Transition from the reduced to oxidised/semi-oxidised states. The structure of the reduced state (as shown in panel A) is overlaid with the oxidised/semi-oxidised state outlined in black and white. The proposed movement of the atoms is symbolised by arrows, except for movements below 1Å. (C) The oxidised/semi-oxidised state of the hybrid cluster (*MthCP_{ox}* extracted from *MthCP_{mix}*). The structures are represented in wheat (reduced state) and light pink (oxidised/semi-oxidised) transparent cartoon, with the cofactors and residues coordinating/interacting with the cluster being shown as balls and sticks. Oxygen, nitrogen, sulphur, and iron atoms are coloured red, blue, yellow, and orange, respectively. The atom numbering originates from (Hagen, 2022).

singular chemistry of the cofactor has been intensively dissected in a few model *Pseudomonadota* species, but the enzyme has rarely been studied in archaea. The HCPs from *Methanococcales*, including the thermophilic methanogen *M. thermolithotrophicus*, belong to class I and branch closer to the HCPs extensively studied from *Desulfovibrio* species. It explains the singular sequence and structure conservation between *DdHCP*, *DvHCP*, and *MthHCP*. Such a phylogenetic organisation led van Lis and colleagues to propose that this group of bacterial enzymes have an archaeal origin, acquired by the bacteria by inter-domain horizontal gene transfer (van Lis et al., 2020). The trees presented in Figure 7 and Supplementary Figure S11 are coherent with this proposal, as they show the branch of bacterial class I HCP emerging from the tree of archaeal HCPs. The HCPs from bacteria and *Methanococcales* species are harboured by the same branch in which they form monophyletic groups. This branch emerges from a tree mainly composed of sequences from *Methanomicrobia* and *Halobacteria*, yet belonging to a different archaeal clade compared to *Methanococcales*. Hence, the trees suggest horizontal gene transfer between different archaeal phyla but also from archaea to bacteria. As HCPs from *Methanococcales* species and bacteria cluster separately, one can hypothesise that these organisms acquired the gene from a similar

archaeal ancestor or that the transfer of the protein from one group to the other occurred during their early evolution. Anyway, the proximity between archaeal and bacterial enzymes favours the hypothesis of a similar role within the organisms, coherently with the overall conservation of the active site architecture.

4.2. Physiological roles of HCP in methanogens

The different studies tend to agree on a physiological NO reductase role of HCP in bacteria, detoxifying the highly reactive gas in the provenance of side reactions of nitrate and nitrite reduction or from hosts in the case of pathogenic organisms. The NO reduction might be coupled with a regulatory protein trans-nitrosylation activity (Hagen, 2022). The similar kinetic parameters for the hydroxylamine reductase and the identical active site architecture support that *MthHCP*, *DdHCP*, and *DvHCP* share catalytic properties. It also implies that *MthHCP* should, in principle, perform NO reduction. However, most methanogens, especially *Methanococcales*, do not infect eukaryotic hosts, and

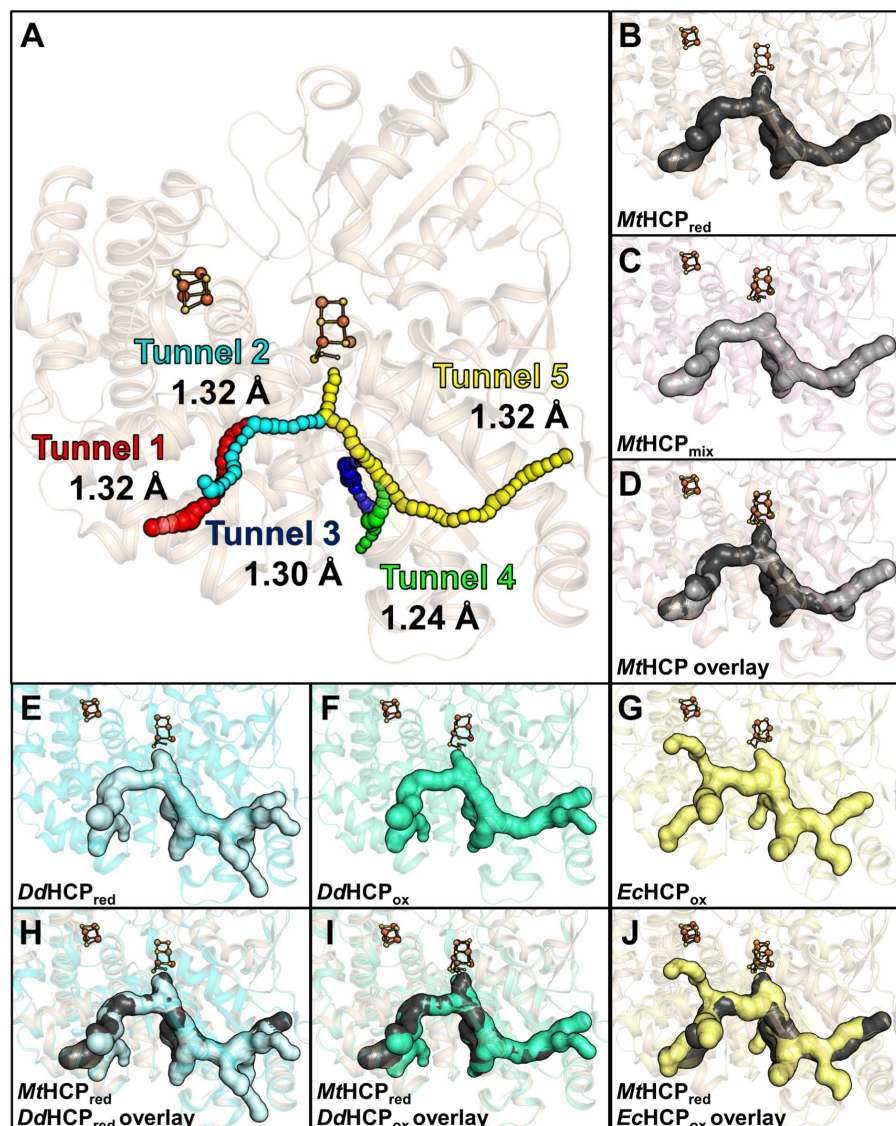


FIGURE 6

Channelling system in HCPs. (A) Tunnel analysis of *MtHCP_{red}* by the CAVER programme. Spheres in different colours represent the tunnels predicted by the analysis, and the bottleneck radius is given in Ångström. (B–D) Prediction of the channelling system in *MtHCP* structures. (E–G) Prediction of the channelling system in *DdHCP_{red}* (PDB 1OA0), *DdHCP_{ox}* (PDB 1GNL), and *EcHCP_{ox}* (PDB 7DE4) structures, respectively. (H–J) Superimposition of *MtHCP_{red}* and bacterial structures and their channelling systems. (A–J) The structures are represented as transparent cartoons. (B–J) The predicted channelling systems are in surfaces with the clusters, and the coordinating cysteine 402 (*MtHCP* numbering) are presented in balls and sticks, with sulphur and iron coloured yellow and orange, respectively. The structures of *MtHCP_{red}*, *MtHCP_{mix}*, *DdHCP_{red}*, *DdHCP_{ox}*, and *EcHCP_{ox}* are coloured wheat, light pink, light cyan, green cyan, and light yellow, respectively.

only a few reduce nitrate (Belay et al., 1990; Jeanthon et al., 1998). In our culture conditions, *M. thermolithotrophicus* produced a sufficient quantity of HCP for native purification and crystallisation, yet the culture depends exclusively on ammonium as a nitrogen source. Accordingly, our recent transcriptomic analysis performed on *M. thermolithotrophicus* showed that the *hcp* gene expression is not impacted by the nitrogen source, unlike nearly all genes involved in nitrogen acquisition (Maslać et al., 2022). Therefore, it is rather less probable that *MtHCP* is used for nitrate assimilation.

If *MtHCP* physiologically serves as detoxifying NO reductase, one can thus wonder about the NO origin. Methanogens genomes

do not harbour genes encoding for the NO-production pathway, such as NO synthase or NO-generating nitrite reductase. An alternative enzyme or pathway would be necessary. Nitrite reduction to ammonia can be performed by methanogens [e.g., by the F₄₂₀-dependent sulfite/nitrite reductase (Jespersen et al., 2023)]. The enzymatic reduction of nitrite may, in theory, be a source of NO, similar to the nitrate reducers. Under these growing conditions, *M. thermolithotrophicus* does not have any exogenous nitrite source and therefore NO-production as a side product of nitrite reduction would depend on an endogenous nitrite source (e.g., by the anabolism), incoherent with the abundance of the HCP in the cytoplasm. An abiotic chemical reaction might occur

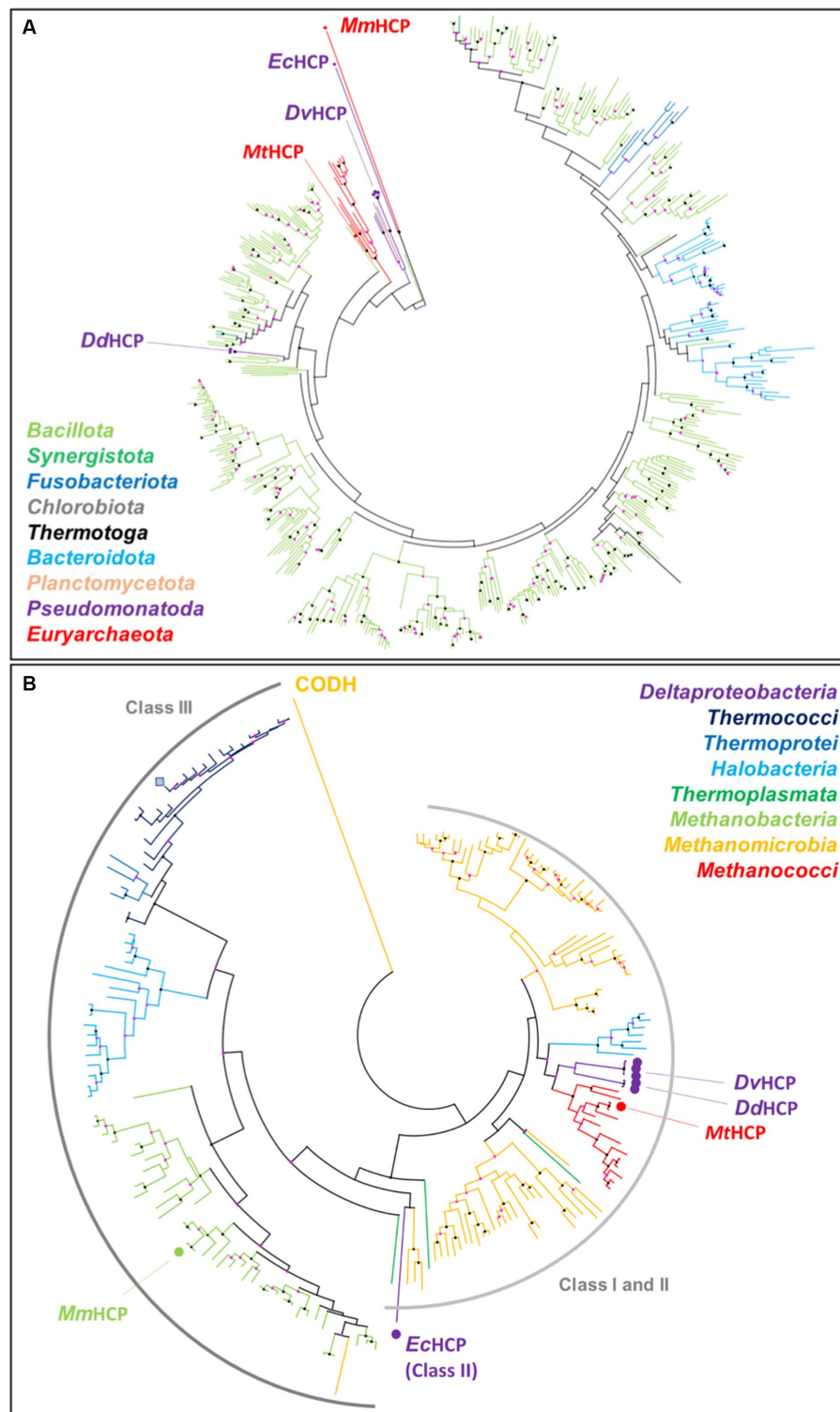


FIGURE 7

MtHCP position in the phylogenetic tree of HCPs. (A) HCP phylogenetic tree from the PDB (highlighted by a dot and labelled) and the 500 closest sequences to *MtHCP* from the RefSeq database. The branches are coloured by phylum. (B) HCP phylogenetic tree from the PDB (highlighted by a dot and labelled) and the 500 closest sequences to *MtHCP* from the RefSeq database, restricted to the *Archaea* domain. The branches are coloured by class. The sequence of *P. furiosus* is marked as a blue square. The labelled *CODH* is used as an outgroup. Node statistics (200 replicates) between 50–90 and 90–100 are represented by pink and black dots, respectively.

in the sulfidic and metal-rich medium at 60°C as a source of nitrite or NO. However, numerous methanogens, including mesophilic species, contain a gene coding for HCP (class I or III).

The protein was also shown to be relatively abundant in mesophilic *Bacillota* growing with ammonium as a nitrogen source (Valgepea et al., 2018). In the case of methanogens and

other anaerobes, the natural abundance of the protein is, therefore, inadequate with the apparent absence of a notable biological or abiotic NO source. Either a source of NO escapes our understanding, or the HCP has an anabolic function more important and possibly more ancient than initially thought. Nitrate/nitrite-reducing organisms and pathogens would then have acquired or derived HCP to survive the NO concentrations encountered in situ. The HCPs from *Desulfovibrio* species accordingly branch separately in the bacterial HCP phylogenetic tree, which could indicate more recent inter-bacteria horizontal gene transfers (Supplementary Figure S11). Protein nitrosylation would be a plausible function for HCP in other anaerobes, yet it has not been shown to occur in archaea to our knowledge. Future studies will answer if protein nitrosylation has an important role in archaea and if the HCP is still involved in this process. Genetically tractable methanogens such as *Methanococcus maripaludis* (Susanti et al., 2019) or *Methanosarcina acetivorans* (Nayak and Metcalf, 2017) would be an appropriate tool for assessing this question.

4.3. Electron donors of the HCP

The electron donor for the reaction is unknown in class I HCPs. In contrast, the HCP reductase is the physiological partner of class II, providing the electrons for the reaction via NADH oxidation (Wang et al., 2016). Fujishiro and colleagues exemplified a possible role of the N-terminal insertion that would serve as a docking site to bind the HCP reductase in EcHCP class II (Figure 2; Fujishiro et al., 2021). MtHCP and its structural homologues do not present this extension or anything similar, and no gene coding for an HCP reductase exists in the organism. As is often the case for classes I and III HCPs, MtHCP is encoded by an isolated gene with no related genes in the nearby genomic environment, which gives no indication of a putative partner. The systematic absence of other genes in synteny with *hcp* can be seen as an argument towards the involvement of a general electron donor. The reduced to semi oxidised state transition of the hybrid cluster exhibits a redox potential ranging from -250 to -150 mV in HCP class I, significantly lower than in class II (van Lis et al., 2020; Hagen, 2022). Assuming a similar redox potential in MtHCP and approximately 40 mV lower potential for the $[4\text{Fe-4S}]$ electron transferring cluster (van Lis et al., 2020), the enzyme reduction would be exergonic with NAD(P)H [$E_0' = -320$ mV (Walsh, 1986)] but also various substrates and electron shuttles of *M. thermolithotrophicus* such as ferredoxin [physiological potential approximately -500 mV (Schuchmann and Müller, 2014)] or F_{420} [$E_0' = -360$ mV (Walsh, 1986)]. If the HCP reduction by ferredoxin would be largely exergonic, it would also represent a substantial energy loss, improbable in such an energy-limited organism, whilst coupling the reaction with NAD(P)H and $F_{420}\text{H}_2$ would require an oxidase partner.

The role of the hydrated polar cavity connecting the cubane $[4\text{Fe-4S}]$ cluster to the surface is unknown. Using soluble molecules that would deliver electrons directly to the cluster would require a catalytic mechanism that has not been described yet. Since such cubane clusters are considered to serve as electron relays rather than catalytic centres, this hypothesis is unlikely. Furthermore, a similar

cavity is present in EcHCP that is reduced by another protein (Figure 2). It could also affect the structural stabilisation or regulation of enzyme activity.

5. Conclusion

The last years of research highlighted the NO reductase activity of HCP, used for NO detoxification and protein nitrosylation. If this is in perfect agreement with the enzyme production during nitrate and nitrite reduction or a defence mechanism during pathogenicity, its role is elusive in bacteria and archaea growing without nitrate, nitrite, or any specific source of NO. That is the case of the methanogen *M. thermolithotrophicus*, abundantly producing its HCP. The structural characterisation of this enzyme highlighted a striking similarity with bacterial enzymes, and for a good reason: HCPs from *Methanococcales* are closest to bacterial HCPs than to any others. As the growth conditions contain no apparent source of NO, it is plausible that the HCP is not used for NO detoxification but rather for an important and overlooked function in the anabolism of methanogens and probably in acetogens or other *Bacillota* species. Future genetic studies on model organisms will have to pursue this quest to extend the physiological role of HCPs in anaerobes.

Data availability statement

The data presented in the study are deposited in the PDB repository, accession number 8CNR and 8CNS.

Author contributions

OL and TW designed the research, collected the X-ray data, solved the structures, built the models, and analysed the structures. OL purified and crystallised the proteins. MB performed the activity measurements, electrophoreses, and size-exclusion chromatography experiments. OL and MB performed the bioinformatics analyses. All authors participated in the manuscript writing.

Funding

This study was funded by the Max-Planck-Gesellschaft and the Deutsche Forschungsgemeinschaft priority programme 1927, “Iron-Sulfur for Life” WA 4053/1-1.

Acknowledgments

We thank the Max Planck Institute for Marine Microbiology and the Max Planck Society for their continuous support. We also thank Christina Probian and Ramona Appel for their continuous support in the Microbial Metabolism laboratory and Marion Jespersen for the cultivation of the biomass used in this study and for carrying out the first purification step. We thank Dennis Jakob for his contribution to the project. We thank the Swiss Light Source and DESY Synchrotrons and especially the staff of beamline PXIII from SLS and P11 at PETRA III.

Conflict of interest

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.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated

organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2023.1179204/full#supplementary-material>

References

- Adam, P. S., Borrel, G., Brochier-Armanet, C., and Gribaldo, S. (2017). The growing tree of Archaea: new perspectives on their diversity, evolution and ecology. *ISME J.* 11, 2407–2425. doi: 10.1038/ismej.2017.122
- Almeida, C. C., Romão, C. V., Lindley, P. F., Teixeira, M., and Saraiva, L. M. (2006). The role of the hybrid cluster protein in oxidative stress defense. *J. Biol. Chem.* 281, 32445–32450. doi: 10.1074/jbc.M605888200
- Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., et al. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389–3402. doi: 10.1093/nar/25.17.3389
- Aragão, D., Macedo, S., Mitchell, E. P., Romão, C. V., Liu, M. Y., Frazão, C., et al. (2003). Reduced hybrid cluster proteins (HCP) from *Desulfovibrio desulfuricans* ATCC 27774 and *Desulfovibrio vulgaris* (Hildenborough): X-ray structures at high resolution using synchrotron radiation. *J. Biol. Inorganic Chem.* 8, 540–548. doi: 10.1007/s00775-003-0443-x
- Aragão, D., Mitchell, E. P., Frazão, C. F., Carrondo, M. A., and Lindley, P. F. (2008). Structural and functional relationships in the hybrid cluster protein family: structure of the anaerobically purified hybrid cluster protein from *Desulfovibrio vulgaris* at 1.35 Å resolution. *Acta Crystallogr. Sect. D* 64, 665–674. doi: 10.1107/S0907444908009165
- Belay, N., Jung, K.-Y., Rajagopal, B. S., Kremer, J. D., and Daniels, L. (1990). Nitrate as a sole nitrogen source for *Methanococcus thermolithotrophicus* and its effect on growth of several methanogenic bacteria. *Curr. Microbiol.* 21, 193–198. doi: 10.1007/BF02092121
- Beliaev, A. S., Klingeman, D. M., F., Castillo, F., Roldán, M. D., Romine, M. F., Tiedje, J. M., et al. (2005). Global transcriptome analysis of *Shewanella oneidensis* MR-1 exposed to different terminal electron acceptors. *J. Bacteriol.* 187, 7138–7145. doi: 10.1128/JB.187.20.7138-7145.2005
- Biestler, A., Dementin, S., and Drennan, C. L. (2022). Visualizing the gas channel of a monofunctional carbon monoxide dehydrogenase. *J. Inorg. Biochem.* 230:111774. doi: 10.1016/j.jinorgbio.2022.111774
- Briolat, V., and Reyssat, G. (2002). Identification of the *Clostridium perfringens* genes involved in the adaptive response to oxidative stress. *J. Bacteriol.* 184, 2333–2343. doi: 10.1128/JB.184.9.2333-2343.2002
- Bulut, S., Audebert, S., Pieulle, L., Seduk, F., Baudelet, E., Espinosa, L., et al. (2019). Clustering as a means to control nitrate respiration efficiency and toxicity in *Escherichia coli*. *mBio* 10:e01832-19. doi: 10.1128/mBio.01832-19
- Cabello, P., Pino, C., Olmo-Mira, M. F., Castillo, F., Roldán, M. D., and Moreno-Vivián, C. (2004). Hydroxylamine assimilation by *Rhodobacter capsulatus* EIF1. Requirement of the *hcp* gene (hybrid cluster protein) located in the nitrate assimilation *nas* gene region for hydroxylamine reduction. *J. Biol. Chem.* 279, 45485–45494. doi: 10.1074/jbc.M404417200
- Cadby, I. T., Faulkner, M., Cheneby, J., Long, J., van Helden, J., Dolla, A., et al. (2017). Coordinated response of the *Desulfovibrio desulfuricans* 27774 transcriptome to nitrate, nitrite and nitric oxide. *Sci. Rep.* 7:16228. doi: 10.1038/s41598-017-16403-4
- Chen, V. B., Arendall, W. B. 3rd, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., et al. (2010). MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* 66, 12–21. doi: 10.1107/S0907444909042073
- Cole, J. A. (2021). Anaerobic bacterial response to nitric oxide stress: widespread misconceptions and physiologically relevant responses. *Mol. Microbiol.* 116, 29–40. doi: 10.1111/mmi.14713
- Cooper, S. J., Garner, C. D., Hagen, W. R., Lindley, P. F., and Bailey, S. (2000). Hybrid-cluster protein (HCP) from *Desulfovibrio vulgaris* (Hildenborough) at 1.6 Å resolution. *Biochemistry* 39, 15044–15054. doi: 10.1021/bi001483m
- Edgar, R. C. (2004). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinform.* 5:113. doi: 10.1186/1471-2105-5-113
- Emmsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010). Features and development of *Coot*. *Acta Crystallogr. D Biol. Crystallogr.* 66, 486–501. doi: 10.1107/S0907444910007493
- Faeder, E. J., Davis, P. S., and Siegel, L. M. (1974). Reduced nicotinamide adenine dinucleotide phosphate-sulfite reductase of enterobacteria: V. STUDIES WITH THE
- ESCHERICHIA COLI HEMOFLAVOPROTEIN DEPLETED OF FLAVIN MONONUCLEOTIDE: DISTINCT ROLES FOR THE FLAVIN ADENINE DINUCLEOTIDE AND FLAVIN MONONUCLEOTIDE PROSTHETIC GROUPS IN CATALYSIS. *J. Biol. Chem.* 249, 1599–1609. doi: 10.1016/S0021-9258(19)42923-2
- Filenko, N. A., Browning, D. F., and Cole, J. A. (2005). Transcriptional regulation of a hybrid cluster (prismene) protein. *Biochem. Soc. Trans.* 33, 195–197. doi: 10.1042/BST0330195
- Filenko, N., Spiro, S., Browning, D. F., Squire, D., Overton, T. W., Cole, J., et al. (2007). The NsrR regulon of *Escherichia coli* K-12 includes genes encoding the hybrid cluster protein and the periplasmic, respiratory nitrite reductase. *J. Bacteriol.* 189, 4410–4417. doi: 10.1128/JB.00080-07
- Fujishiro, T., Ooi, M., and Takaoka, K. (2021). Crystal structure of *Escherichia coli* class II hybrid cluster protein, HCP, reveals a [4Fe-4S] cluster at the N-terminal protrusion. *FEBS J.* 288, 6752–6768. doi: 10.1111/febs.16062
- Fujishiro, T., and Takaoka, K. (2022). Class III hybrid cluster protein (HCP) from *Methanothermobacter marburgensis*. *Jpn. Soc. Prom. Sci.* doi: 10.2210/pdb7E0L/pdb
- Gao, S.-H., Ho, J. Y., Fan, L., Nouwens, A., Hoelzle, R. D., Schulz, B., et al. (2019). A comparative proteomic analysis of *Desulfovibrio vulgaris* Hildenborough in response to the antimicrobial agent free nitrous acid. *Sci. Total Environ.* 672, 625–633. doi: 10.1016/j.scitotenv.2019.03.442
- Hagen, W. R. (2019). EPR spectroscopy of putative enzyme intermediates in the NO reductase and the auto-nitrosylation reaction of *Desulfovibrio vulgaris* hybrid cluster protein. *FEBS Lett.* 593, 3075–3083. doi: 10.1002/1873-3468.13539
- Hagen, W. R. (2022). Structure and function of the hybrid cluster protein. *Coord. Chem. Rev.* 457:214405. doi: 10.1016/j.ccr.2021.214405
- Hagen, W. R., Pierik, A. J., and Veeger, C. (1989). Novel electron paramagnetic resonance signals from an Fe/S protein containing six iron atoms. *J. Chem. Soc. Faraday Trans. 1* 85, 4083–4090.
- Hagen, W. R., van den Berg, W. A., van Dongen, W. M., Reijerse, E. J., and van Kan, P. J. (1998). EPR spectroscopy of biological iron–sulfur clusters with spin-admixed $S = 3/2$. *J. Chem. Soc. Faraday Trans.* 94, 2969–2973. doi: 10.1039/a803059f
- Haveman, S. A., Greene, E. A., Stilwell, C. P., Voordouw, J. K., and Voordouw, G. (2004). Physiological and gene expression analysis of inhibition of *Desulfovibrio vulgaris* Hildenborough by nitrite. *J. Bacteriol.* 186, 7944–7950. doi: 10.1128/JB.186.23.7944-7950.2004
- He, Q., Huang, K. H., He, Z., Alm, E. J., Fields, M. W., Hazen, T. C., et al. (2006). Energetic consequences of nitrite stress in *Desulfovibrio vulgaris* Hildenborough, inferred from global transcriptional analysis. *Appl. Environ. Microbiol.* 72, 4370–4381. doi: 10.1128/AEM.02609-05
- Heo, J., Wolfe, M. T., Staples, C. R., and Ludden, P. W. (2002). Converting the NiFeS carbon monoxide dehydrogenase to a hydrogenase and a hydroxylamine reductase. *J. Bacteriol.* 184, 5894–5897. doi: 10.1128/JB.184.21.5894-5897.2002
- Jarrell, K. F., and Sprott, G. D. (1981). The transmembrane electrical potential and intracellular pH in methanogenic bacteria. *Can. J. Microbiol.* 27, 720–728. doi: 10.1139/m81-110
- Jeanthon, C., L'Haridon, S., Reysenbach, A. L., Vernet, M., Messner, P., Sleytr, U. B., et al. (1998). *Methanococcus infernus* sp. nov., a novel hyperthermophilic lithotrophic methanogen isolated from a deep-sea hydrothermal vent. *Int. J. Syst. Evol. Microbiol.* 48, 913–919.
- Jeoung, J.-H., Fessler, J., Domnik, L., Klemke, F., Sinnreich, M., Teutloff, C., et al. (2022). A morphing [4Fe-3S-nO]-cluster within a carbon monoxide dehydrogenase scaffold. *Angew. Chem. Int. Ed.* 61:e202117000. doi: 10.1002/anie.202117000
- Jespersen, M., Pierik, A. J., and Wagner, T. (2023). Structures of the sulfite detoxifying F_{420} -dependent enzyme from *Methanococcales*. *Nat. Chem. Biol.* doi: 10.1038/s41589-022-01232-y
- Kim, C. C., Monack, D., and Falkow, S. (2003). Modulation of virulence by two acidified nitrite-responsive loci of *Salmonella enterica* Serovar Typhimurium. *Infect. Immun.* 71, 3196–3205. doi: 10.1128/IAI.71.6.3196-3205.2003

- Klüber, H. D., and Conrad, R. (1998). Inhibitory effects of nitrate, nitrite, NO and N₂O on methanogenesis by *Methanosarcina barkeri* and *Methanobacterium bryantii*. *FEMS Microbiol. Ecol.* 25, 331–339. doi: 10.1016/S0168-6496(97)00102-5
- Kozlikova, B., Sebestova, E., Sustr, V., Brezovsky, J., Strnad, O., Daniel, L., et al. (2014). CAVER analyst 1.0: graphic tool for interactive visualization and analysis of tunnels and channels in protein structures. *Bioinformatics* 30, 2684–2685. doi: 10.1093/bioinformatics/btu364
- Krissinel, E., and Henrick, K. (2004). Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr. D Biol. Crystallogr.* 60, 2256–2268. doi: 10.1107/S0907444904026460
- Kröckel, M., Trautwein, A. X., Arendsen, A. F., and Hagen, W. R. (1998). The prismatic protein resolved—Mössbauer investigation of a 4Fe cluster with an unusual mixture of bridging ligands and metal coordinations. *Eur. J. Biochem.* 251, 454–461. doi: 10.1046/j.1432-1327.1998.2510454.x
- 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
- Lemaire, O. N., Infossi, P., Ali Chaouche, A., Espinosa, L., Leimkühler, S., Giudici-Orticoni, M. T., et al. (2018). Small membranous proteins of the TorE/NapE family, crutches for cognate respiratory systems in *Proteobacteria*. *Sci. Rep.* 8:13576. doi: 10.1038/s41598-018-31851-2
- Lemaire, O. N., and Wagner, T. (2021). Gas channel rerouting in a primordial enzyme: structural insights of the carbon-monoxide dehydrogenase/acetyl-CoA synthase complex from the acetogen *Clostridium autoethanogenum*. *Biochim. Biophys. Acta Bioenerg.* 1862:148330. doi: 10.1016/j.bbabi.2020.148330
- Liebschner, D., Afonine, P. V., Baker, M. L., Bunkóczi, G., Chen, V. B., Croll, T. I., et al. (2019). Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr. D Struct. Biol.* 75, 861–877. doi: 10.1107/S2059798319011471
- Macedo, S., Araújo, D., Mitchell, E. P., and Lindley, P. (2003). Structure of the hybrid cluster protein (HCP) from *Desulfovibrio desulfuricans* ATCC 27774 containing molecules in the oxidized and reduced states. *Acta Crystallogr. D Biol. Crystallogr.* 59, 2065–2071. doi: 10.1107/S0907444903025861
- Macedo, S., Mitchell, E. P., Romão, C. V., Cooper, S. J., Coelho, R., Liu, M. Y., et al. (2002). Hybrid cluster proteins (HCPs) from *Desulfovibrio desulfuricans* ATCC 27774 and *Desulfovibrio vulgaris* (Hildenborough): X-ray structures at 1.25 Å resolution using synchrotron radiation. *J. Biol. Inorg. Chem.* 7, 514–525. doi: 10.1007/s00775-001-0326-y
- Mager, J. (1960). A TPNH-linked sulfite reductase and its relation to hydroxylamine reductase in *Enterobacteriaceae*. *Biochim. Biophys. Acta* 41, 553–555. doi: 10.1016/0006-3002(60)90065-2
- Marriott, S. J., Farrar, A., Breton, J. L., Hagen, W. R., and Thomson, A. J. (1995). Characterization of the prismatic protein from *Desulfovibrio vulgaris* (Hildenborough) by low-temperature magnetic circular dichroic spectroscopy. *Eur. J. Biochem.* 232, 501–505. doi: 10.1111/j.1432-1033.1995.501zz.x
- Maslač, N., Sidhu, C., Teeling, H., and Wagner, T. (2022). Comparative transcriptomics sheds light on remodeling of gene expression during diazotrophy in the thermophilic methanogen *Methanothermococcus thermolithotrophicus*. *mBio* 13:e0244322. doi: 10.1128/mbio.02443-22
- Moura, I., Tavares, P., Moura, J. J., Ravi, N., Huynh, B. H., Liu, M. Y., et al. (1992). Direct spectroscopic evidence for the presence of a 6Fe cluster in an iron-sulfur protein isolated from *Desulfovibrio desulfuricans* (ATCC 27774). *J. Biol. Chem.* 267, 4489–4496. doi: 10.1016/S0021-9258(18)42859-1
- Nayak, D. D., and Metcalf, W. W. (2017). Cas9-mediated genome editing in the methanogenic archaeon *Methanosarcina acetivorans*. *Proc. Natl. Acad. Sci. U. S. A.* 114, 2976–2981. doi: 10.1073/pnas.1618596114
- Okinaka, Y., Yang, C. H., Perna, N. T., and Keen, N. T. (2002). Microarray profiling of *Erwinia chrysanthemi* 3937 genes that are regulated during plant infection. *Mol. Plant Microbe Interact.* 15, 619–629. doi: 10.1094/MPMI.2002.15.7.619
- Ostrowski, J., Barber, M. J., Rueger, D. C., Miller, B. E., Siegel, L. M., and Kredich, N. M. (1989). Characterization of the flavoprotein moieties of NADPH-sulfite reductase from *Salmonella typhimurium* and *Escherichia coli*. Physicochemical and catalytic properties, amino acid sequence deduced from DNA sequence of *cysJ*, and comparison with NADPH-cytochrome P-450 reductase. *J. Biol. Chem.* 264, 15796–15808. doi: 10.1016/S0021-9258(18)1547-0
- Overeijnder, M. L., Hagen, W. R., and Hagedoorn, P.-L. (2009). A thermostable hybrid cluster protein from *Pyrococcus furiosus*: effects of the loss of a three helix bundle subdomain. *J. Biol. Inorganic Chem.* 14, 703–710. doi: 10.1007/s00775-009-0483-y
- Pereira, A. S., Tavares, P., Krebs, C., Huynh, B. H., Rusnak, F., Moura, I., et al. (1999). Biochemical and spectroscopic characterization of overexpressed fuscoredoxin from *Escherichia coli*. *Biochem. Biophys. Res. Commun.* 260, 209–215. doi: 10.1006/bbrc.1999.0748
- Pierik, A. J., Hagen, W. R., Dunham, W. R., and Sands, R. H. (1992). Multi-frequency EPR and high-resolution Mössbauer spectroscopy of a putative [6Fe-6S] prismatic-cluster-containing protein from *Desulfovibrio vulgaris* (Hildenborough). Characterization of a supercluster and superspin model protein. *Eur. J. Biochem.* 206, 705–719. doi: 10.1111/j.1432-1033.1992.tb16977.x
- Pierik, A. J., Wolbert, R. B., Mutsaers, P. H., Hagen, W. R., and Veeger, C. (1992). Purification and biochemical characterization of a putative [6Fe-6S] prismatic-cluster-containing protein from *Desulfovibrio vulgaris* (Hildenborough). *Eur. J. Biochem.* 206, 697–704. doi: 10.1111/j.1432-1033.1992.tb16976.x
- Robert, X., and Gouet, P. (2014). Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* 42, W320–W324. doi: 10.1093/nar/gku316
- Roos, V., and Klemm, P. (2006). Global gene expression profiling of the asymptomatic bacteriuria *Escherichia coli* strain 83972 in the human urinary tract. *Infect. Immun.* 74, 3565–3575. doi: 10.1128/IAI.01959-05
- Schuchmann, K., and Müller, V. (2014). Autotrophy at the thermodynamic limit of life: a model for energy conservation in acetogenic bacteria. *Nat. Rev. Microbiol.* 12, 809–821. doi: 10.1038/nrmicro3365
- Seth, D., Hess, D. T., Hausladen, A., Wang, L., Wang, Y.-J., and Stamler, J. S. (2018). A multiplex enzymatic machinery for cellular protein S-nitrosylation. *Mol. Cell* 69, 451–464.e6. doi: 10.1016/j.molcel.2017.12.025
- 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
- Spiro, S. (2007). Regulators of bacterial responses to nitric oxide. *FEMS Microbiol. Rev.* 31, 193–211. doi: 10.1111/j.1574-6976.2006.00061.x
- Stokkermans, J. P. W. G., Houba, P. H. J., Pierik, A. J., Hagen, W. R., van Dongen, W. M. A. M., and Veeger, C. (1992). Overproduction of prismatic protein in *Desulfovibrio vulgaris* (Hildenborough): evidence for a second S= 1/2-spin system in the one-electron reduced state. *Eur. J. Biochem.* 210, 983–988. doi: 10.1111/j.1432-1033.1992.tb17503.x
- Susanti, D., Frazier, M. C., and Mukhopadhyay, B. (2019). A genetic system for *Methanocaldococcus jannaschii*: an evolutionary deeply rooted hyperthermophilic methanarchaeon. *Front. Microbiol.* 12:5610. doi: 10.3389/fmicb.2019.01256
- Tavares, P., Pereira, A. S., Krebs, C., Ravi, N., Moura, J. J., Moura, I., et al. (1998). Spectroscopic characterization of a novel tetranuclear Fe cluster in an iron-sulfur protein isolated from *Desulfovibrio desulfuricans*. *Biochemistry* 37, 2830–2842. doi: 10.1021/bi9723008
- Tickle, IJ, Flensburg, C., Keller, P., Paciorek, W., Sharff, A., Vornrhein, C., et al. STARANISO (Cambridge, United Kingdom: Global Phasing), (2018).
- Usón, I., and Sheldrick, G. (2018). An introduction to experimental phasing of macromolecules illustrated by SHELX; new autotracing features. *Acta Crystallogr. Sec. D Struct. Biol.* 74, 106–116. doi: 10.1107/S2059798317015121
- Valgepea, K., De Souza Pinto Lemgruber, R., Abdalla, T., Binos, S., Takemori, N., Takemori, A., et al. (2018). H₂ drives metabolic rearrangements in gas-fermenting *Clostridium autoethanogenum*. *Biotechnol. Biofuels* 11:55. doi: 10.1186/s13068-018-1052-9
- van den Berg, W. A., Hagen, W. R., and van Dongen, W. M. (2000). The hybrid-cluster protein ('prismatic protein') from *Escherichia coli*. Characterization of the hybrid-cluster protein, redox properties of the [2Fe-2S] and [4Fe-2S-2O] clusters and identification of an associated NADH oxidoreductase containing FAD and [2Fe-2S]. *Eur. J. Biochem.* 267, 666–676. doi: 10.1046/j.1432-1327.2000.01032.x
- van den Berg, W. A., Stevens, A. A., Verhagen, M. F., van Dongen, W. M., and Hagen, W. R. (1994). Overproduction of the prismatic protein from *Desulfovibrio desulfuricans* ATCC 27774 in *Desulfovibrio vulgaris* (Hildenborough) and EPR spectroscopy of the [6Fe-6S] cluster in different redox states. *Biochim. Biophys. Acta Protein Struct. Mol. Enzymol.* 1206, 240–246. doi: 10.1016/0167-4838(94)90214-3
- van Lis, R., Brugière, S., Baffert, C., Couté, Y., Nitschke, W., and Atteia, A. (2020). Hybrid cluster proteins in a photosynthetic microalga. *FEBS J.* 287, 721–735. doi: 10.1111/febs.15025
- Vornrhein, C., Flensburg, C., Keller, P., Sharff, A., Smart, O., Paciorek, W., et al. (2011). Data processing and analysis with the autoPROC toolbox. *Acta Crystallogr. Sect. D* 67, 293–302. doi: 10.1107/S0907444911007773
- Walsh, C. (1986). Naturally occurring 5-deazaflavin coenzymes: biological redox roles. *Acc. Chem. Res.* 19, 216–221. doi: 10.1021/ar00127a004
- Wang, Y., Cai, X., Fan, J., Wang, D., and Mao, Y. (2022). Transcriptome analysis provides new insights into the tolerance and aerobic reduction of *Shewanella decolorationis* N11-3 to bromate. *Appl. Microbiol. Biotechnol.* 106, 4749–4761. doi: 10.1007/s00253-022-12006-w
- Wang, J., Vine, C. E., Balasany, B. K., Rizk, J., Bradley, C. L., Tinajero-Trejo, M., et al. (2016). The roles of the hybrid cluster protein, Hcp and its reductase, Hcr, in high affinity nitric oxide reduction that protects anaerobic cultures of *Escherichia coli* against nitrosative stress. *Mol. Microbiol.* 100, 877–892. doi: 10.1111/mmi.13356
- Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., et al. (2011). Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.* 67, 235–242. doi: 10.1107/S0907444910045749

Wolfe, M. T., Heo, J., Garavelli, J. S., and Ludden, P. W. (2002). Hydroxylamine reductase activity of the hybrid cluster protein from *Escherichia coli*. *J. Bacteriol.* 184, 5898–5902. doi: 10.1128/JB.184.21.5898-5902.2002

Yu, Z., Pesesky, M., Zhang, L., Huang, J., Winkler, M., and Chistoserdova, L. (2020). A complex interplay between nitric oxide, quorum sensing, and the unique secondary metabolite tundrenone constitutes the hypoxia response in *Methylobacter*. *mSystems* 5:e00770-19. doi: 10.1128/mSystems.00770-19

Yurkiw, M. A., Voordouw, J., and Voordouw, G. (2012). Contribution of rubredoxin: oxygen oxidoreductases and hybrid cluster proteins of *Desulfovibrio vulgaris* Hildenborough to survival under oxygen and nitrite stress. *Environ. Microbiol.* 14, 2711–2725. doi: 10.1111/j.1462-2920.2012.02859.x

Zhang, W., Culley, D. E., Hogan, M., Vitiritti, L., and Brockman, F. J. (2006). Oxidative stress and heat-shock responses in *Desulfovibrio vulgaris* by genome-wide transcriptomic analysis. *Antonie Van Leeuwenhoek* 90, 41–55. doi: 10.1007/s10482-006-9059-9



OPEN ACCESS

EDITED BY

Marleen van Wolferen,
University of Freiburg, Germany

REVIEWED BY

Doug Bartlett,
University of California, San Diego,
United States
Changyi Zhang,
University of Illinois at Urbana-Champaign,
United States

*CORRESPONDENCE

Michail M. Yakimov
✉ mikhail.iakimov@cnr.it

†These authors have contributed equally to this work

RECEIVED 08 March 2023

ACCEPTED 09 May 2023

PUBLISHED 31 May 2023

CITATION

Reva O, Messina E, La Cono V, Crisafi F, Smedile F, La Spada G, Marturano L, Selivanova EA, Rohde M, Krupovic M and Yakimov MM (2023) Functional diversity of nanohaloarchaea within xylan-degrading consortia. *Front. Microbiol.* 14:1182464. doi: 10.3389/fmicb.2023.1182464

COPYRIGHT

© 2023 Reva, Messina, La Cono, Crisafi, Smedile, La Spada, Marturano, Selivanova, Rohde, Krupovic and Yakimov. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). 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.

Functional diversity of nanohaloarchaea within xylan-degrading consortia

Oleg Reva^{1†}, Enzo Messina^{2†}, Violetta La Cono³,
Francesca Crisafi³, Francesco Smedile³, Gina La Spada³,
Laura Marturano³, Elena A. Selivanova⁴, Manfred Rohde⁵,
Mart Krupovic⁶ and Michail M. Yakimov^{3*}

¹Department of Biochemistry, Genetics and Microbiology, Centre for Bioinformatics and Computational Biology, University of Pretoria, Pretoria, South Africa, ²National Council of Research, CNR, Rome, Italy, ³Extreme Microbiology, Biotechnology and Astrobiology Group, Institute of Polar Research, ISP-CNR, Messina, Italy, ⁴Institute for Cellular and Intracellular Symbiosis, Ural Branch, Russian Academy of Sciences, Orenburg, Russia, ⁵Central Facility for Microbiology, Helmholtz Centre for Infection Research, Braunschweig, Germany, ⁶Archaeal Virology Unit, Institut Pasteur, Université Paris Cité, Paris, France

Extremely halophilic representatives of the phylum *Candidatus* Nanohaloarchaeota (members of the DPANN superphyla) are obligately associated with extremely halophilic archaea of the phylum *Halobacteriota* (according to the GTDB taxonomy). Using culture-independent molecular techniques, their presence in various hypersaline ecosystems around the world has been confirmed over the past decade. However, the vast majority of nanohaloarchaea remain uncultivated, and thus their metabolic capabilities and ecophysiology are currently poorly understood. Using the (meta)genomic, transcriptomic, and DNA methylome platforms, the metabolism and functional prediction of the ecophysiology of two novel extremely halophilic symbiotic nanohaloarchaea (*Ca. Nanohalococcus occultus* and *Ca. Nanohalovita haloferacivicina*) stably cultivated in the laboratory as members of a xylose-degrading binary culture with a haloarchaeal host, *Haloferax lucentense*, was determined. Like all known DPANN superphylum nanoorganisms, these new sugar-fermenting nanohaloarchaea lack many fundamental biosynthetic repertoires, making them exclusively dependent on their respective host for survival. In addition, given the cultivability of the new nanohaloarchaea, we managed to discover many unique features in these new organisms that have never been observed in nano-sized archaea both within the phylum *Ca. Nanohaloarchaeota* and the entire superphylum DPANN. This includes the analysis of the expression of organism-specific non-coding regulatory (nc)RNAs (with an elucidation of their 2D-secondary structures) as well as profiling of DNA methylation. While some ncRNA molecules have been predicted with high confidence as RNAs of an archaeal signal recognition particle involved in delaying protein translation, others resemble the structure of ribosome-associated ncRNAs, although none belong to any known family. Moreover, the new nanohaloarchaea have very complex cellular defense mechanisms. In addition to the defense mechanism provided by the type II restriction-modification system, consisting of Dcm-like DNA methyltransferase and Mrr restriction endonuclease, *Ca. Nanohalococcus* encodes an active type I-D CRISPR/Cas system, containing 77 spacers divided into two loci. Despite

their diminutive genomes and as part of their host interaction mechanism, the genomes of new nanohaloarchaea do encode giant surface proteins, and one of them (9,409 amino acids long) is the largest protein of any sequenced nanohaloarchaea and the largest protein ever discovered in cultivated archaea.

KEYWORDS

nanohaloarchaeal-haloarchaeal symbioses, ecology of nanohaloarchaea, CRISPR, ncRNA, methylomics

Introduction

The last decade has revealed the astonishing diversity of ultrasmall archaea (<500 nm in size, with cell volumes as low as $0.009 \mu\text{m}^3$) in nature, and the number of studies emphasizing their importance in the ecology, phylogeny, and evolutionary history of archaea is constantly growing. These nano-sized organisms belong to the tentative archaeal superphylum DPANN (an acronym of the names of the first described candidate phyla Diapherotrites, Parvarchaeota, Aenigmarchaeota, Nanoarchaeota, and Nanohaloarchaeota), which currently accommodates additional six candidate phyla (Huberarchaeota, Micrarchaeota, Mamarchaeota, Pacearchaeota, Undinarchaeota, and Woesearchaeota) (Castelle et al., 2018; Dombrowski et al., 2020). Members of DPANN are ubiquitous in nature and inhabit extremely diverse environments worldwide, including acid mine drainage (Golyshina et al., 2017) and highly alkaline sites (Vavourakis et al., 2016), brackish/freshwater, ocean sediments, hydrothermal vents, permafrost, and human microbiome (Dombrowski et al., 2019; for further references). The phylum “*Candidatus* Nanohaloarchaeota” is a representative extremely halophilic lineage within the DPANN superphylum. They inhabit various hypersaline ecosystems around the world: from natural salt lakes (Ghai et al., 2011; Narasingarao et al., 2012; Selivanova et al., 2018) and crystallizer ponds of solar salterns (La Cono et al., 2020; Leoni et al., 2020) to polyextreme habitats, such as salt crusts of the hyperarid deserts (Crits-Christoph et al., 2016; Finstad et al., 2017), Antarctic permanently cold hypersaline lakes (Hamm et al., 2019), and Eastern Mediterranean magnesium-saturated deep-sea brine lakes (La Cono et al., 2019). With cell volumes approaching the theoretical size limit for the unicellular form of life ($0.008 \mu\text{m}^3$; Comolli et al., 2009) and genome sizes of approximately 1.0 Mbp, nanohaloarchaea are interesting in that they have multiple auxotrophies and so appear to depend, as ectosymbionts, on haloarchaeal hosts. However, as is true for all DPANN archaea, the most-basic questions regarding cellular metabolism and physiology, the nature of their trophic interaction with hosts, and their ecophysiological functions in the environment remain unanswered, as only a handful members of *Ca.* Nanohaloarchaeota have been grown and cultured under laboratory conditions in binary cultures. While an Antarctic haloarchaeon *Halorubrum lacusprofundi* is required for true parasitic growth of Antarctic nanohaloarchaea (proposed as *Ca.* Nanohaloarchaeum antarcticus) (Hamm et al., 2019), the nanohaloarchaeon *Ca.* Nanohalobium constans coexists with the chitinolytic haloarchaeon *Halomicrobium* sp. LC1Hm as a

mutualistic ectosymbiont, an unprecedented finding (La Cono et al., 2020).

Here, we present a comparative study of metabolic functions, a repertoire of surface proteins likely involved in interaction with hosts, and the ecophysiology of two extremely halophilic nanohaloarchaea, *Ca.* Nanohalococcus occultus and *Ca.* Nanohalovita haloferacivicina. Our recent study to stably cultivate these nanohaloarchaea as members of xylan-degrading haloarchaeal consortia, obtained from brines and sediments of two distant hypersaline ecosystems (La Cono et al., 2023), has enabled the current study of their functional diversity that is key to understanding their ecology in nature. Metabolism, functional prediction of unique features, and both interaction with hosts and defense mechanisms were inferred from their cultivation experiments in binary cultures, whole genomes, transcriptomics, and DNA methylome analyses. Despite highly reduced, streamlined genomes and the absence of many fundamental biosynthetic repertoires, making them exclusively dependent on their respective host for survival, new nanohaloarchaeal strains have a complex set of genes (which occupy more than 10% of their entire genomes), involved in elaborated cell surface construction and adaptation to the ectosymbiotic lifestyle. Overall, they exhibit remarkable physiological and adaptive diversity that parallels what was recently described for other *Ca.* Nanohaloarchaeota members (Xie et al., 2022; Zhao et al., 2022). The present study greatly expands our understanding of the functional and genetic relationships between nanohaloarchaeal symbionts and their haloarchaeal hosts.

Results and discussion

General genome and defense mechanism characteristics of novel cultivated nanohaloarchaea

The genome of *Ca.* Nanohalococcus occultus SVXNc (SVXNc for short) consists of a single circular chromosome of 948,351 bp with molGC content of 47.4%. It contains single copies of 5S, 16S, and 23S rRNA genes located in three different loci, as well as 39 tRNA genes, 15 of which have an intron. Of the 1,046 protein-coding genes annotated in strain SVXNc, 620 (59.3% of the total) could be assigned to one of the NCBI Clusters of Orthologous Groups (COG) categories and 667 (63.8% of the total) to Archaeal Clusters of Orthologous Genes (arCOG). Notably, the 28,230 bp gene is present in this genome encoding for giant surface protein SVXNc_0300, similar to those found in genomes of other

DPANN representatives, including nanohaloarchaea (see Hamm et al., 2019 for further references). The genome of *Ca. Nanohalovita haloferacivicina* BNXNv (BNXNv for short) consists of a single circular chromosome of 1,046,959 bp with molGC content of 45.0%. It contains single copies of 5S, 16S, and 23S rRNA genes located in three different loci, as well as 38 tRNA genes, four of which have an intron. Of the 1,170 protein-coding genes annotated in the strain BNXNv, 675 (57.7% of the total) could be assigned to one of the NCBI COG categories and 738 (63.1% of the total) to an arCOGs. Like *Ca. Nanohalococcus occultus* SVXNc, the large 26,643 bp gene encoding the giant surface protein was also found in the BNXNv genome and is discussed below.

Circos ribbon plots (Krzywinski et al., 2009) were used to visualize synteny between genomes of newly cultivated nanohaloarchaea and *Ca. Nanohalobium constans* LC1Nh (La Cono et al., 2020) used as a reference and to find clusters of orthologous genes. High overall collinearity and high similarity in terms of the gene context (>70% of amino acid identity) were found between all of them along with the presence of a few organism-specific genome rearrangements, such as the giant proteins and several genomic islands (Figure 1). These findings strongly suggest the common origin of all these nanohaloarchaea, even though they were isolated from geographically distinct sites and have different haloarchaeal hosts (La Cono et al., 2020, 2023).

Both genomic islands, found in the SVXNc genome, had a relatively lower GC content (41.5–44 mol%) compared to the 47.5 mol% calculated for the genome as a whole. A BLASTP search showed that these genomic islands may be acquired horizontally from other archaeal microorganisms. Noteworthy, no genomic islands were found within the *Ca. Nanohalovita* BNXNh genome, which is generally uncommon in nanohaloarchaea and other DPANN members. In the SVXNc genome, CRISPRFinder detected six potential CRISPRs regions (clusters of regularly interspaced short palindromic repeats), although a deeper analysis confirmed the presence of only one intact CRISPR/Cas system (SVXNc_0336-46) in the region from 328,855 bp to 342,986 bp, which corresponds to the location of one of the two genomic islands found. According to the CRISPR classification (Makarova et al., 2020), this system belongs to the type I-D CRISPR/Cas system and contains 77 spacers divided into two loci containing 31 and 46 spacers, respectively, and separated by the Cas2 (SVXNc_0336) gene and two genes encoding hypothetical proteins (SVXNc_0337-8) (Figure 2A). Apart from the MAG-assembled *Ca. Nanopetraeus* sp. SG9 (Crits-Christoph et al., 2016) and nanohaloarchaeon E09 (GCA_012927765) found in hypersaline salterns near Alicante (Spain) (Feng et al., 2021), we did not find evidence of other CRISPR/Cas systems in any other publicly available nanohaloarchaeal genomes. Remarkably, the 16,166 bp-long CRISPR region had a significantly lower G+C content (41.4 mol%) than the entire SVXNc genome (47.5%). To elucidate the origin of the CRISPR/Cas system in SVXNc, the *cas1* gene product was aligned with Cas1 proteins from various archaeal genomes. The resulting phylogeny showed that the SVXNc Cas1 protein, together with the E09 Cas1 protein, form a sister clade to Cas1 homologs from archaea of the classes *Methanomicrobia* and *Halobacteria* (Figure 2B). However, the genomes of *Halobacteria*, with the exception of that of *Haloquadratum walsbyi* (G+C % of 47.9) (Bolhuis et al.,

2006), are characterized by high G+C% (>60 mol%), while the genomes of some extremely halophilic methanogens, such as *Methanohalarchaeum* and *Methanohalophilus*, have low G+C% (38 mol%) as does the CRISPR locus of SVXNc (Spring et al., 2010; Sorokin et al., 2017). These results support the potential origin of the nanohaloarchaeal CRISPR locus from halophilic methanogens.

Although the role of CRISPR/Cas systems as protective agents against viruses and other potential mobile agents that can threaten the integrity of the genome is well known, the analysis of spacer sequences as records of possible past encounters is still very challenging. To date, many haloviruses have been identified that attack members of halophilic microbial communities (see Atanasova et al., 2015; Truitt and Deole, 2021 for further references). However, we did not find any similarity between the SVXNc spacers and the known halovirus sequences, suggesting that this nanohaloarchaeon has dealt with or encountered unknown viruses or other mobile genetic elements in the past. Alternatively, given that *Ca. Nanohalococcus* SVXNc, which has a very limited anabolism, appears to obtain the nucleotides required for the growth by the uptake of DNA from the environmental (eDNA), an attractive possibility is that its CRISPR/Cas system participates in the degradation of the imported eDNA, which has recently been proposed for other nano-sized prokaryotes, namely members of proposed new order *Nucleotidisoterales* within *Ca. Nanohaloarchaeota* (Xie et al., 2022) and bacterial “vampires” of the Gracilibacteria lineage within Candidate Phyla Radiation (Moreira et al., 2021; Yakimov et al., 2022). In this context, and as if to confirm this assumption, a direct BLASTn search (Altschul et al., 1997) was carried out against the entire collection of nr/nt nucleotides in the NCBI database (August 10, 2022) using the more dissimilar sequences (discontiguous megablast) option. Only one 91.11% identity match, namely the spacer 39 against a *Haloplanus* sp. GDY1 plasmid (CP098517), was found.

In addition to CRISPR/Cas system, *Ca. Nanohalococcus* SVXNc encodes another defense mechanism courtesy of the type II restriction-modification (RM) system, consisting of a Dcm-like DNA methyltransferase (SVXNc_0752) and a restriction endonuclease Mrr (SVXNc_0757). These genes do not form an operon, and both are expressed in the SVXNc transcriptome (Supplementary Table S1). Bipartite methylation of cytosine residues at GDGCHC type II recognition sites was confirmed for *Ca. Nanohalococcus* SVXNc by PacBio sequencing of the genome in three repeats (Figure 3, the methylated cytosine residue is shown in bold and highlighted by color; the guanine residue opposing the methylated cytosine on the complementary strand is italicized). Out of 140 palindromic motifs GDGCHC found in the genome on both DNA strands, 138 motifs were methylated at both DNA strands (Supplementary Table S2). Only one site, GTGCAC, located at the 576132-576137 region within the *ppsA* coding sequence was not methylated on either strand in all three repeats of sequencing. Either this site was somehow protected from methylation, or it could be that this sequence had not been recognized by the methyltransferase as it was the only GTGCAC sequence in the whole genome of *Ca. Nanohalococcus*. Methylation of all other recognition motifs in the genome suggested that this restriction-modification system was fully functional and could protect the genome from phage invaders. The BNXNh nanohaloarchaeon,

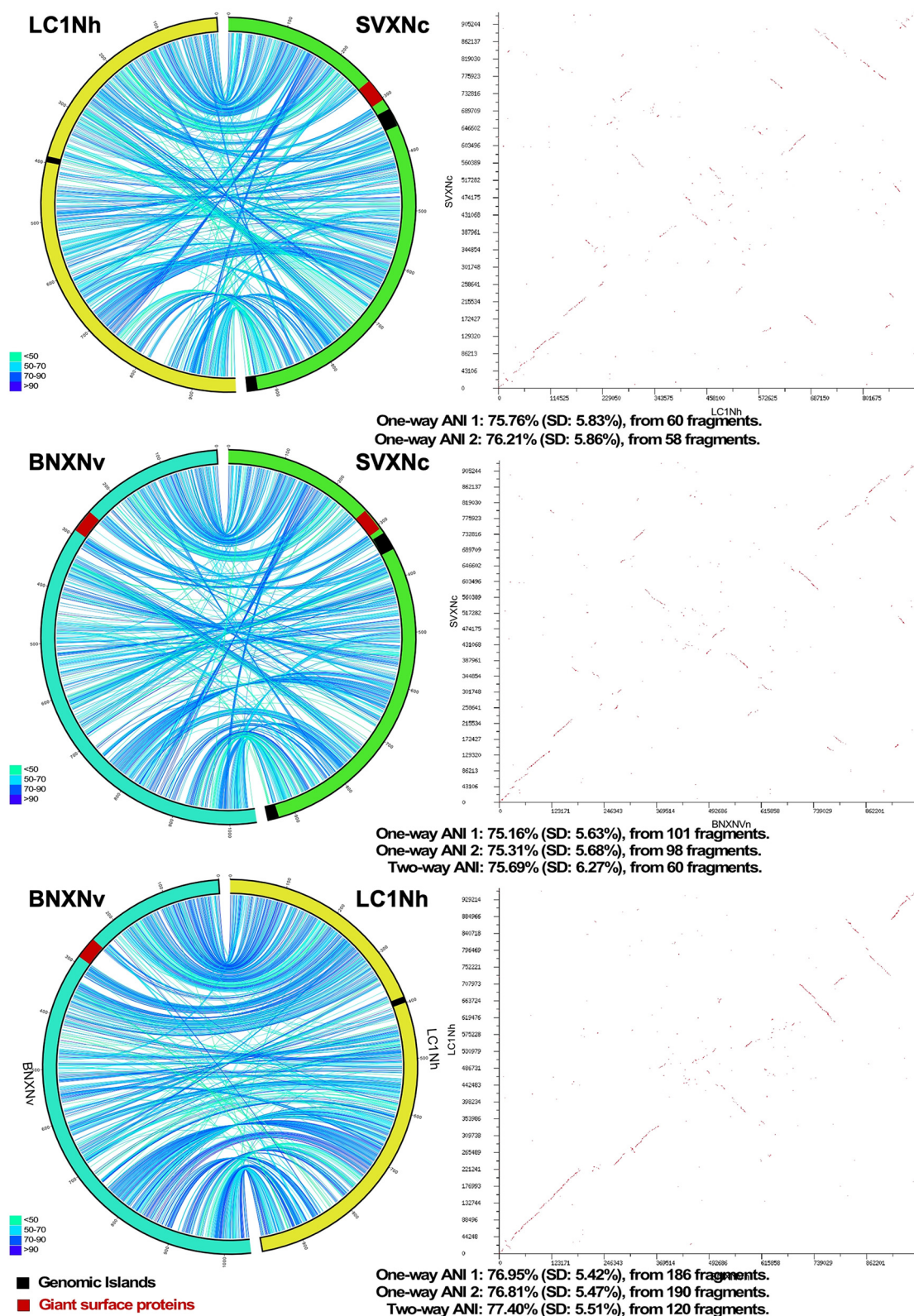
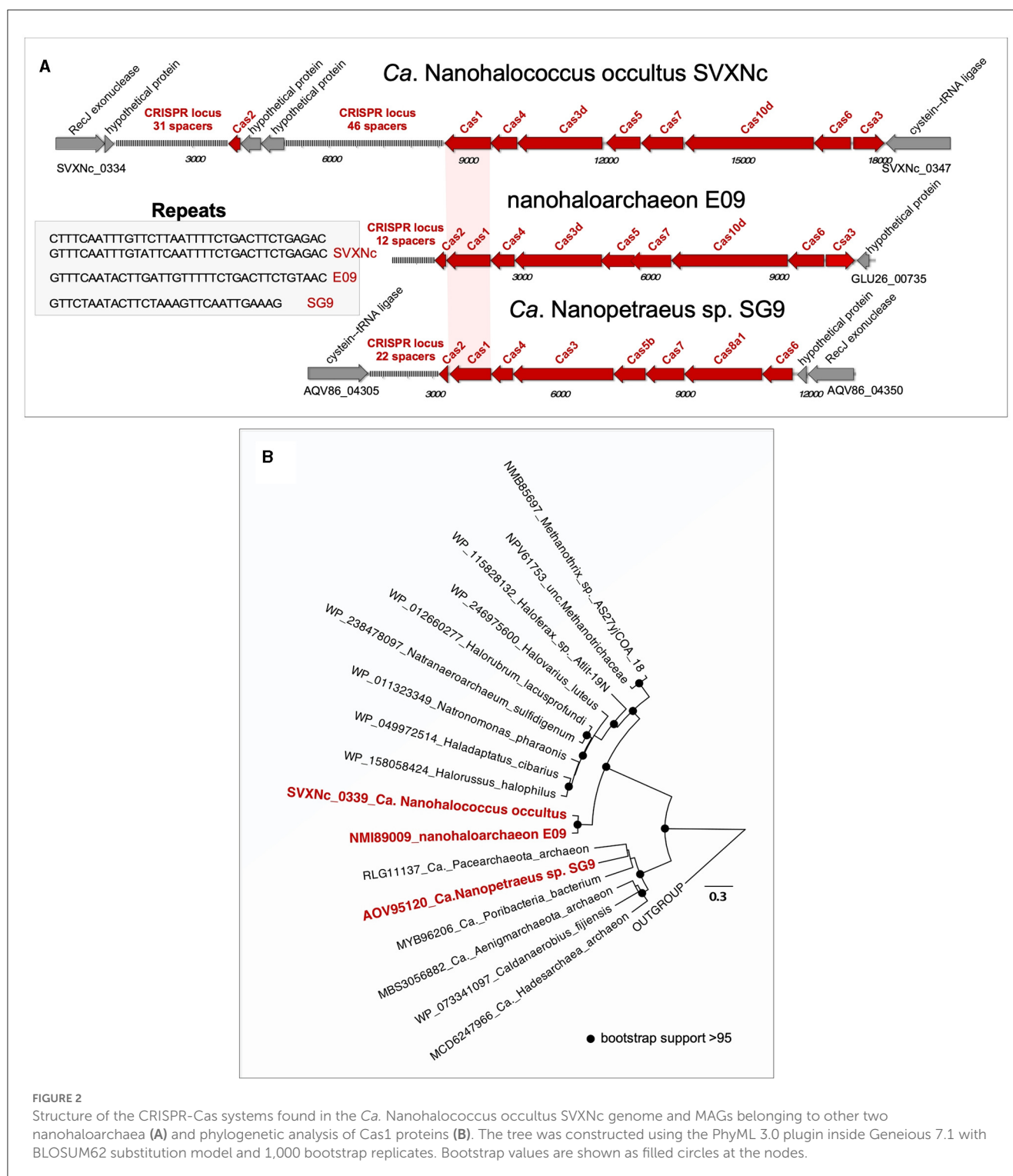


FIGURE 1

Pairwise genome comparison between *Ca. Nanohalococcus occultus* SVXNc, *Ca. Nanohalovita haloferacivicina* BNXNv, and *Ca. Nanohalobium constant* LC1Nh. Circos-based genome alignments and Dot-plot alignments are present in the left and right columns, respectively. Links indicate pairs of orthologous genes between the genomes, and the color is scaled to the percentage of amino acid identity levels (shown as the bottom-left inserts). The genomic islands predicted with IslandViewer4 are shown as black septa, while the positions of giant SPEARE proteins found in the BNXNv and SVXNc genomes are shown as red septa. For the SVXNc and LC1Nh couple, the similarity was too low for the two-way ANI index calculation, and only one-way ANI was shown.



in turn, has another, the type III RM system, consisting of Mod DNA methyltransferase (BNXNv_0801) and Res restriction endonuclease (BNXNv_0802), co-located in the same operon. For this nanohaloarchaeon, methylation of adenine residues can be expected. To our knowledge, these are the first indications of experimentally confirmed DNA methylation patterns and an active RM system across the entire DPANN superphylum.

Reconstruction of the core metabolism of novel cultivated nanohaloarchaea

Protein translocation systems, membrane-associated proteases, and transporters. An overview of the metabolic potential and the core metabolism is illustrated in [Supplementary Figure S1](#). Among the totals of 1,088 and 1,199 proteins predicted from

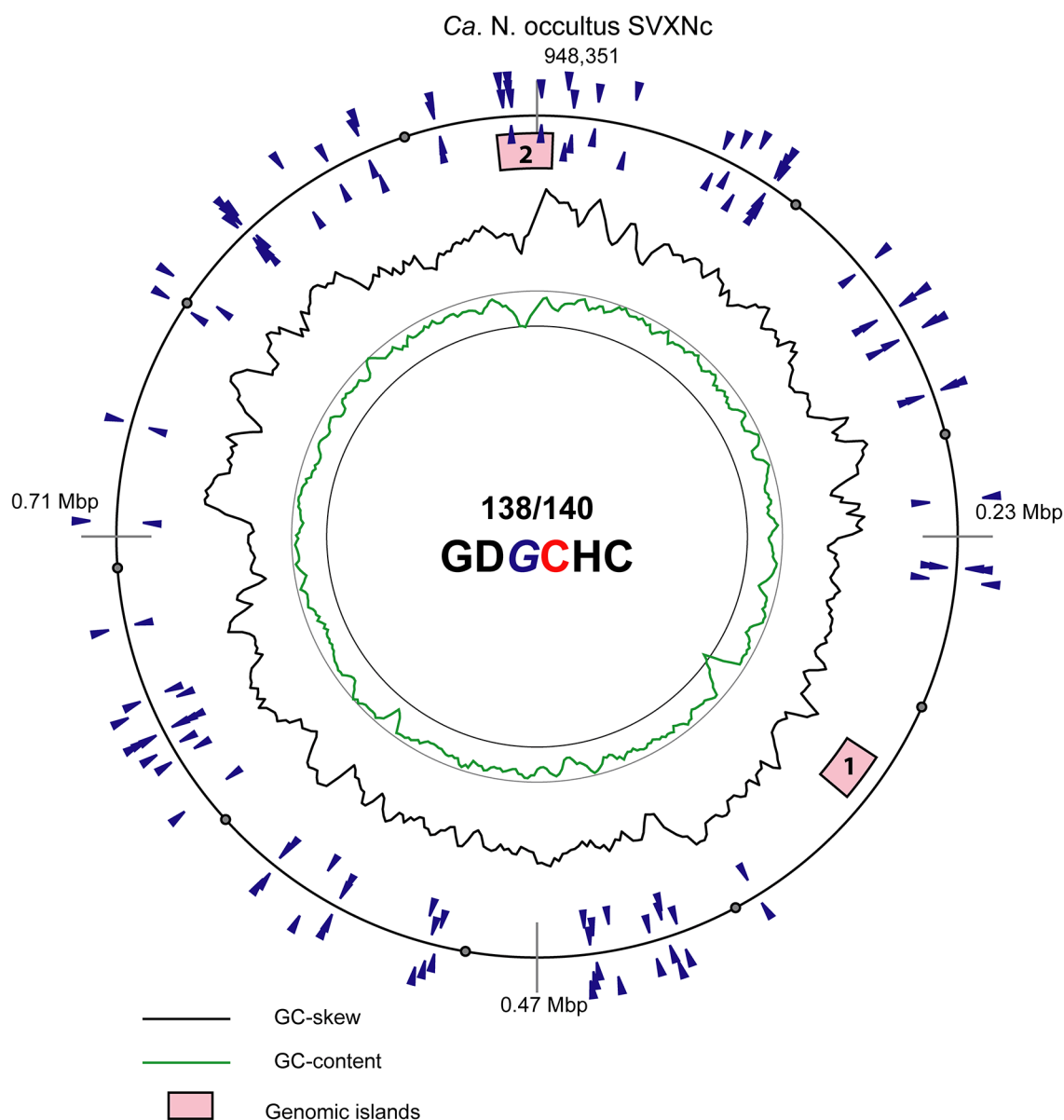


FIGURE 3

Atlas view of the circular chromosome of the strain SVXNc. Histogram graphs show GC-skew and GC-composition fluctuations in a 5,000 bp sliding window. Locations of motifs GDGCHC with bipartite methylated cytosine residues on the direct and reverse complement DNA strands are depicted by blue triangles outside and inside of the outermost ring representing the chromosome. In the motif sequence GDGCHC, the methylated cytosine (C) is shown red, and the guanine residues (G) opposing the methylated cytosine on the complementary strand are italicized. Locations of the two identified genomic islands are depicted by numbered pink boxes.

the SVXNc and BNXNv genomes, respectively, some were annotated by BlastKOALA (Kanehisa et al., 2016) as putative members of the “Membrane Transport” category. Apparently, both organisms possess the general secretion system Sec as the major means of protein secretion, which includes the signal recognition particle complex (SRP), and six different genes belonging to this protein-trafficking system were recognized: SecD (SVXNc_1062 and BNXNv_1183), SecF (SVXNc_1063 and BNXNv_1184), SecG (SVXNc_0958 and BNXNv_1035), SecY (SVXNc_0611,0894 and BNXNv_1117), Sss1 (SVXNc_0114 and BNXNv_0098), the SRP receptor FtsY (SVXNc_0393 and

BNXNv_0691), and the targeting protein Ffh (SVXNc_0398 and BNXNv_0686) as well as the gene for non-coding SRP RNA. The twin-arginine translocation (Tat) system, that exports folded proteins from cells and is represented by a single TatC protein in the SVXNc genome (SVXNc_0793) only, and thus, is likely inactive in both nanohaloarchaea or this gene has another function. Signal peptidases of the SppA type (SVXNc_0702 and BNXNv_0859) and archaeal type I (SVXNc_0770 and BNXNv_0034,0327,0746) are the principal intra-membrane peptidases responsible for processing the majority of exported proteins. Two rhomboid-family proteases (SVXNc_0214 and BNXNv_0161) are the intra-membrane serine

proteases that cleave other proteins, including S-layer glycoproteins (SVXNc_0478,0977 and BNXNv_0277,0530,0625,0988) within their transmembrane domains.

A total of 11 and 12 genes were annotated as components of ABC-type transporter systems in the SVXNc and BNXNv genomes, respectively. Among them, there are genes encoding a putative peptide ABC transport system SalXY (SVXNc_0976,0978-9 and BNXNv_0529,0531-32) and three uncharacterized ABC-2 type transporter complexes. Sugars can be imported into the cytoplasm either by one of these uncharacterized ABC-type transporters, putative permeases PerM (SVXNc_0693 and BNXNv_0869,0978), and/or by major facilitator superfamily (MFS) permeases found in the genomes. Notably, the SVXNc genome harbors only one gene of MFS permease (SVXNc_0852) (similar to *Ca. Nanohalobium constans*), whereas six MFS-coding genes (BNXNv_0572,0893,0955,0995-6,1020) were found in the BNXNv genome. As we mentioned previously (La Cono et al., 2020), this putative transporter system seems to be specific to nanohaloarchaea since we failed to find any close homologs (>25% of aa sequence identity) with annotated functions.

Transport of amino acids and citrate(isocitrate) into the cells of the strains SVXNc and BNXNv likely occurs *via* amino acid permeases PerM (SVXNc_0693 and BNXNv_0869) and tripartite tricarboxylate transporters (SVXNc_1057-8 and BNXNv_1180-1), respectively. The phosphate/sulfate Pit permeases (SVXNc_0158,0539 and BNXNv_0146,0553) likely participate in the transportation of inorganic ions, osmotic homeostasis, and resistance to heavy metals along with K⁺-dependent mechano-sensitive MscS channels (SVXNc_0010,0482,0867 and BNXNv_0365,0395,0547,0911), K⁺-dependent Na⁺-Ca²⁺ ECM27 exchanger (SVXNc_0053 and BNXNv_0458), Na⁺/K⁺: H⁺ antiporter of Kef type (SVXNc_0165-6,0900 and BNXNv_0923-4,0968), NADH-dependent K⁺ transport system of Trk type (SVXNc_0862-3 and BNXNv_0906-7), zinc/iron permease of the ZIP family (SVXNc_0931 and BNXNv_0999), Fe²⁺/Mn²⁺ transporters MgtA and CCC1 (SVXNc_0063,0637 and BNXNv_0226,0424,1013), an iron complex transporter FepB (BNXNv_0613, found in the BNXNv genome only), and P-type heavy metal (cations)-transporting ATPases (SVXNc_0063,0584,0689,0931,1046 and BNXNv_0038,0226,0242,0251,0499,0999,1013). Apart from this transporter machinery, the *Ca. Nanohalococcus occultus* SVXNc genome harbors a fully operational multicomponent sodium/proton antiporter MnhA-G system (SVXNc_0517-23) that utilizes the proton motive force to expel intracellular sodium ions to the extracellular milieu. As revealed by a BLAST search, out of all members of the DPANN superphylum, the Mnh antiporter system was found in only four nanohaloarchaeal genomes. Phylogenetic analysis confirmed that this system was apparently acquired by horizontal gene transfer from haloarchaea.

Energy production and catabolism

The genomes of the strains SVXNc and BNXNv contain no genes associated with the carbon-fixation pathways, pointing toward their strict heterotrophic lifestyle. Both genomes lacked pivotal enzymes needed for the pentose phosphate pathway and the

ribulose monophosphate pathway, so these microbes are incapable of metabolizing pentose sugars. Given that they also lack genes encoding most components of the tricarboxylic acid cycle and any of the respiratory complexes (NADH dehydrogenase complex, functional cytochrome oxidases, and terminal reductases), we infer a strictly anaerobic fermentation-based lifestyle. Membrane-bound proton-translocating pyrophosphatases were also not found, so the maintenance of the chemiosmotic membrane potential, osmotic stress response, and pH homeostasis should fully rely on an archaeal-type H⁺-ATPase (SVXNc_0673,0675-82 and BNXNv_0873-81), Kef-type K⁺/H⁺ antiporters (SVXNc_0165-6,0900 and BNXNv_0924,0968), seven-subunit Mnh-type Na⁺/H⁺ secondary antiporter (SVXNc_0517-23, found only in the SVXNc genome), and possibly several other unidentified systems functioning as proton-translocating membrane pumps.

A complete set of genes of the archaeal version of the dissimilative Embden–Meyerhof–Parnas (EMP) pathway of glycolysis was identified in both genomes. None of the genes of the Entner–Doudoroff pathway were found in any of the oxidative and non-oxidative variants of the pentose phosphate pathway (Supplementary Figure S1). Thus, in the absence of membrane respiratory complexes, the EMP pathway is the only way of gaining energy by fermentation, namely by substrate-level phosphorylation. As it was demonstrated recently (La Cono et al., 2020), the EMP pathway of energy production with some variations at the upstream steps of glucose metabolism exists in all sequenced nanohaloarchaea and also in some halophilic methanogenic archaea but it was missed in haloarchaea (Gonzalez-Ordenes et al., 2018). Employing ADP as the phosphoryl donor, the phosphorylation of both glucose and fructose-6-phosphate is likely catalyzed by only one enzyme, the bifunctional ADP-dependent phosphofructokinase/glucokinase PfkC (SVXNc_0128 and BNXNv_0111) of the ribokinase family (EC 2.7.1.146; EC 2.7.1.147). BNXNv has an additional phosphofructokinase FruK (BNXNv_0774) to phosphorylate fructose-6-phosphate. Fructose-1,6-biphosphate is further converted *via* a fructose-biphosphate aldolase FbaB (SVXNc_0147 and BNXNv_0133) to dihydroxyacetone phosphate and glyceraldehyde-3-phosphate (GAP). These intermediates enter the lower portion of the EMP pathway and are transformed by G6pA dehydrogenase (SVXNc_0140 and BNXNv_0123), phosphoglycerate kinase Pglk (SVXNc_0217 and BNXNv_0164), phosphoglycerate mutase Gmpl (SVXNc_1002 and BNXNv_1066), and enolase Eno (SVXNc_0290 and BNXNv_0311) ending up with phosphoenolpyruvate (PEP). The final step of the glycolysis consisting of the conversion of PEP to pyruvate can be catalyzed by a pyruvate kinase PykF (SVXNc_0488 and BNXNv_0446) and two AMP/Pi-dependent phosphoenolpyruvate synthases/pyruvate water dikinases PpsA (SVXNc_0254,0565,0650 and BNXNv_0201,1153). The synthesis of pyruvate from PEP by the water dikinase is energetically more favorable than by the pyruvate kinase since it can directly synthesize ATP from phosphate and AMP, produced by the ADP-dependent PfkC (Falb et al., 2008; Atomi and Reeve, 2019; La Cono et al., 2020). As mentioned above, the genomes of SVXNc and BNXNv encode the complete archaeal-type H⁺-ATPase complex (SVXNc_0673,0675-82 and BNXNv_0873-81). Thus, in addition to many important homeostatic and anabolic reactions,

the energy generated during the glycolysis can be used to maintain the cytoplasmic pH within a biocompatible range through the F_0 rotor of the A-type ATPase, resembling the metabolism of other strictly fermentative organisms that lack the electron transport chains and are incapable of respiration.

As in many DPANN organisms, the novel nanohaloarchaea have many genes designated for pyruvate metabolism, which include a pyruvate dehydrogenase complex (PDH) consisting of AcoAB, AceF, and Lpd enzymes (SVXNc_0064-67,0758 and BNXNv_0048-51,0239), lactate dehydrogenase LdhA (SVXNc_0992 and BNXNv_0542), NAD-dependent malic enzyme SfcA (SVXNc_0069 and BNXNv_0060), and a short-chain alcohol dehydrogenase FabG (SVXNc_1061 and BNXNv_0428). The PDH complex decarboxylates pyruvate to acetyl-CoA. An ADP-forming acetyl-CoA synthetase AcdA (SVXNc_0068 and BNXNv_0052) likely completes the oxidative pathway of pyruvate metabolism leading to the formation of ATP and acetate. The presence of the other two pyruvate-transforming enzymes indicates that this metabolite could also be used in several reductive pathways using pyruvate as an electron acceptor in its conversion to malate and ethanol possibly to reduce the excess of NADH formed during pyruvate oxidation. It is noteworthy that all these enzymes of pyruvate metabolism were highly expressed in the nanohaloarchaeal strain SVXNc (Supplementary Table S1).

The genes of the EMP pathway in the SVXNc and BNXNv genomes can be used also in the gluconeogenic pathway due to the presence of a fructose-1,6-bisphosphatase Fbp (SVXNc_0148 and BNXNv_0134) that switches between the glycogenesis and the gluconeogenesis. The strains SVXNc and BNXNv have enzymes necessary for the synthesis of glycogen: a phosphomannomutase/phosphoglucomutase ManB (SVXNc_0127,0558 and BNXNv_0110,1160), nucleotidyltransferase (SVXNc_0671 and BNXNv_0476,0884), glycogen synthase RfaG (SVXNc_0131 and BNXNv_0114), and a glycogenin-like protein BNXNv_1210 found only in BNXNv. This finding supports the hypothesis that both these nanohaloarchaea can perform a *de novo* synthesis of glucose and further convert it for intracellular storage in the form of glycogen, a process proposed as a hallmark not only of all sequenced members of *Ca.* Nanohaloarchaeota but also of most other DPANN organisms (Castelle et al., 2015, 2018; Dombrowski et al., 2019). Remarkably, this mode of carbon storage is absent in all known members of haloarchaea (La Cono et al., 2020). In accordance with the ability to synthesize glycogen, both nanohaloarchaea strains used in this study were also capable of the decomposition of glycogen into glucose. This process involves a glycogen debranching enzyme GDB1 (SVXNc_0130 and BNXNv_0113) and various alpha-glucan hydrolases: alpha-amylase of the glycosyl hydrolase (GH) 57 family (SVXNc_0138 and BNXNv_0120), alpha-amylase of the GH13 family (BNXNv_0413, found only in the BNXNv genome), and a glucoamylase of the GH15 family (SVXNc_0760, found only in the SVXNc genome). All the enzymes involved in the metabolism of pyruvate and synthesis/decomposition of glycogen were actively expressed in *Nanohalococcus occultus* SVXNc as revealed by the transcriptome analysis (Supplementary Table S1). This suggests a highly dynamic turnover of glucose and the storage compounds in this organism.

ROS sensing and redox homeostasis

Redox homeostasis is important for all living organisms. In the absence of any of the respiratory complexes, both strains of nanohaloarchaea must stick to a strictly anaerobic fermentative lifestyle. However, they are obligatory-dependent on their aerobic host *Haloferax lucertense* (La Cono et al., 2023). Inhabiting aerobic environments requires an extended tolerance to oxidative stress. In both these archaeal nanoorganisms, thioredoxin most likely plays the most important role in the maintenance of redox homeostasis and in catalyzing various biochemical redox reactions controlled by a thiol-disulfide isomerase TrxA (SVXNc_0146,0837 and BNXNv_0132,0394,0617), thioredoxin reductase TrxB (SVXNc_0516 and BNXNv_0612), ribonucleotide reductase NrdA (SVXNc_0533 and BNXNv_0637), peptide methionine sulfoxide reductase MsrA (SVXNc_0607,0994 and BNXNv_0501,0935,1059), and a peroxiredoxin Bcp (SVXNc_0159,0582 and BNXNv_0618). The thioredoxin system is involved in peroxide removal by a superoxide dismutase SodA (SVXNc_0514 and BNXNv_0610). Both strains have an additional alkyl hydroperoxide reductase AhpF (SVXNc_0616 and BNXNv_0455) that confers a more active peroxide scavenging during oxidative stresses (Kashima and Ishikawa, 2003). Other important mechanisms of redox homeostasis involve the glutathione (GSH)-dependent systems (Zhang and Forman, 2012). GSH is synthesized by glutamate-cysteine ligase and glutathione synthase GshAB (SVXNc_0869-70 and BNXNv_0913,0915). GSH serves as an oxidizing substrate in reactions controlled by glutaredoxin GrxC (SVXNc_0734 and BNXNv_0819). GSH may play an important role in the removal of toxic methylglyoxal compounds resulting from the isomerization of several glycolysis intermediates to avert oxidation stress (Rawat and Maupin-Furrow, 2020). This process is controlled by a GSH-dependent lactoylglutathione lyase GloA (SVXNc_0045 and BNXNv_0036). The strain SVXNc has an additional hydroxyacylglutathione hydrolase SVXNc_0657 that presumably controls a similar reaction of conversion of methylglyoxal compounds to (R)-S-lactoylglutathione with a further degradation to lactate. All the abovementioned genes are expressed in SVXNc (Supplementary Table S1).

The limited anabolic capability of the cultivated nanohaloarchaea

The characteristic genomic features of *Ca.* *Nanohalococcus occultus* SVXNc and *Ca.* *Nanohalovita haloferacivicina* BNXNv are consistent with the predictions made before for uncultured and cultured nanohaloarchaea and also for other DPANN organisms (Castelle et al., 2015, 2018; Dombrowski et al., 2019; Hamm et al., 2019; La Cono et al., 2020). Despite the presence of a full set of genes for chromosome maintenance and replication, a significant genome reduction was observed, making these nanoorganisms unable to synthesize the most necessary metabolic precursors including the majority of amino acids and lipids, nucleotides, and cofactors. Indeed, the enzymatic suite of *de novo* biosynthesis of amino acids is limited in both strains (Supplementary Figure S1) with only a few genes present including an asparagine synthase AsnB (SVXNc_0203 and BNXNv_0940) and a threonine

dehydratase Tdh (SVXNc_0578), found only in the SVXNc genome. BNXNv has an operon of genes involved in aromatic acid biosynthesis that includes two prephenate dehydratases PheA and TyrA (BNXNv_0067,69), aspartate/tyrosine/aromatic acid aminotransferase AspB (BNXNv_0068), and a chorismate mutase (BNXNv_0070). Both nanohaloarchaea have an amino acid transporter PotE of the APC permease family (SVXNc_0483 and BNXNv_0526) that likely takes part in the uptake of aromatic amino acids from the extracellular environment. Remarkably, the phylogenetic analysis of the PotE revealed that in addition to nanohaloarchaeon M3_22 (Feng et al., 2021), all other DPANN members lack this type of transporter suggesting that this permease was likely acquired by horizontal gene transfer from haloarchaea.

Regarding the metabolism of purine and pyrimidine, we identified only a few genes encoding various kinases involved in the inter-conversion of nucleoside phosphates, adenylate kinase, thymidylate synthases, kinases, and thymidine kinases, enzymes involved in the downstream stages of nucleotides synthesis or nucleotide salvage (Supplementary Figure S1). Similarly, the capability for *de novo* synthesis of cofactors is virtually absent, except for several genes involved in completing the synthesis and maturation of the most-common cofactors. For example, both nanohaloarchaea have a molybdopterin-synthase MoeB (SVXNc_0838 and BNXNv_0396). Molybdopterin is an enzymatic cofactor of several archaeal oxidases and a component of the tungsten cofactor that catalyzes ferredoxin-dependent oxidoreductases oxidizing aldehydes, formaldehyde, and glyceraldehyde-3-phosphate (Buessecker et al., 2022). Pterin-4a-carbinolamine dehydratase PhhB (SVXNc_0711 and BNXNv_0848) synthesizes tetrahydrobiopterin that is another important cofactor of amino acid metabolism. The final step of the synthesis of the redox cofactor NAD⁺ is controlled by a nicotinamide mononucleotide adenyltransferase NadR (SVXNc_0143 and BNXNv_0126), which was found in both genomes. In addition to this gene, the BNXNv genome has a phosphoribosyltransferase/nicotinamidase PncAB (BNXNv_0901-2) contributing to additional steps of NAD⁺ biosynthesis from nicotinate. Both genomes harbor a thymidylate synthase ThyA (SVXNc_0152 and BNXNv_0141) and a glycine/serine hydroxymethyltransferase GlyA (SVXNc_0154 and BNXNv_0142) involved in the turnover of tetrahydrofolate derivatives. However, this system is more developed in the SVXNc genome due to the presence of a dihydrofolate reductase FoaA (SVXNc_0151) and a methylenetetrahydrofolate dehydrogenase FodD (SVXNc_1072), which are absent in the BNXNv genome.

Tetrahydrofolate derivatives are the donors or acceptors of C₁-groups in many biosynthetic pathways, particularly in the synthesis of S-adenosyl-methionine (SAM) and in serine-glycine turnover. The SAM synthesis requires cobalamin as a cofactor, and the genomes of both strains have a PduO-based adenosylcobalamin salvage system (SVXNc_0941 and BNXNv_1010); however, an archaeal SAM-synthetase MetK was found only in the BNXNv genome (BNXNv_0766). SAM is the universal donor of methyl groups used by the two SAM-dependent methyltransferases found in the SVXNc and four such genes found in the BNXNv genomes (SVXNc_0183,1081 and BNXNv_0119,0321,0522,0971). One important SAM-based biosynthetic pathway shared by both nanohaloarchaea is the synthesis of diphthine controlled by DPH5

diphthine synthase/diphthamide methyltransferases (SVXNc_1032 and BNXNv_1096). Diphthine is a compound found in archaeal and eukaryotic cells that is indispensable for the functionality of the elongation factor 2 (Su et al., 2013). SAM also provides with methyl groups in the DNA restriction-modification systems, found in both microorganisms.

As was already mentioned when analyzing the *Ca. Nanohalobium constans* LC1Nh genome (La Cono et al., 2020), a caveat in this and other reconstructions are that more than one-third of the proteins encoded in the SVXNc and BNXNv genomes (35.9 and 37.0%, respectively) were annotated as hypothetical proteins and could not be assigned to any functional category. This raises the question of whether these organisms might encode some novel enzymes that drive either known or yet unknown metabolic pathways.

Cell-surface structures of the cultivated nanohaloarchaea and their predicted role in the interaction with the host

In sharp contrast to their simplified core metabolism and limited anabolic capabilities, the nanohaloarchaeal strains SVXNc and BNXNv possess a sophisticated set of genes related to the construction of an elaborated cell surface, which seems to be a common hallmark of many DPANN archaea (Huber et al., 2002; Jahn et al., 2008; Golyshina et al., 2017; Jarett et al., 2018; Hamm et al., 2019; La Cono et al., 2020). The SVXNc and BNXNv genomes encode various glycosyl transferases (11 and 12 genes, respectively) which belong to the GT1, GT2, and GT8 families. According to transcriptomic data, the majority of them are expressed in the SVXNc cells (Supplementary Table S1). This indicates that this nanohaloarchaeon expends substantial amounts of energy on the biosynthesis of precursors for glycosylation processes and the synthesis of polysaccharides and glycoproteins as principal components of the extracellular matrix, including N-glycosylated S-layer proteins (SVXNc_0478,0977 and BNXNv_0277,0530,0625,0988) and many other predicted surface proteins.

The BNXNv genome contains an operon comprising genes of phytoene dehydrogenase CrtD, squalene synthetase CrtB, and geranylgeranyl pyrophosphate synthase IsdA (BNXNv_0461-3). Using alkyldecaprenyl pyrophosphate as a substrate, these enzymes might participate in the synthesis of extracellular squalene-containing lipid monolayer, similar to that found in many extremely halo(natrono)philic archaea (Gilmore et al., 2013). None of the known homologs of these genes were found in the SVXNc genome and only genomes of three uncultured nanohaloarchaea, namely *Ca. Nanosalina* (Narasimgarao et al., 2012) and two unclassified strains, recovered from salterns near Alicante (Spain) (Feng et al., 2021), had these genes. The SVXNc and BNXNv genomes have 15 and 10 genes, respectively, likely involved in the synthesis and assembly of archaeella and filament proteins. Four of these genes (SVXNc_1016-7 and BNXNv_1081-2) were situated in large operons (SVXNc_1003-31 and BNXNv_1067-93) that host up to 22 hypothetical proteins resembling the structure of euryarchaeal archaeella operons (Albers and Jarrell, 2018). It remains

to be determined whether these flagellar structures are used for motility of detached SVXNc and BNXNv cells, or whether they facilitate attachment of these archaea to the host cells.

As it was demonstrated recently (La Cono et al., 2023), both transmission and scanning electron microscopy revealed the spatial distribution of nanohaloarchaeal cells on a single *Haloferax lucentense* host cell, suggesting a strong intercellular interaction (Figure 4). Given this, we looked for genomic evidence of putative mediators of these cross-species cellular interactions. At least six strain-specific secreted or membrane-bound surface proteins, containing either a polycystic kidney disease (PKD) domain (SVXNc_0035) or multiple (up to five) concanavalin A-like/lectin laminin (LamG) domains (SVXNc_0801,0806,0807,0813,0831), were found in the SVXNc genome. The surface proteins containing these domains could serve a variety of purposes by mediating interactions with carbohydrates or glycosylated proteins, and they were also predicted to be involved in the specific recognition and fixation of DPANN organisms to their hosts (Castelle et al., 2015, 2018; Golyshina et al., 2017). Remarkably, the genes encoding three of the LamG domain-containing proteins are organized in one operon together with the genes of 1,693 amino acids long trimeric autotransporter adhesin-like protein (SVXNc_0803) and 1,386 amino acids long protein, containing Microbial Surface Components Recognizing Adhesive Matrix Molecules (MSCRAMM) clumping factor A domain and knob domain of R-type pyocins (SVXNc_0798). MSCRAMM and knob-like domains occur in host-recognition and binding proteins, where the knob-like domain may interact with sialic acid (Buth et al., 2018; Foster, 2019; Redero et al., 2020). These nine genes together occupy 5.85% (55,476 bp) of the entire SVXNc genome. Correspondingly, in the BNXNv genome, there are at least 14 genes encoding the proteins harboring the PKD and LamG domains, which occupy 6.15% (64,343 bp) of the entire BNXNv genome.

Giant membrane-anchored extracellular proteins

We believe that the giant surface proteins found in both genomes of the cultured nanohaloarchaea deserve special attention. Although these giant proteins were absent in the *Ca. Nanohalobium constans* genome (La Cono et al., 2020), their presence in many other nanohaloarchaea has been documented (see Hamm et al., 2019 for further references). The SVXNc_0300 cell-bound protein (9,409 amino acids) is the largest protein among all of the sequenced nanohaloarchaea and the largest protein ever found in any of the archaea that have been cultivated (Bolhuis et al., 2006). The BNXNv_0298 giant protein, also presumed to have a surface localization in the cellular envelope, is slightly shorter but is nevertheless enormous (8,880 amino acids). The predicted structures of these two giant proteins are somewhat similar each to other, but the BNXNv giant protein lacks 529 amino acids long stretch of tandem immunoglobulin (Ig-like)/fibronectin-like domains (indicated with a hatched box in Figure 5). Thus, for a detailed structural analysis, we have focused on the longest protein.

The N-terminus of the SVXNc_0300 protein apparently contains a Sec signal peptide, which would translocate the gigantic ectodomain to the outside of the cell. The signal peptide is followed by a single immunoglobulin-like (Ig-like) domain and then by

two β -transducin (WD40-like) β -propeller domains, which contain pyrroloquinoline-quinine (PQQ)-like motifs. The WD40 repeat is a short structural motif of approximately 40 amino acids, often terminating in a tryptophan-aspartic acid (W-D) dipeptide. These repeats are found in extremely functionally diverse proteins and typically mediate interactions with various substrates, including proteins and DNA (Xu and Min, 2011). Nevertheless, the bulk of the SVXNc_0300 protein consists of tandem repetitions of Ig-like domains, and such tandem repeats are known (in invasins, intimins and S-layer proteins, and other proteins) to mediate adhesion (Nesta et al., 2012). We used AlphaFold2 to model a 700-amino acids region accommodating seven Ig-like domains (100 amino acids per one Ig-like domain), so there should be around 75 Ig-like domains in total. This extended region is followed by a transmembrane domain and then by the VirB4/TraG-like ATPase domain, which is one of the essential components of type IV secretion systems. The topology is consistent with the ATPase domain facing the cytosol (as would be expected). Finally, at the tip of the protein, there is a coiled-coil domain. Hamm et al. (2019) suggested that this domain is a restriction endonuclease, but this does not seem to be the case. In the homology searches, these authors obtained a match to a helicase domain of a type III restriction-methylation system and concluded that the corresponding domain of their SPEARE protein functions as a restriction enzyme. We reanalyzed the SPEARE protein of *Ca. Nanohaloarchaeum antarcticus* and found no endonuclease domain; instead, as in the cases of SVXNc_0300 and BNXNv_0298, the protein contains a well-conserved VirB4/TraG-like ATPase domain. Thus, the giant proteins in our nanohaloarchaea, along with the SPEARE proteins of *Ca. Nanohaloarchaeum antarcticus* and *Ca. Halopetraeus SG9* (Oren and DiRuggiero, 2015), appear to be a part of the type IV secretion system, and further supports that it is a secretion system comes from the fact that the gene immediately downstream of the SVXNc_0300 giant protein gene (as well as BNXNv_0298) encodes another AAA+ ATPase (SVXNc_0299 and BNXNv_0297) of the VirB4/TraG- family. In addition to the AAA+ ATPase domain, the latter protein contains an N-terminal domain (PF03135.16), found in VirB4/TraG of *Rhizobium radiobacter* (P0A3W0), leaving little doubt that it is a secretion ATPase, rather than some other AAA+ ATPase. We then looked for other proteins encoded transcriptionally downstream of the SVXNc_0300 giant protein. Of special note is a proteasome activator protein SVXNc_0296, which could regulate the degradation of host proteins. Thus, we propose that the giant cell-bound proteins in nanohaloarchaea play a key role in subjugating the host cells by binding to their surface via the Ig-like domains and secreting yet unknown effectors into the host cells using the VirB4-like ATPase domain. We hypothesize that, in addition to the VirB4/TraG-like ATPase domain mentioned above, these giant proteins lack other enzymatic domains, in particular, the serine protease and restriction endonuclease domains, predicted in the previous study (Hamm et al., 2019). Accordingly, the naming of these giant proteins as SPEARE (after serine protease, adhesion, and restriction endonuclease) does not reflect their actual function. Rather than being the instrument responsible for intracellular invasion into the host (Hamm et al., 2023), these proteins likely function as secretion channels (straws) through

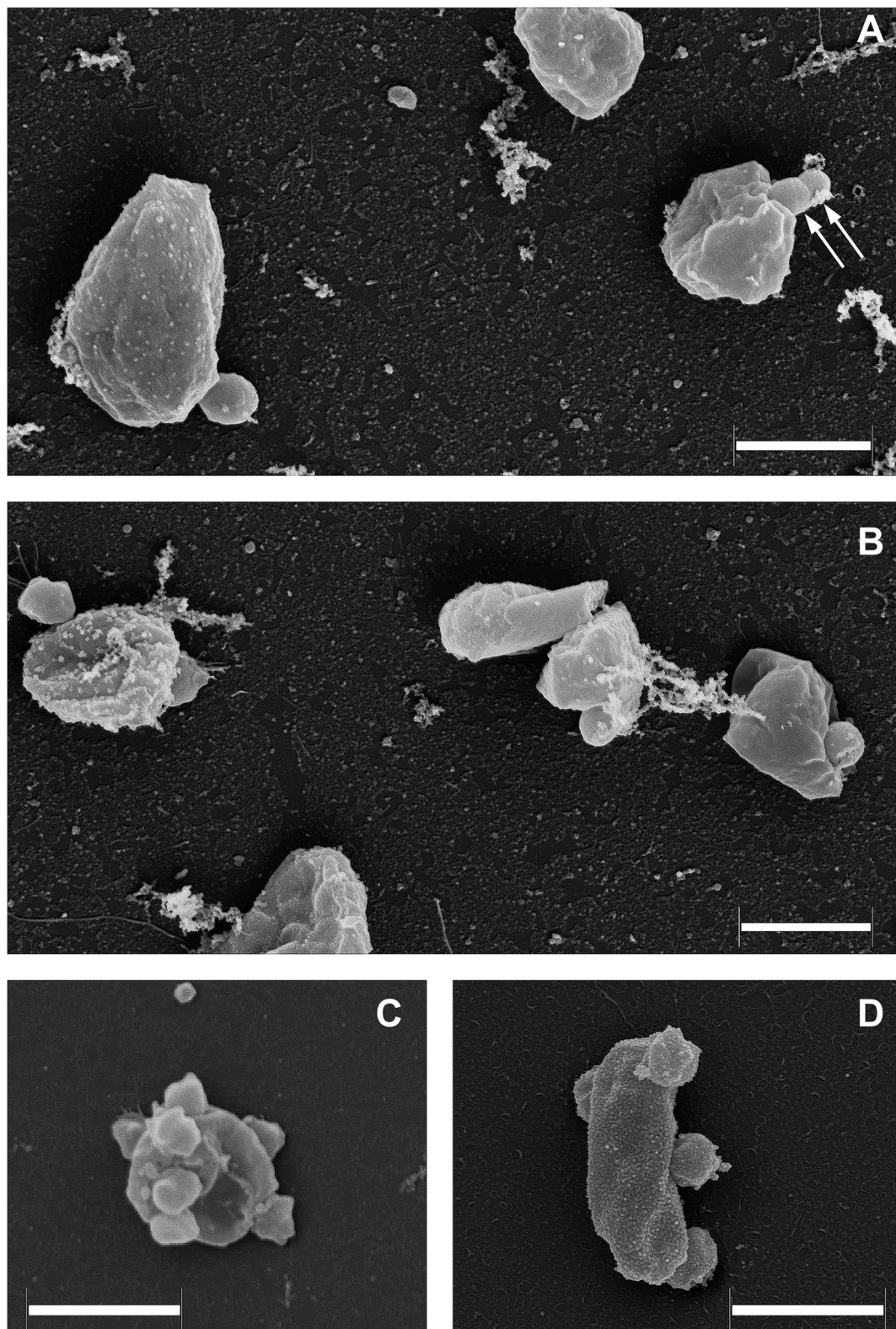
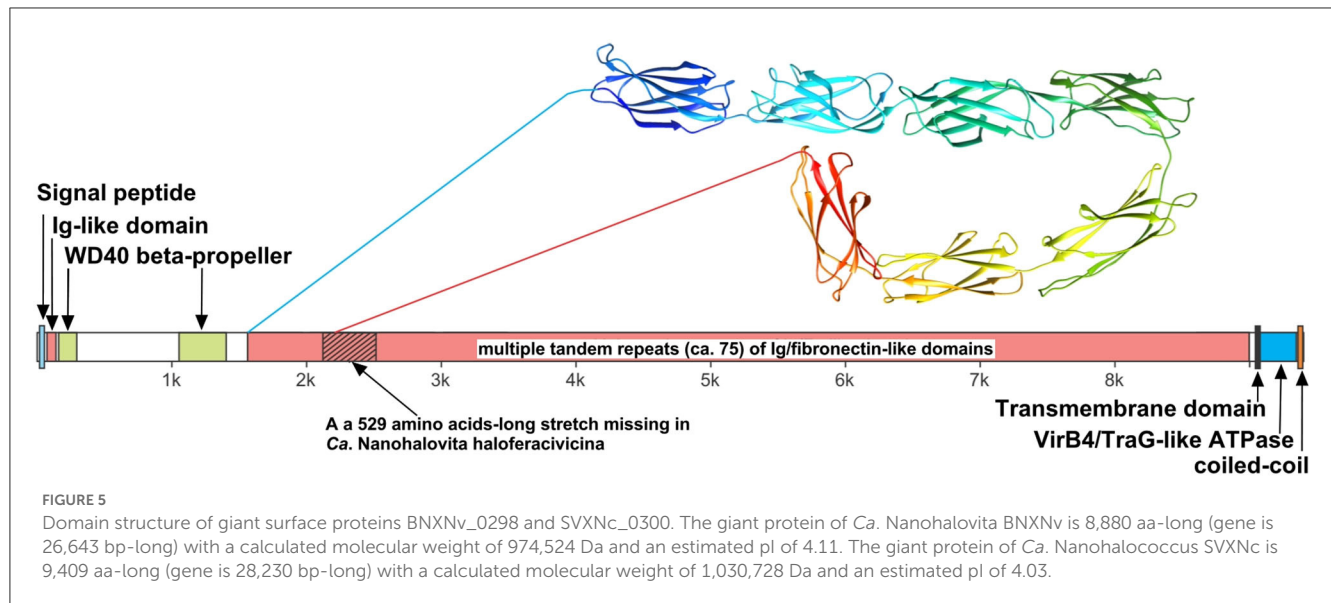


FIGURE 4

Field emission scanning electron micrographs of *Ca. Nanohalovita haloferacivicina* BNXNv (A, B) and *Ca. Nanohalococcus occultus* SVXNc (C, D) attached to their respective hosts, *Haloferax lucertense* BNX82 and SVX82. The images show a strong difference in the median multiplicity of host-attached nanosymbionts: *Ca. Nanohalovita* BNXNv—1–2 cells/host cell; *Ca. Nanohalococcus* SVXNc—4–7 cells/host cell. Dividing BNXNv cells are indicated by white arrows (A). Scale bars represent 1,000 nm.



which the nanohaloarchaeon can inject effector(s) that favorably affect the host metabolism. The latter possibility is supported by our recent results demonstrating a 50-fold increase in the production of glucose-6-phosphate in *Haloferax lucentense* cells associated with the SVXNc symbiont compared with axenic culture of the host (La Cono et al., 2023).

Functional diversity emphasizing the uniqueness of the cultivated nanohaloarchaea

The findings of the current study double the number (from 2 to 4) of the cultivated nanohaloarchaea and enrich the knowledge of their functional diversity. Although all four exhibited very similar fermentative-type catabolism, the novel cultivated nanohaloarchaea harbor unique genomic loci, neither observed before in the *Candidate* Nanohaloarchaeota phylum nor in any other members of the DPANN superphylum. Among them, noteworthy is the phosphoenolpyruvate phosphotransferase sugar transport system (PTS) in the *Ca. Nanohalovita* genome (BNXNv_0773-8). The PTS is present in many halophilic archaea, but it is not common for the entire DPANN superphylum. The only exception is the uncultivated nanohaloarchaeon M3_22 (Feng et al., 2021), which likely belongs to the same genus *Ca. Nanohalovita* (La Cono et al., 2023). In analogy to what is known as haloarchaeal PST (Eichler and Maupin-Furlow, 2013), the PTS of *Ca. Nanohalovita* was predicted to mediate the transfer of phosphoryl groups from phosphoenolpyruvate to imported sugars (Figure 6A). Although it was demonstrated that the PTS system in *Haloferax volcanii* is involved in the metabolism of fructose and/or galacticol (Eichler and Maupin-Furlow, 2013), participation of this system in transportation/metabolism of other monosaccharides in BNXNv cannot be ruled out.

Of particular interest is a complete set of genes responsible for the synthesis of sialic acid from glucose

and UDP-*N*-acetylglucosamine found in the SVXNc genome (SVXNc_0559,0561,1069,1077-9) (Figure 6B). Apart from some mostly pathogenic bacteria, the ability to synthesize sialic acid has not been found in archaeal genomes, with the exception of the human intestinal methanogen *Methanobrevibacter smithii* and eight haloarchaeal species. The latter finding is very intriguing as it suggests that *Ca. Nanohalococcus* has developed mechanisms to disguise its surface with carbohydrate moieties that mimic those encountered in its haloarchaea host's glycan landscape, as was proposed elsewhere (Samuel et al., 2007). All these genes were found to be expressed in the SVXNc transcriptome (Supplementary Table S1), indicating that the synthesis of sialic acid occurs in this organism. Interestingly, the operon harboring the glycosyltransferase and *N*-acetylneuraminate synthases that are involved in the biosynthesis of sialic acid was found within the predicted horizontally transferred second genomic island (Figure 3).

The ability to adapt to environmental changes, not least in relation to stress responses and interactions with host organisms, requires ectosymbiotic nanohaloarchaea to control the expression of various host genes. It appears that in the newly characterized nanohaloarchaea, this occurs *via* at least two different mechanisms that modulate certain cellular pathways. *Ca. Nanohalococcus* has an actively expressed *cyaB* gene (SVXNc_0489, value of 344 ± 76 FPKM [fragments per kilobase of gene per million mapped fragments]) encoding an adenylate cyclase class IV, which synthesizes the intracellular signal transmitter cAMP. The role of cyclic nucleotides as second messengers for intracellular signal transduction is well described in bacteria and archaea, including haloarchaea (see Braun et al., 2019 for further references), where it was shown that cAMP serves as a catabolite de-repression signal that triggers the expression of many metabolic pathways associated with the cell cycle and osmotic adjustment (Baumann et al., 2007). To our knowledge, the presence of adenylate cyclase class IV has not been reported before in any of the genomes of DPANN members.

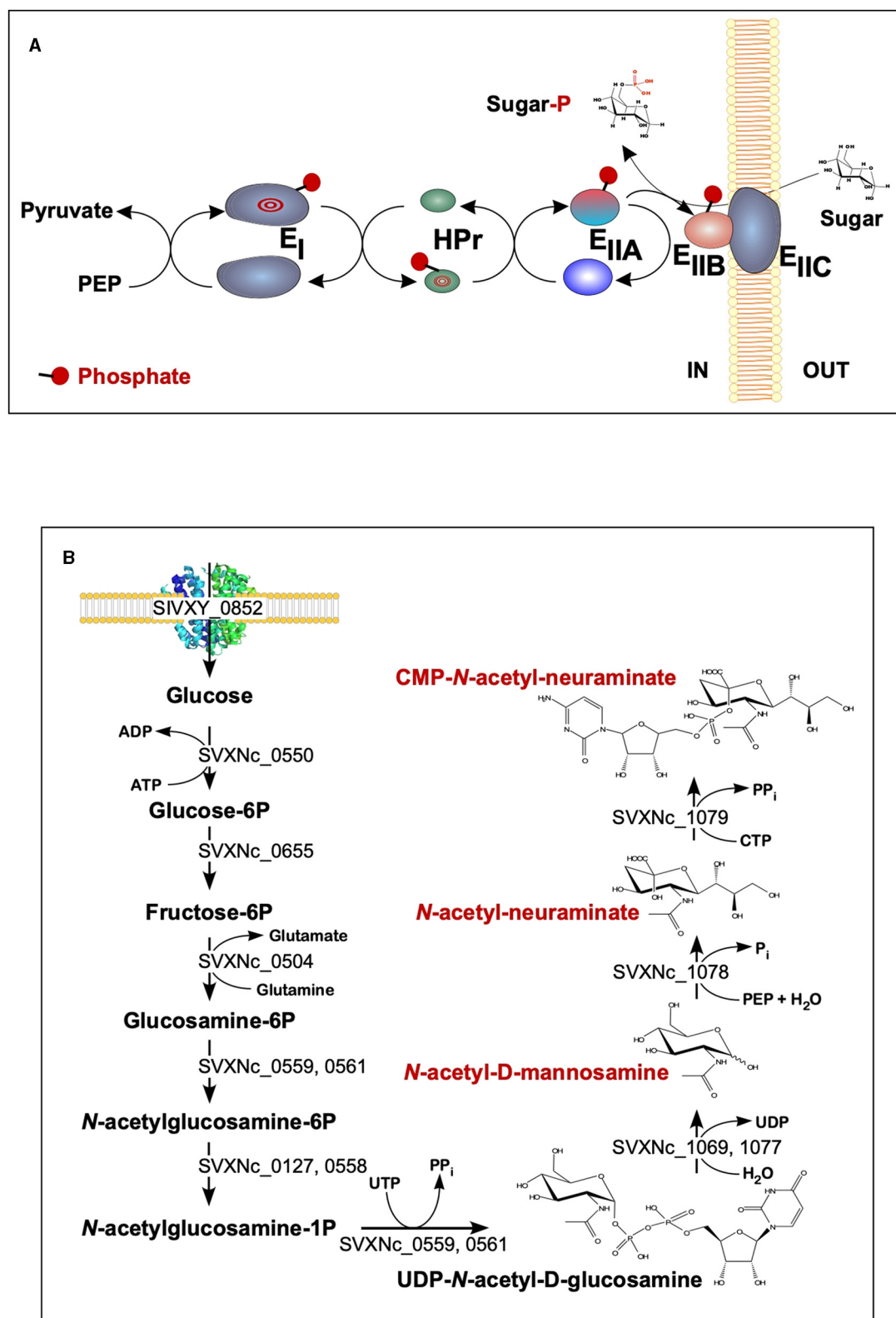


FIGURE 6

Schematic diagram of the BNXNv phosphotransferase (PTS) system (A) and the proposed pathway for sialic acid synthesis by SVXNc inferred from its genome analysis and transcriptome data (B). The BNXNv PTS system seems to be responsible for the simultaneous transport and phosphorylation of sugar substrates. A series of enzyme intermediates, including E1 (BNXNv_0770), HPr (BNXNv_0769), EIIA (BNXNv_0768), EIIB (BNXNv_0765), and EIIC (BNXNv_0767), were predicted to be phosphorylated.

Analyzing both genomic and transcriptomic data, at least five genes of non-coding regulatory RNA (ncRNA) were predicted in the SVXNc genome by two-dimensional structures typical of ncRNA molecules (Supplementary Table S3). They exhibited an extremely high level of expression that was two to three orders of magnitude higher than the average level calculated for all other genes. Homologous versions of the genes SVXNc_nc0001 and SVXNc_nc0002 with characteristic 2D structures were also found in the *Ca. Nanohalovita* genome. However, the search for similar ncRNAs using BLAST did not give significant matches with any of the prokaryotes that have been sequenced, clearly indicating their organism-specific nature. Non-coding regulatory RNAs play an important role in the regulation of gene transcription and mRNA translation in archaea (Gelsinger and DiRuggiero, 2018). Using the StructRNAfinder web server (Arias-Carrasco et al., 2018), only SVXNc_nc0002 was predicted with high confidence (E value of $2e^{-32}$) to be an archaeal signal recognition particle RNA (SRP). The function of SRP is to delay protein translation until the ribosome-bound SRP has an opportunity to associate with the membrane-resident SRP receptor, such as the Sec-translocating system. The other four ncRNAs were not affiliated with any known families of ncRNAs, although their 2D structure resembled the structures of ribosome-associated ncRNA (rncRNA) molecules. All of them have typical rncRNA H/ACA regions, consisting of two hairpins and two single-stranded regions termed a hairpin-hinge-hairpin-tail structure (Supplementary Table S3).

Conclusion

This study has doubled the number of cultivated genera representing the phylum *Ca. Nanohaloarchaeota*, an extremely halophilic archaeal lineage within the DPANN superphylum. Omics analyses of new nanohaloarchaea showed that, despite their rudimentary catabolic capabilities and the absence of a minimal set of enzymes required for the biosynthesis of nucleotides, amino acids, lipids, vitamins, and cofactors, they possess numerous unique adaptive mechanisms, necessary for their ectosymbiotic lifestyle. All genomic loci for the latter have likely been acquired through the horizontal gene transfer from other members of an extremely halophilic microbial community, suggesting a high evolutionary dynamism and plasticity of their reduced genomes. In addition, all our cultured nanohaloarchaea illustrate their common evolutionary history of adaptation to the hypersaline environment and interspecies interactions with polysaccharidolytic haloarchaea, which can now be confidentially assigned to the classical types of either mutualistic symbiosis or pure commensalism.

Experimental procedures

Laboratory cultivation of archaeal strains

Both nanohaloarchaeal strains, *Ca. Nanohalococcus occultus* SVXNc and *Ca. Nanohalovita haloferacivicina* BNXNv, were cultivated in binary cultures on the LC medium (La Cono et al., 2020), supplemented with D-xylose (2 g L^{-1}) at 37°C without shaking together with their corresponding hosts, *Haloferax*

lucentense SVX82 and BNX82, respectively (La Cono et al., 2023). The presence of nanohaloarchaea in the culture was monitored by light microscopy and PCR amplification with the nanohaloarchaea-specific primers as described previously (La Cono et al., 2020, 2023).

Field emission scanning electron microscopy

The grown cells were fixed with 2% (v/v, final concentration) freshly prepared paraformaldehyde. The fixative was removed by washing twice with LC mineral medium. Following that, a final fixation with aqueous osmium tetroxide was carried out (four parts LC mineral medium and one part 5% [w/v] aqueous osmium tetroxide) for 30 min at 25°C . The fixed material was then washed with LC mineral medium and placed onto poly-L-lysine-coated coverslips for 10 min, followed by treatment with 1% (v/v) glutaraldehyde to cross-link microbes with poly-L-lysine coating. This step prevents cells from being washed away during the dehydration and critical-point drying of the attached microorganisms. Dehydrating was achieved using a series of acetone–water mixtures and pure acetone (10, 30, 50, 70, 90, and 100% [v/v] acetone) on ice for 10 min for each step. Once in the 100% acetone, samples were allowed to reach 25°C (evaporation of acetone has cooled the sample), replenishing with fresh 100% acetone. Samples were then subjected to critical-point drying with liquid CO_2 (CPD 030, Bal-Tec, Liechtenstein). Dried samples were covered with a gold–palladium film by sputter coating (SCD 500 Bal-Tec, Liechtenstein) before examination in a field-emission scanning electron microscope Zeiss Merlin (Carl Zeiss, Oberkochen) using the Everhart-Thornley SE detector and the in-lens secondary electron detector in a 50:50 ratio with an acceleration voltage of 5 kV. Contrast and brightness were adjusted with Adobe Photoshop CS5.

DNA extraction and sequencing

Genome sequencing assembly and curation as well as the phylogenetic, phylogenomic, and proteomic analyses are described in detail in our precedent study (La Cono et al., 2020, 2023). Shortly, the DNA for genome sequencing was extracted from 2.0 ml of corresponding co-cultures collected at a fixed time (120 h), corresponding to mid-log phases of growth using a GNome DNA kit (MP Biomedicals, USA). Extracted DNA was dissolved in $50\text{ }\mu\text{l}$ of TE buffer (10 mM Tris-HCl, 1 mM EDTA [pH 7.5]) and quantified using a NanoDrop ND-1000 spectrophotometer (Celbio). The quality of the extracted DNA was checked by electrophoresis in a 1.0% agarose gel. Whole-genome shotgun sequencing of the binary, binary, and axenic cultures was done by FISABIO (Valencia, Spain) using the Illumina[®] MiSeq System platform (San Diego, CA, United States) with $2 \times 300\text{ bp}$ short insert paired-end libraries (MiSeq Reagent Kit v3). FISABIO also performed the quality assessment and the sequence joining (forward R1 and reverse R2). Quality assessment was performed with the PRINSEQ-lite program using the following parameters:

Min_length:50 bp trim_qual_right:30 bp trim_qual_type:mean, trim_qual_window:20 bp. Forward and reverse reads were joined by the FLASH program applying default parameters. All two genomes were assembled as single circular chromosomes and analyzed as described previously (La Cono et al., 2020; Sorokin et al., 2021).

Total RNA extraction and sequencing

For total RNA extraction, 70 ml of the binary culture of *Ca. Nanohalococcus occultus* SVXNc + *Haloferax lucertense* SVX82 (three replicates) was cultivated, using the LC medium containing D-xylose (2 g L⁻¹), statically at 37°C for 120 h, corresponding to the mid-log phase of growth. After centrifugation at 10,000 × g for 120 min, the biomass from 5 ml of grown culture was collected and total RNA was extracted using MasterPure Complete DNA and RNA purification RNA kits (Epicenter). Total RNA was stored in isopropanol at -20°C before precipitation. RNA was resuspended in 50 µl of RNase-free water. Extracted RNA was treated with a TURBO DNA-free kit (Ambion) to eliminate any residual DNA from the final elution. The quality and concentration of RNA samples were determined using Qubit 3.0 fluorometer (Thermo Fisher Scientific, Italy). The metatranscriptome analysis was performed by FISABIO (Valencia, Spain) using the Illumina® NextSeq Mid Output platform (San Diego, CA, United States) with 2 × 100 bp short insert paired-end libraries (NextSeq Reagent Kit v2.5). FISABIO also performed the quality assessment and the sequence joining (forward R1 and reverse R2). Quality assessment was performed with the PRINSEQ-lite program using the following parameters: min_length: 50 bp; trim_qual_right: 30 bp; trim_qual_type: mean; and trim_qual_window: 20 bp.

Whole-genome sequence analysis and annotation

Metabolic pathways were predicted based on the annotated whole-genome sequences of the strains using the PathoLogic tool of the program Pathway Tools 26.0 (Paley et al., 2021). Information about enzyme encoding genes and the related metabolic pathways was extracted using the SmartTables function of the program. The network of metabolic pathways was visualized using Cytoscape 3.9.1 (<https://cytoscape.org/>). Horizontally acquired genomic islands (GIs) were identified using SeqWord Genomic Island Sniffer (Bezuidt et al., 2009). Genomic comparisons and synteny were visualized using Circos software (Krzywinski et al., 2009) and Blast Dot Plot graph obtained by the RAST server (Aziz et al., 2008; Overbeek et al., 2014; Brettin et al., 2015). Percentages of amino acid identity levels used as the input for Circos visualization and the Blast Dot Plot graph were obtained by the RAST server. Average nucleotide identity (ANI) was calculated for each genome couple compared using the ANI calculator (<http://enve-omics.ce.gatech.edu/ani/index>) as described previously (Goris et al., 2007; Rodriguez and Konstantinidis, 2014). The CRISPRfinder online program (<https://crispr.i2bc.paris-saclay.fr/cgi-bin/crispr/advRunCRISPRFinder.cgi>) was used to detect CRISPRs system in

all sequenced and annotated genomes of nanohaloarchaea. The NetNGlyc-1.0 online program (<https://services.healthtech.dtu.dk/services/NetNGlyc-1.0>) was used to predict N-glycosylation sites in nanohaloarchaeal proteins using artificial neural networks that examine the sequence context of Asn-Xaa-Ser/Thr sequons (Gupta and Brunak, 2002). The whole-genome reference sequence of the strain *Ca. Nanohalococcus occultus* SVXNc [CP107255] was indexed using the *buildindex* function of the Rsubreads package (Bioconductor, www.bioconductor.org). The obtained Illumina RNA reads were mapped to the reference sequence by the function *align* of the Rsubread package. The numbers of reads aligned against each coding sequence of the reference genomes were counted by the function *featureCounts* of the Rsubread package and then normalized by the lengths of the coding sequences to Reads per Kilobase per Million reads (RPKM) values shown in Supplementary Table S1. The prediction of ncRNA was confirmed by modeling the 2D structures of the ncRNA molecules using the RNAfold WebServer (<http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi>).

Genome methylation profiling

For the prediction of genome methylation profiles, the PacBio reads were mapped against the whole-genome sequence of the strain *Ca. Nanohalococcus occultus* SVXNc using pbmm2 aligner of the software package SMRT Link v10.1.0.119588 (<https://www.pacb.com/support/software-downloads/>) followed by the prediction of methylated nucleotides and methylation motifs using ipdSummary and motifMaker tools of the SMRT Link package. The following command lines were used to run the pipelines on the server:

```
pbmm2 index path_to_reference_FASTA_file path_to_reference_index_MMI_file.
pbmm2 align -sort path_to_reference_index_MMI_file path_to_read_database_XML_file path_to_output_BAM_file.
ipdSummary path_to_output_BAM_file -reference path_to_reference_FASTA_file -identify m6A, m4C -gff path_to_output_GFF_file.
motifMaker find -fasta path_to_reference_FASTA_file -gff path_to_output_GFF_file -minScore 30 -output path_to_motif_output_CSV_file.
```

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary material.

Author contributions

MY, FC, VL, ES, and LM were responsible for microbiology work. OR, EM, and MY designed the research. OR, EM, FS, MK, GL, and MY carried out bioinformatics analyses and data

interpretation. MR performed FESEM analysis. OR, EM, FS, VL, FC, and GL were responsible for genome sequencing, assembling, and curation. MY, OR, and MK wrote the manuscript, with the contribution of all authors. All authors contributed to the article and approved the submitted version.

Funding

This study was supported by a grant from the FUTURENZYMES Project (Contract 101000327), funded by the European Union's Horizon 2020 Research Program. MK was supported by Agence Nationale de la Recherche grant ANR-20-CE20-0009. This work has been additionally supported by the project "Emergence of virulence and antibiotic-resistance vectors in coastal and deep sea marine environments and analysis of the mechanisms and conditions underlying their spread and evolution" funded by the MIUR Progetti di Ricerca di Rilevante Interesse Nazionale (PRIN) Bando 2017 - grant 201728ZA49_001 and by the project "Innovazione tecnologica per la tutela e valorizzazione degli ecosistemi marini (ITEM)" funded by MIUR CTN02_00059_9948371.

References

- Albers, S. V., and Jarrell, K. F. (2018). The archaeellum: an update on the unique archaeal motility structure. *Trends Microbiol.* 26, 351–362. doi: 10.1016/j.tim.2018.01.004
- Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., et al. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucl. Acids Res.* 25, 3389–3402. doi: 10.1093/nar/25.17.3389
- Arias-Carrasco, R., Vázquez-Morán, Y., Nakaya, H. I., and Maracaja-Coutinho, V. (2018). StructRNAfinder: an automated pipeline and web server for RNA families prediction. *BMC Bioinform.* 19, 55. doi: 10.1186/s12859-018-2052-2
- Atanasova, N. S., Oksanen, H. M., and Bamford, D. H. (2015). Haloviruses of archaea, bacteria, and eukaryotes. *Curr. Opin. Microbiol.* 25, 40–48. doi: 10.1016/j.mib.2015.04.001
- Atomi, H., and Reeve, J. (2019). Microbe Profile: *Thermococcus kodakarensis*: the model hyperthermophilic archaeon. *Microbiology (Reading)* 165, 1166–1168. doi: 10.1099/mic.0.000839
- Aziz, R. K., Bartels, D., Best, A. A., DeJongh, M., Disz, T., Edwards, R. A., et al. (2008). The RAST Server: rapid annotations using subsystems technology. *BMC Genom.* 9, 1–15. doi: 10.1186/1471-2164-9-75
- Baumann, A., Lange, C., and Soppa, J. (2007). Transcriptome changes and cAMP oscillations in an archaeal cell cycle. *BMC Cell. Biol.* 8, 21. doi: 10.1186/1471-2121-8-21
- Bezuidt, O., Lima-Mendez, G., and Reva, O. N. (2009). SeqWord Gene Island Sniffer: a program to study the lateral genetic exchange among bacteria. *Int. J. Comp. Inf. Eng.* 3, 2399–2404.
- Bolhuis, H., Palm, P., Wende, A., Falb, M., Rampp, M., Rodriguez-Valera, F., et al. (2006). The genome of the square archaeon *Haloquadratum walsbyi*: life at the limits of water activity. *BMC Genom.* 7, 1–12. doi: 10.1186/1471-2164-7-169
- Braun, F., Thomalla, L., van der Does, C., Quax, T. E. F., Allers, T., Kaever, V., et al. (2019). Cyclic nucleotides in archaea: Cyclic di-AMP in the archaeon *Haloferax volcanii* and its putative role. *Microbiologyopen* 8, e00829. doi: 10.1002/mb.03.829
- Brettin, T., Davis, J. J., Disz, T., Edwards, R. A., Gerdes, S., Olsen, G. J., et al. (2015). RASTtk: a modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. *Sci. Rep.* 5, 8365. doi: 10.1038/srep08365
- Buessecker, S., Palmer, M., Lai, D., Dimapilis, J., Mayali, X., Mosier, D., et al. (2022). An essential role for tungsten in the ecology and evolution of a previously uncultivated lineage of anaerobic, thermophilic Archaea. *Nat. Commun.* 13, 3773. doi: 10.1038/s41467-022-31452-8
- Buth, S. A., Shneider, M. M., Scholl, D., and Leiman, P. G. (2018). Structure and analysis of R1 and R2 pyocin receptor-binding fibers. *Viruses* 10, 427. doi: 10.3390/v10080427
- Castelle, C. J., Brown, C. T., Anantharaman, K., Probst, A. J., Huang, R. H., and Banfield, J. F. (2018). Biosynthetic capacity, metabolic variety and unusual biology in the CPR and DPANN radiations. *Nat. Rev. Microbiol.* 16, 629–645. doi: 10.1038/s41579-018-0076-2
- Castelle, C. J., Wrigton, K. C., Thomas, B. C., Hug, L. A., Brown, C. T., Wilkins, M. J., et al. (2015). Genomic expansion of domain archaea highlights roles for organisms from new phyla in anaerobic carbon cycling. *Curr. Biol.* 25, 690–701. doi: 10.1016/j.cub.2015.01.014
- Comolli, L. R., Baker, B. J., Downing, K. H., Siegerist, C. E., and Banfield, J. F. (2009). Three-dimensional analysis of the structure and ecology of a novel, ultra-small archaeon. *ISME J.* 3, 159–167. doi: 10.1038/ismej.2008.99
- Crits-Christoph, A., Gelsinger, D. R., Ma, B., Wierzbos, J., Ravel, J., Davila, A., et al. (2016). Functional interactions of archaea, bacteria and viruses in a hypersaline endolithic community. *Environ. Microbiol.* 18, 2064–2077. doi: 10.1111/1462-2920.13259
- Dombrowski, N., Lee, J.-H., Williams, T. A., Offe, P., and Spang, A. (2019). Genomic diversity, lifestyles and evolutionary origins of DPANN archaea. *FEMS Microbiol. Lett.* 366, fnz008. doi: 10.1093/femsle/fnz008
- Dombrowski, N., Williams, T. A., Sun, J., Woodcroft, B. J., Lee, J. H., Minh, B. Q., et al. (2020). Undinarchaeota illuminate DPANN phylogeny and the impact of gene transfer on archaeal evolution. *Nat. Commun.* 11, 3939. doi: 10.1038/s41467-020-17408-w
- Eichler, J., and Maupin-Furlow, J. (2013). Post-translation modification in Archaea: lessons from *Haloferax volcanii* and other haloarchaea. *FEMS Microbiol. Rev.* 37, 583–606. doi: 10.1111/1574-6976.12012
- Falb, M., Müller, K., Königsmaier, L., Oberwinkler, T., Horn, P., von Gronau, S., et al. (2008). Metabolism of halophilic archaea. *Extremophiles* 12, 177–196. doi: 10.1007/s00792-008-0138-x
- Feng, Y., Neri, U., Gosselin, S., Louyakis, A. S., Papke, R. T., Gophna, U., et al. (2021). The evolutionary origins of extreme halophilic archaeal lineages. *Genome Biol. Evol.* 13, evab166. doi: 10.1093/gbe/evab166
- Finstad, K. M., Probst, A. J., Thomas, B. C., Andersen, G. L., Demergasso, C., Echeverría, A., et al. (2017). Microbial community structure and the persistence of cyanobacterial populations in salt crusts of the hyperarid Atacama Desert from genome-resolved metagenomics. *Front. Microbiol.* 8, 1435. doi: 10.3389/fmicb.2017.01435

Conflict of interest

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.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2023.1182464/full#supplementary-material>

- Foster, T. J. (2019). The MSCRAMM family of cell-wall-anchored surface proteins of gram-positive cocci. *Trends Microbiol.* 27, 927–941. doi: 10.1016/j.tim.2019.06.007
- Gelsinger, D. R., and DiRuggiero, J. (2018). The non-coding regulatory RNA revolution in archaea. *Genes (Basel)* 9, 141. doi: 10.3390/genes9030141
- Ghai, R., Pašić, L., Fernández, A. B., Martín-Cuadrado, A. B., Mizuno, C. M., McMahon, K. D., et al. (2011). New abundant microbial groups in aquatic hypersaline environments. *Sci. Rep.* 1, 135. doi: 10.1038/srep00135
- Gilmore, S. F., Yao, A. I., Tietel, Z., Kind, T., Facciotti, M. T., and Parikh, A. N. (2013). Role of squalene in the organization of monolayers derived from lipid extracts of *Halobacterium salinarum*. *Langmuir* 29, 7922–7930. doi: 10.1021/la401412t
- Golyshina, O. V., Toshchakov, S. V., Makarova, K. S., Gavrillov, S. N., Korzhnikov, A. A., La Cono, V., et al. (2017). 'ARMAN' archaea depend on association with euryarchaeal host in culture and *in situ*. *Nat. Commun.* 8, 60. doi: 10.1038/s41467-017-00104-7
- Gonzalez-Ordenes, F., Cea, P. A., Fuentes-Ugarte, N., Muñoz, S. M., Zamora, R. A., Leonardo, D., et al. (2018). ADP-dependent kinases from the archaeal order Methanosarcinales adapt to salt by a non-canonical evolutionarily conserved strategy. *Front. Microbiol.* 9, 1305. doi: 10.3389/fmicb.2018.01305
- Goris, J., Konstantinidis, K. T., Klappenbach, J. A., Coenye, T., Vandamme, P., and Tiedje, J. M. (2007). DNA–DNA hybridization values and their relationship to whole-genome sequence similarities. *Int. J. Syst. Evol. Microbiol.* 57, 81–91. doi: 10.1099/ijs.0.64483-0
- Gupta, R., and Brunak, S. (2002). Prediction of glycosylation across the human proteome and the correlation to protein function. *Pac. Symp. Biocomput.* 2002, 310–322. doi: 10.1142/9789812799623_0029
- Hamm, J. N., Erdmann, S., Eloe-Fadrosh, E. A., Angeloni, A., Zhong, L., Brownlee, C., et al. (2019). Unexpected host dependency of Antarctic Nanohaloarchaeota. *Proc. Natl. Acad. Sci. U S A* 116, 14661–14670. doi: 10.1073/pnas.1905179116
- Hamm, J. N., Liao, Y., von Kuegelgen, A., Dombrowski, N., Landers, E., Brownlee, C., et al. (2023). The intracellular lifestyle of an archaeal symbiont. *bioRxiv* 2023–02. doi: 10.1101/2023.02.24.529834
- Huber, H., Hohn, M. J., Rachel, R., Fuchs, T., Wimmer, V. C., and Stetter, K. O. (2002). A new phylum of Archaea represented by a nanosized hyperthermophilic symbiont. *Nature* 417, 63–67. doi: 10.1038/417063a
- Jahn, U., Gallenberger, M., Paper, W., Junglas, B., Eisenreich, W., Stetter, K. O., et al. (2008). *Nanoarchaeum equitans* and *Ignicoccus hospitalis*: new insights into a unique, intimate association of two archaea. *J. Bacteriol.* 190, 1743–1750. doi: 10.1128/JB.01731-07
- Jarett, J. K., Nayfach, S., Podar, M., Inskeep, W., Ivanova, N. N., Munson-McGee, J., et al. (2018). Single-cell genomics of co-sorted *Nanoarchaeota* suggests novel putative host associations and diversification of proteins involved in symbiosis. *Microbiome* 6, 1–14. doi: 10.1186/s40168-018-0539-8
- Kanehisa, M., Sato, Y., and Morishima, K. (2016). BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *J. Mol. Biol.* 428, 726–731. doi: 10.1016/j.jmb.2015.11.006
- Kashima, Y., and Ishikawa, K. (2003). Alkyl hydroperoxide reductase dependent on thioredoxin-like protein from *Pyrococcus horikoshii*. *J. Biochem.* 134, 25–29. doi: 10.1093/jb/mvg109
- Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., et al. (2009). Circos: an information aesthetic for comparative genomics. *Genome Res.* 19, 1639–1645. doi: 10.1101/gr.092759.109
- La Cono, V., Bortoluzzi, G., Messina, E., La Spada, G., Smedile, F., Giuliano, L., et al. (2019). The discovery of Lake Hephæstus, the youngest athalassohaline deep-sea formation on Earth. *Sci. Rep.* 9, 1679. doi: 10.1038/s41598-018-38444-z
- La Cono, V., Messina, E., Reva, O., Smedile, F., La Spada, G., Crisafi, F., et al. (2023). Nanohaloarchaea as beneficiaries of xylan degradation by haloarchaea. *Microb. Biotechnol.* In Press doi: 10.1111/1751-7915.14272
- La Cono, V., Messina, E., Rohde, M., Arcadi, E., Ciordia, S., Crisafi, F., et al. (2020). Symbiosis between nanohaloarchaeon and haloarchaeon is based on utilization of different polysaccharides. *Proc. Natl. Acad. Sci. U S A* 117, 20223–20234. doi: 10.1073/pnas.2007232117
- Leoni, C., Volpicella, M., Fosso, B., Manzari, C., Piancone, E., Dileo, M. C. G., et al. (2020). A differential metabarcoding approach to describe taxonomy profiles of bacteria and archaea in the saltern of Margherita di Savoia (Italy). *Microorganisms* 8, 936. doi: 10.3390/microorganisms8060936
- Makarova, K. S., Wolf, Y. I., Iranzo, J., Shmakov, S. A., Alkhnbashi, O. S., Brouns, S. J., et al. (2020). Evolutionary classification of CRISPR–Cas systems: a burst of class 2 and derived variants. *Nat. Rev. Microbiol.* 18, 67–83. doi: 10.1038/s41579-019-0299-x
- Moreira, D., Zivanovic, Y., López-Archilla, A. I., Iniesto, M., and López-García, P. (2021). Reductive evolution and unique predatory mode in the CPR bacterium *Vampirococcus lugosii*. *Nat. Commun.* 12, 2454. doi: 10.1038/s41467-021-22762-4
- Narasimangarao, P., Podell, S., Ugalde, J. A., Brochier-Armanet, C., Emerson, J. B., Brocks, J. J., et al. (2012). *De novo* metagenomic assembly reveals abundant novel major lineage of Archaea in hypersaline microbial communities. *ISME J.* 6, 81–93. doi: 10.1038/ismej.2011.78
- Nesta, B., Spraggon, G., Alteri, C., Gomes Moriel, D., Rosini, R., Veggi, D., et al. (2012). FdeC, a novel broadly conserved *Escherichia coli* adhesin eliciting protection against urinary tract infections. *MBi* 3, e00010–12. doi: 10.1128/mBio.00010-12
- Oren, A., and DiRuggiero, J. (2015). "Candidatus Nanopettraeus," in *Bergey's Manual of Systematics of Archaea and Bacteria* 1–4. doi: 10.1002/9781118960608.fbm00148
- Overbeek, R., Olson, R., Pusch, G. D., Olsen, G. J., Davis, J. J., Disz, T., et al. (2014). The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Res.* 42, D206–D214. doi: 10.1093/nar/gkt1226
- Paley, S., Billington, R., Herson, J., Krummenacker, M., and Karp, P. D. (2021). Pathway tools visualization of organism-scale metabolic networks. *Metabolites* 11, 64. doi: 10.3390/metabo11020064
- Rawat, M., and Maupin-Furlow, J. A. (2020). Redox and thiols in archaea. *Antioxidants* 9, 381. doi: 10.3390/antiox9050381
- Redero, M., Aznar, J., and Prieto, A. I. (2020). Antibacterial efficacy of R-type pyocins against *Pseudomonas aeruginosa* on biofilms and in a murine model of acute lung infection. *J. Antimicrob. Chemother.* 75, 2188–2196. doi: 10.1093/jac/dkaa121
- Rodriguez, R. L. M., and Konstantinidis, K. T. (2014). Bypassing cultivation to identify bacterial species. *Microbe* 9, 111–118. doi: 10.1128/microbe.9.111.1
- Samuel, B. S., Hansen, E. E., Manchester, J. K., Coutinho, P. M., Henrissat, B., Fulton, R., et al. (2007). Genomic and metabolic adaptations of *Methanobrevibacter smithii* to the human gut. *Proc. Natl. Acad. Sci. U S A* 104, 10643–10648. doi: 10.1073/pnas.0704189104
- Selivanova, E. A., Poshvina, D. V., Khlopko, Y. A., Gogoleva, N. E., and Plotnikov, A. O. (2018). Diversity of prokaryotes in planktonic communities of saline Sol-Ilets lakes (Orenburg Oblast, Russia). *Microbiology* 87, 569–582. doi: 10.1134/S0026261718040161
- Sorokin, D. Y., Messina, E., Smedile, F., La Cono, V., Hallsworth, J. E., and Yakimov, M. M. (2021). Carbohydrate-dependent sulfur respiration in halo(alkali)philic archaea. *Environ. Microbiol.* 23, 3789–3808. doi: 10.1111/1462-2920.15421
- Sorokin, D. Y., Makarova, K. S., Abbas, B., Ferrer, M., Golyshin, P. N., Galinski, E. A., et al. (2017). Discovery of extremely halophilic, methyl-reducing euryarchaea provides insights into the evolutionary origin of methanogenesis. *Nat. Microbiol.* 2, 1–11. doi: 10.1038/nmicrobiol.2017.81
- Spring, S., Scheuner, C., Lapidus, A., Lucas, S., Glavina Del Rio, T., Tice, H., et al. (2010). The genome sequence of *Methanohalophilus mahii* SLPT reveals differences in the energy metabolism among members of the *Methanosarcinaceae* inhabiting freshwater and saline environments. *Archaea* 2010, 690737. doi: 10.1155/2010/690737
- Su, X., Lin, Z., and Lin, H. (2013). The biosynthesis and biological function of diphthamide. *Crit. Rev. Biochem. Mol. Biol.* 48, 515–521. doi: 10.3109/10409238.2013.831023
- Truitt, C., and Deole, R. (2021). "Viruses of Extremely Halophilic Prokaryotes," in *Bacteriophages in Therapeutics*, ed. S. B. Bhardwaj (London, United Kingdom: IntechOpen). doi: 10.5772/intechopen.96720
- Vavourakis, C. D., Ghai, R., Rodriguez-Valera, F., Sorokin, D. Y., Tringe, S. G., Hugenholtz, P., et al. (2016). Metagenomic insights into the uncultured diversity and physiology of microbes in four hypersaline soda lake brines. *Front. Microbiol.* 7, 211. doi: 10.3389/fmicb.2016.00211
- Xie, Y. G., Luo, Z. H., Fang, B. Z., Jiao, J. Y., Xie, Q. J., Cao, X. R., et al. (2022). Functional differentiation determines the molecular basis of the symbiotic lifestyle of *Ca. Nanohaloarchaeota*. *Microbiome* 10, 1–13. doi: 10.1186/s40168-022-01376-y
- Xu, C., and Min, J. (2011). Structure and function of WD40 domain proteins. *Protein Cell.* 2, 202–214. doi: 10.1007/s13238-011-1018-1
- Yakimov, M. M., Merkel, A. Y., Gaisin, V. A., Pilhofer, M., Messina, E., Hallsworth, J. E., et al. (2022). Cultivation of a vampire: 'Candidatus Absconditococcus praedator'. *Environ. Microbiol.* 24, 30–49. doi: 10.1111/1462-2920.15823
- Zhang, H., and Forman, H. J. (2012). Glutathione synthesis and its role in redox signaling. *Semin. Cell. Dev. Biol.* 23, 722–772. doi: 10.1016/j.semcdb.2012.03.017
- Zhao, D., Zhang, S., Zhou, H., Xue, Q., Sun, W., et al. (2022). Comparative genomic insights into the evolution of Halobacteria-associated 'Candidatus Nanohaloarchaeota'. *Msystems* 7, e00669–e00622. doi: 10.1128/msystems.00669-22



OPEN ACCESS

EDITED BY

Solenne Ithurbide,
University of Freiburg, Germany

REVIEWED BY

Alejandra Recalde,
University of Freiburg, Germany
Elżbieta Kraszewska,
Polish Academy of Sciences, Poland

*CORRESPONDENCE

Lennart Randau
✉ lennart.randau@staff.uni-marburg.de

RECEIVED 31 March 2023

ACCEPTED 01 June 2023

PUBLISHED 15 June 2023

CITATION

Breuer R, Gomes-Filho JV, Yuan J and
Randau L (2023) Transcriptome profiling of
Nudix hydrolase gene deletions in the
thermoacidophilic archaeon *Sulfolobus*
acidocaldarius.
Front. Microbiol. 14:1197877.
doi: 10.3389/fmicb.2023.1197877

COPYRIGHT

© 2023 Breuer, Gomes-Filho, Yuan and
Randau. 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.

Transcriptome profiling of Nudix hydrolase gene deletions in the thermoacidophilic archaeon *Sulfolobus acidocaldarius*

Ruth Breuer¹, José Vicente Gomes-Filho¹, Jing Yuan^{2,3} and
Lennart Randau^{1,3*}

¹Prokaryotic RNA Biology, Department of Biology, Philipps-Universität Marburg, Marburg, Germany,

²Max Planck Institute for Terrestrial Microbiology, Marburg, Germany, ³SYNMIKRO, Center for Synthetic Microbiology, Marburg, Germany

Nudix hydrolases comprise a large and ubiquitous protein superfamily that catalyzes the hydrolysis of a nucleoside diphosphate linked to another moiety X (Nudix). *Sulfolobus acidocaldarius* possesses four Nudix domain-containing proteins (SACI_RS00730/Saci_0153, SACI_RS02625/Saci_0550, SACI_RS00060/Saci_0013/Saci_NudT5, and SACI_RS00575/Saci_0121). Deletion strains were generated for the four individual Nudix genes and for both Nudix genes annotated to encode ADP-ribose pyrophosphatases (SACI_RS00730, SACI_RS00060) and did not reveal a distinct phenotype compared to the wild-type strain under standard growth conditions, nutrient stress or heat stress conditions. We employed RNA-seq to establish the transcriptome profiles of the Nudix deletion strains, revealing a large number of differentially regulated genes, most notably in the Δ SACI_RS00730/SACI_RS00060 double knock-out strain and the Δ SACI_RS00575 single deletion strain. The absence of Nudix hydrolases is suggested to impact transcription via differentially regulated transcriptional regulators. We observed downregulation of the lysine biosynthesis and the archaeum formation iModulons in stationary phase cells, as well as upregulation of two genes involved in the *de novo* NAD⁺ biosynthesis pathway. Furthermore, the deletion strains exhibited upregulation of two thermosome subunits (α , β) and the toxin-antitoxin system VapBC, which are implicated in the archaeal heat shock response. These results uncover a defined set of pathways that involve archaeal Nudix protein activities and assist in their functional characterization.

KEYWORDS

Nudix hydrolase, transcriptomics, thermophile, RNA processing, gene regulation

1. Introduction

Nudix hydrolases constitute an evolutionary conserved protein superfamily of functionally versatile proteins present in all three domains of life. They catalyze the hydrolysis of a wide range of small nucleotide substrates composed of a nucleoside diphosphate linked to another moiety X (Nudix) (Figure 1) identifiable by the conserved Nudix motif with the consensus sequence GX₅EX₅U/AXREX₂EEXGU (U for a hydrophobic residue, X for any residue) (Bessman et al., 1996; McLennan, 2006). Initially characterized as “housecleaning enzymes” which cleanse the cell of potentially toxic metabolites (Bessman et al., 1996), it has since been revealed that their biological roles are more diverse than previously thought. In *Escherichia coli*, the Nudix

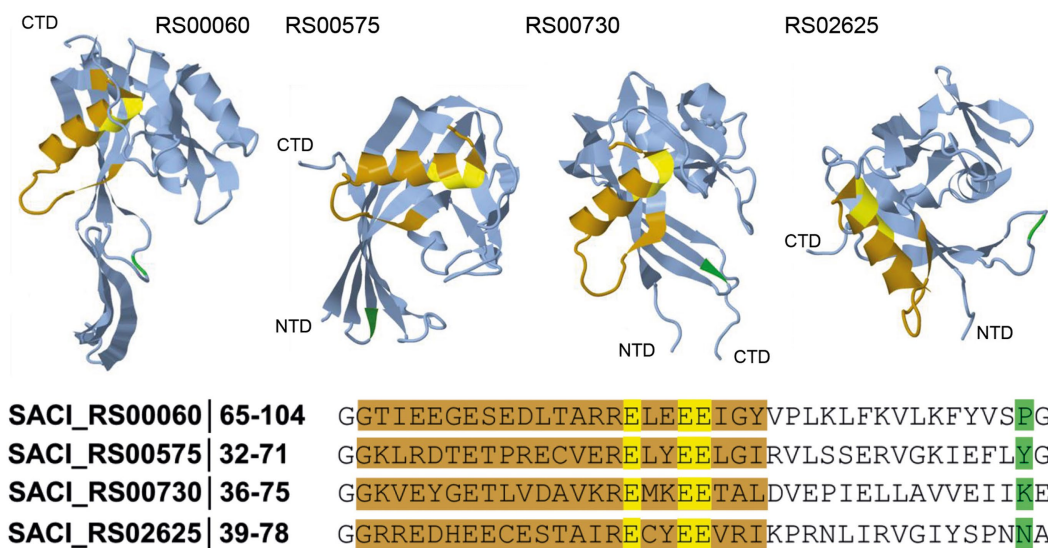


FIGURE 1

Schematic representation of the substrate backbone for Nudix hydrolases. In *S. acidocaldarius*, four Nudix hydrolases were identified, and their tridimensional structure was predicted using AlphaFold (AlphaFoldDB accession codes: Q4JCN6, Q4JCD4, Q4JCA5, Q4JB83, Jumper et al., 2021). The Nudix motif (orange, consensus sequence: GX₂EX₂REUXEEXGU) and the corresponding functional amino acids (yellow) are shown. The amino acid at position 16 (green) after the Nudix motif can be used as an initial guide for substrate identification of newly discovered Nudix hydrolases.

hydrolase NudB hydrolyses 8-oxo-dADP, 8-oxo-dGDP and 2-oxo-dADP and was thus proposed to possess antimutator activity (Hori et al., 2005). Subsequent studies revealed dihydroneopterin triphosphate (DHNTTP), which is structurally similar to GTP, to be the preferred substrate and deletion of *nudB* led to impaired folate synthesis *in vivo*, where DHNTTP plays an integral intermediary role (Gabelli et al., 2007). Recently, Nudix hydrolases have gained attention due to their ability to remove non-canonical metabolite caps from RNA molecules. Two prominent examples are the bacterial Nudix hydrolases NudC and RppH. Initially described as a NAD/H pyrophosphohydrolase, NAD-RNA decapping by NudC results in 5'-monophosphorylated RNA and nicotinamide mononucleotide (NMN) (Frick and Bessman, 1995; Cahová et al., 2015). The RNA pyrophosphohydrolase RppH converts 5'-triphosphorylated RNA into 5'-monophosphorylated RNA triggering RNA degradation by RNase E or RNase J (Deana et al., 2008; Richards et al., 2011) and can also act as NAD-decapping enzyme (Frindert et al., 2018; Grudzien-Nogalska et al., 2019). Furthermore, *E. coli* RppH removes non-methylated 5'-Np_nN-caps from RNA (Hudeček et al., 2020). In eukaryotes, various Nudix hydrolases (termed "NudT") capable of removing different caps from RNA *in vitro* have been identified (Abdelraheim et al., 2003; Song et al., 2010, 2013; Grudzien-Nogalska et al., 2019). NAD-capped RNA has been identified in bacteria and eukaryotes and recently also in the archaeal domain in *Sulfolobus acidocaldarius* and *Haloferax volcanii* (Chen et al., 2009; Cahová et al., 2015; Jiao et al., 2017; Walters et al., 2017; Ruiz-Larrabeiti et al., 2021; Gomes-Filho et al., 2022). Furthermore, methylated and non-methylated dinucleoside polyphosphates (Np_nNs) were identified at the 5' ends of *E. coli* RNA (Hudeček et al., 2020) and most recently, ADPR-capped RNA was identified in human cells (Weixler et al., 2022).

Following the discovery of NAD-capped RNA in *S. acidocaldarius*, we aimed to identify the respective decapping enzyme(s) among the Nudix hydrolases present in this organism and identified four Nudix

domain-containing proteins encoded in the genome. *S. acidocaldarius* is a thermophilic crenarchaeon with an optimal growth temperature between 75 to 80°C and the need for acidic growth medium with an optimal pH of 2–3. It might likely possess more than one decapping enzyme, e.g., mammalian cells possess more than 20 decapping enzymes, each regulating specific subsets of capped RNA (McLennan, 2006; Song et al., 2010; Li et al., 2011). Additionally, it has been reported that the number of Nudix family representatives in bacteria, eukaryotic microorganisms and fungi is linearly correlated with genome size (McLennan, 2006). Currently, *E. coli* presents 13 genes encoding Nudix family hydrolases, *Saccharomyces cerevisiae* 4, *Caenorhabditis elegans* 14, *Drosophila melanogaster* 20 and *Arabidopsis thaliana* 25–28 members (McLennan, 2006; Bessman, 2019). In the archaeal domain, *Haloferax volcanii* exhibits 12 genes encoding Nudix domain-containing proteins, while *Methanococcus jannaschii*, *M. maripaludis* and *Thermococcus kodakarensis* each possess a single Nudix domain protein. Among the other Sulfolobales/Saccharolobales, *Sa. solfataricus* possesses 3, *S. tokodaii* 4 and *S. islandicus* 4 genes encoding Nudix domain-containing proteins. In the organisms presenting only one Nudix protein, it is annotated as a putative ADPR hydrolase (Alm et al., 2005). The Nudix family is known for displaying a large substrate range, especially *in vitro*, encompassing canonical and oxidized nucleotides, nucleotide sugars, dinucleotide coenzymes, diadenosine polyphosphates and capped RNAs, as well as non-nucleotide substrates such as inositol pyrophosphates (McLennan, 2013; Srouji et al., 2017). While the Nudix protein Saci_NudT5 demonstrated ADPR-RNA decapping ability *in vitro*, the other proteins' activities remain elusive (Gomes-Filho et al., 2022). To this end, this study presents the impact of Nudix hydrolase absence on the transcriptome of *S. acidocaldarius*. None of the gene deletions elicited an obvious phenotype, but all strains exhibit a considerable number of differentially regulated transcriptional regulators. Overall, the transcriptome of the Nudix deletion strains resembles that of samples

taken under heat stress and nutrient limitation conditions (Bischof et al., 2019), hence aiding our understanding of the gene network regulating stress response in *S. acidocaldarius*.

2. Results and discussion

2.1. The *Sulfolobus acidocaldarius* genome encodes four Nudix family hydrolases

BLAST analyses and multiple sequence alignments revealed four genes encoding Nudix domain-containing proteins in the genome of *Sulfolobus acidocaldarius*: SACI_RS00730, SACI_RS00060, SACI_RS02625 and SACI_RS00575 (Figure 1). All proteins possess the conserved glutamic acid residues in the Nudix motif crucial to Nudix activity (Cahová et al., 2015; Höfer et al., 2016; Frindert et al., 2018; Grudzien-Nogalska et al., 2019). The residue at position 16 following the G of the Nudix motif correlates with possible substrates for the respective Nudix protein and can serve to identify and distinguish different subsets of Nudix hydrolases (Dunn et al., 1999). In SACI_RS00060, a proline at this position suggests ADP-ribose (ADPR) hydrolysis activity. Indeed, this protein was recently shown to decap ADPR-RNA *in vitro* and hence renamed Saci_NudT5 to match the nomenclature of its human homolog (Gomes-Filho et al., 2022). In SACI_RS00575, a tyrosine at position 16 hints at specificity for dinucleoside polyphosphate substrates, however its substrate specificity remains unclear. Due to high sequence similarity to Saci_NudT5, SACI_RS00730 might represent another ADPR hydrolase. SACI_RS02625 is conserved across many archaeal species, but no specific activity has been predicted thus far. In bacteria, a family of Nudix-related transcriptional regulators (NrtR) regulates NAD⁺ metabolism and interacts with ADPR as their effector molecule. These proteins are characterized by an N-terminal Nudix-like domain homologous to ADPR pyrophosphatases and a C-terminal helix-turn-helix (HTH)-like DNA-binding domain (Rodionov et al., 2008). On this note, an HTH-domain was not identified in any of the Nudix family hydrolases from *S. acidocaldarius*.

Protein structures of the four Nudix hydrolases from *S. acidocaldarius* were modeled by AlphaFold sourced from the UniProt database (Jumper et al., 2021; Varadi et al., 2022; Figure 1). All protein structure predictions reveal similar folding in which the Nudix motif is present as an α -helix located on the outer part of the structure, close to the substrate pocket. In the active site, three conserved glutamic acid residues act as ligands to magnesium ions and are directed toward the inside of the pocket. The indicator residue at position 16 after the G of the Nudix motif is located on the opposite end of the pocket. Saci_NudT5, the largest of the four proteins, additionally exhibits an extended structure which is not present in the other three proteins (Figure 1).

2.2. The Nudix hydrolase genes are not essential in *Sulfolobus acidocaldarius*

The genes encoding all four Nudix hydrolases in *S. acidocaldarius*, SACI_RS00730, SACI_RS00060 (encodes Saci_NudT5), SACI_RS02625 and SACI_RS00575 were individually targeted for deletion using the double crossover method based on plasmid pSVA431

developed by Wagner et al., 2012. With this approach, all four genes were removed from the genome without interrupting their partially overlapping neighboring genes, indicating that none of the Nudix hydrolases is essential for *S. acidocaldarius*. The absence of Nudix gene transcripts was confirmed by RNA-sequencing (Supplementary Figure S1). To account for a possible redundancy between the two most similar genes (Supplementary Figure S2), a double deletion mutant of SACI_RS00730 and SACI_RS00060 was generated. In a phenotypical survey, the deletion strains were grown in parallel to the wild-type strain under standard, nitrogen stress and carbon stress conditions. All strains showed similar growth behavior and no significant deviation from wild-type growth was detected (Figure 2A). Similarly, no significant deviation was observed between the wild-type and deletion strains after submission to 87°C heat shock and cell growth was significantly reduced in all strains except for Δ SACI_RS02625 (Figure 2B). Here, heat-shocked cells showed survival rates between 33 and 65% compared to their control samples, while for Δ SACI_RS02625 80% of cells were still viable after heat shock exposure (Figure 2C). In conclusion, the Nudix hydrolase genes are not essential in *S. acidocaldarius* and their deletion did not elicit a distinct phenotype compared to the wild-type strain under conditions tested.

2.3. Each deletion strain exhibits a distinct transcriptome profile

The Nudix deletion strains were grown in parallel to the wild-type strain and their transcriptome profiles were determined for the mid-logarithmic and the early stationary phase using Illumina RNA-seq. Differential gene expression was analyzed using DESeq2 to enable comparison of the strains' transcriptome profiles (Supplementary Table S1). RT-qPCR was used to independently assess the quality of RNA-seq data and verified gene expression profiles of selected genes (Supplementary Figure S3). Overall, the datasets exhibit more differentially regulated genes in the early stationary compared to the mid-log phase. Both heatmaps show a similar clustering of the strains according to the similarity of their transcriptome profiles with placement of Δ SACI_RS00730 and Δ SACI_RS00575 on opposite ends of the neighborhood tree, clearly establishing these two transcriptomic profiles as most dissimilar (Figure 3). The highest number of differentially regulated genes is present in the Δ SACI_RS00730/SACI_RS00060 double deletion strain and the Δ SACI_RS00575 strain, hence establishing them to be most impactful on the transcriptome. The smallest transcriptomic impact is caused by the individual deletions of SACI_RS00730 and SACI_RS00060 (Figure 3) which were initially believed to possess redundant activities due to high sequence similarity. However, the presence of oppositely regulated gene clusters, as well as distinct activities in the *in vitro* decapping assays refutes this assumption (Gomes-Filho et al., 2022). In summary, clusters of similarly affected genes are rarely shared between more than two strains, giving each strain a unique transcriptomic profile and suggesting unique roles for the enzymes in question. This agrees with the occurrence of only four Nudix hydrolase genes in the genome, as the likelihood of redundancies would be expected to increase with the number of Nudix genes.

To gain insight into the impact of the individual Nudix deletions, the differentially regulated genes of each dataset were assembled into

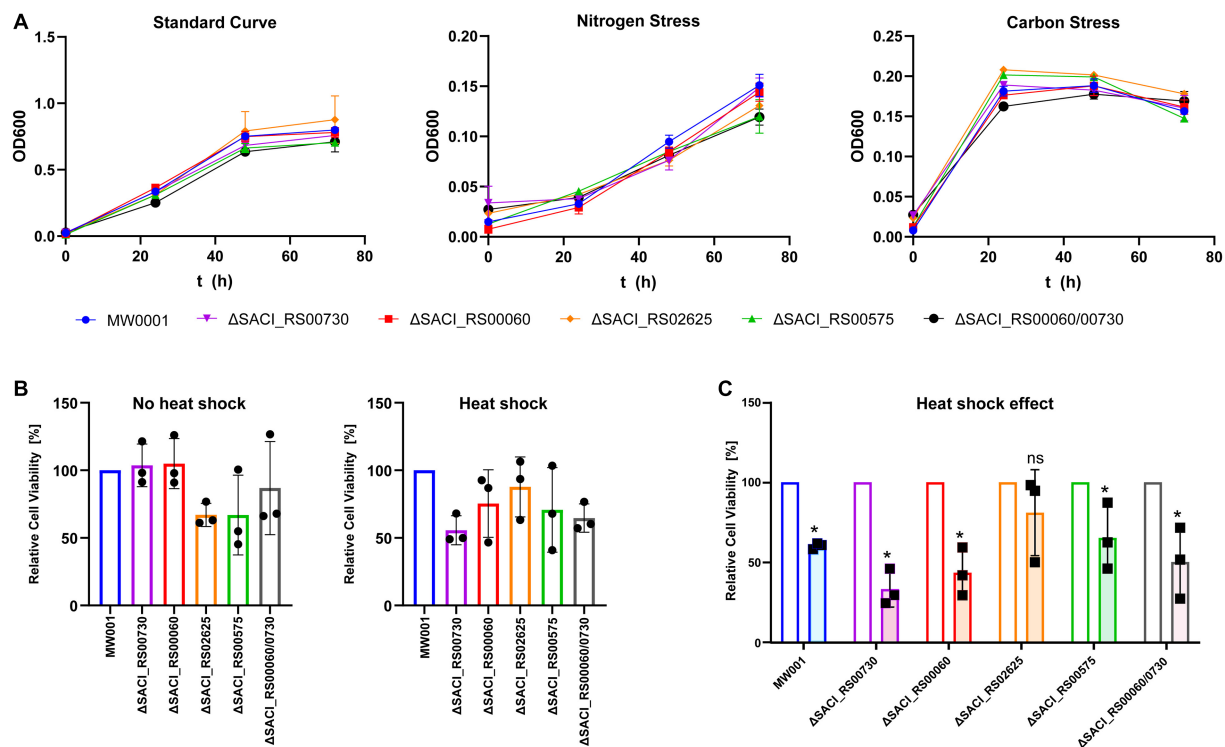


FIGURE 2

Phenotypic characterization of *S. acidocaldarius* Nudix gene deletion strains under standard and stress conditions. (A) Growth curves performed under standard conditions, nitrogen, and carbon stress. (B) Cell viability of each knockout strain relative to the wild-type with and without heat stress. (C) The relative cell viability after heat treatment was normalized by cell viability of non-heat-treated control samples. Each experiment was performed with triplicates, and error bars depict their standard deviation. Asterisks (*) denote Student's *t*-test value <0.05 .

iModulons according to the iModulonDB database (Chauhan et al., 2021; Rychel et al., 2021; Figure 4). An iModulon (independently modulated signal) comprises a group of genes similarly expressed under different (growth) conditions and is hence proposed to be the data-driven analog of a regulon without spatial restriction. These iModulons were identified by observing patterns in transcriptome datasets using unsupervised machine learning and independent component analysis (ICA) (Rychel et al., 2021). The genes encoding the four Nudix family hydrolases are not assigned to any iModulons. Notably, genes can be assigned to more than one iModulon and each iModulon may encompass more genes than currently displayed. Presumably, the iModulons are modulated by a common regulator or related ones which must not necessarily be part of its iModulon and for many iModulons of *S. acidocaldarius* a common regulator has not yet been identified (Chauhan et al., 2021). The DARC (Discovered signal with Absent Regulatory Components) iModulon which consists predominantly of poorly characterized genes, contains two transcription factors, one of which (SACI_RS05830/saci_1223) is upregulated in all Nudix deletion strains. A relation of this iModulon to the cell membrane was proposed, however the large number of uncharacterized genes impedes further predictions (Chauhan et al., 2021).

Between the five Nudix deletion strains, 24 genes were identified to be upregulated in all strains in the early stationary phase (Supplementary Table S2). No common genes were identified for the mid-log phase nor downregulated genes in either growth phase. Roughly half of these genes encode hypothetical proteins or proteins

with domains of unknown function (DUF). Furthermore, several transcriptional regulators were affected in all deletion strains. Most of these genes are attributed to the iModulons of transcriptional regulation by XylR-SoxM or YtrA, as well as cold stress response proteins (Supplementary Table S2).

2.4. Nudix gene deletions indirectly affect transcription via transcriptional regulators

The transcriptional activator XylR regulates genes involved in xylose/arabinose uptake and its respective degradation pathway in *S. acidocaldarius*, while SoxM constitutes a terminal oxidase complex (Komorowski et al., 2002; van der Kolk et al., 2020). The respective iModulon is activated under nutrient-limited conditions, suggesting that it contains genes related to cell growth and starvation (Bischof et al., 2019; Chauhan et al., 2021). A general upregulation of numerous genes responsive to nutrient limitation and environmental stress might explain the lack of a distinct phenotype when the Nudix deletion strains were grown under carbon, nitrogen or heat stress conditions (Figure 2). Indeed, deletion of the Nudix hydrolases, especially SACI_RS00730, Saci_NudT5 and SACI_RS02625, elicits a transcriptome response highly similar to environmental stressors. Accordingly, genes attributed to several of these iModulons are significantly affected in all Nudix deletion strains and found to be mostly upregulated (Figure 4).

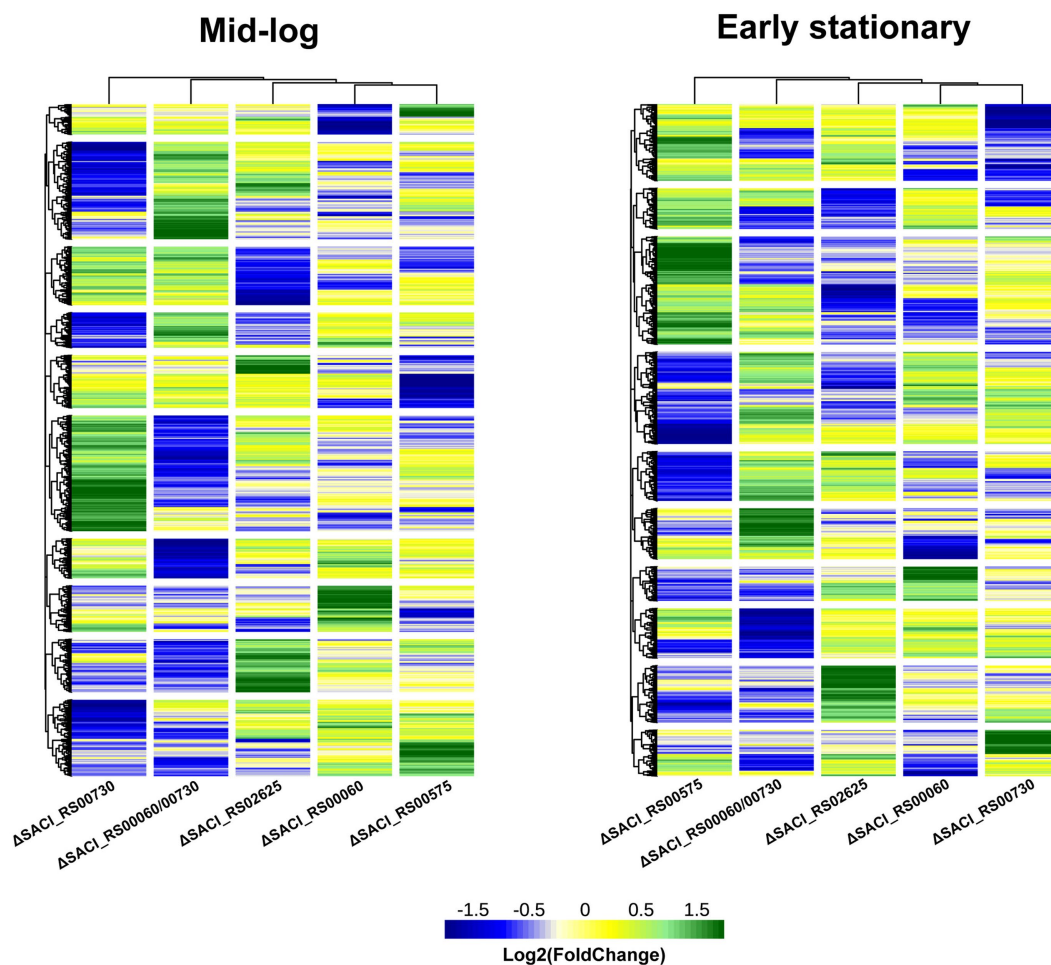


FIGURE 3

Transcriptomes of Nudix gene deletions. Heatmap of the log₂ (FoldChange) of the Nudix gene deletion strains' transcriptome relative to the wild-type strain *S. acidocaldarius* MW001 in the mid-log and early stationary growth phases and clustered according to Euclidean distances. Individual datasets can be found in [Supplementary Table S1](#). Green: upregulated genes, blue: downregulated genes, yellow: unaffected.

The Nudix deletion strains exhibit several upregulated transcriptional regulators, predominantly from the MarR family and four regulators which are present in all strains. The MarR family belongs to the super-group of transcriptional regulators present in bacteria and archaea predating the divergence of the domains (Pérez-Rueda and Collado-Vides, 2001). Its members can act as repressors or activators and their targets comprise genes involved in diverse cellular processes, such as antibiotic resistance, stress response, virulence and catabolism of aromatic compounds (Perera and Grove, 2010; Contursi et al., 2013). Hence, the effect of the Nudix gene deletions is possibly transmitted *via* interconnected transcriptional regulators. Notably, the highest number of differentially regulated transcriptional regulators is found in the datasets that contain the highest number of differentially regulated genes overall. On the same note, the number of differentially regulated genes also correlates with the number of upregulated thermosome subunits and Type II toxin-antitoxin (TA) system genes: one in Δ SACI_RS00730 and Δ SACI_RS00060, four in Δ SACI_RS00575, and five in Δ SACI_RS02625 and the double knock-out. In *Saccharolobus solfataricus*, the toxin VapC6 is a heat-dependent ribonuclease which is inactivated by VapB6 binding. The

ribonucleolytic activity is suggested to aid in the repression of protein synthesis during heat shock response (Maezato et al., 2011). Another study in *Sa. solfataricus* found *vapBC* locus expression to be heat-induced, however some *vapBC* loci were also expressed under normal growth conditions, suggesting roles beyond heat stress response for this system (Tachdjian and Kelly, 2006). The upregulation of a putative *vapBC* locus (encoded by SACI_RS10050/*saci_2079* and SACI_RS10055/*saci_2080*) in two Nudix deletion strains corroborates this idea. Interestingly, the same study not only found a large number of MarR family transcriptional regulators upregulated in the heat-shocked *Sa. solfataricus* samples, but also upregulation of *sso_3167* which encodes a Nudix family hydrolase (Tachdjian and Kelly, 2006). The homolog of *sso_3167* in *S. acidocaldarius* is SACI_RS02625, whose deletion strain did not show significantly reduced growth following the 87°C heat shock (Figure 2). Another study on heat shock response in *S. acidocaldarius* revealed high upregulation of the repressor YtrA (encoded by SACI_RS08880/*saci_1851*) upon heat stress (Baes et al., 2020). Though YtrA itself is not affected in the Nudix deletion strains, its iModulon is predominantly upregulated in Δ SACI_RS02625, Δ SACI_RS00575 and the double KO strain (Figure 4). Notably, YtrA

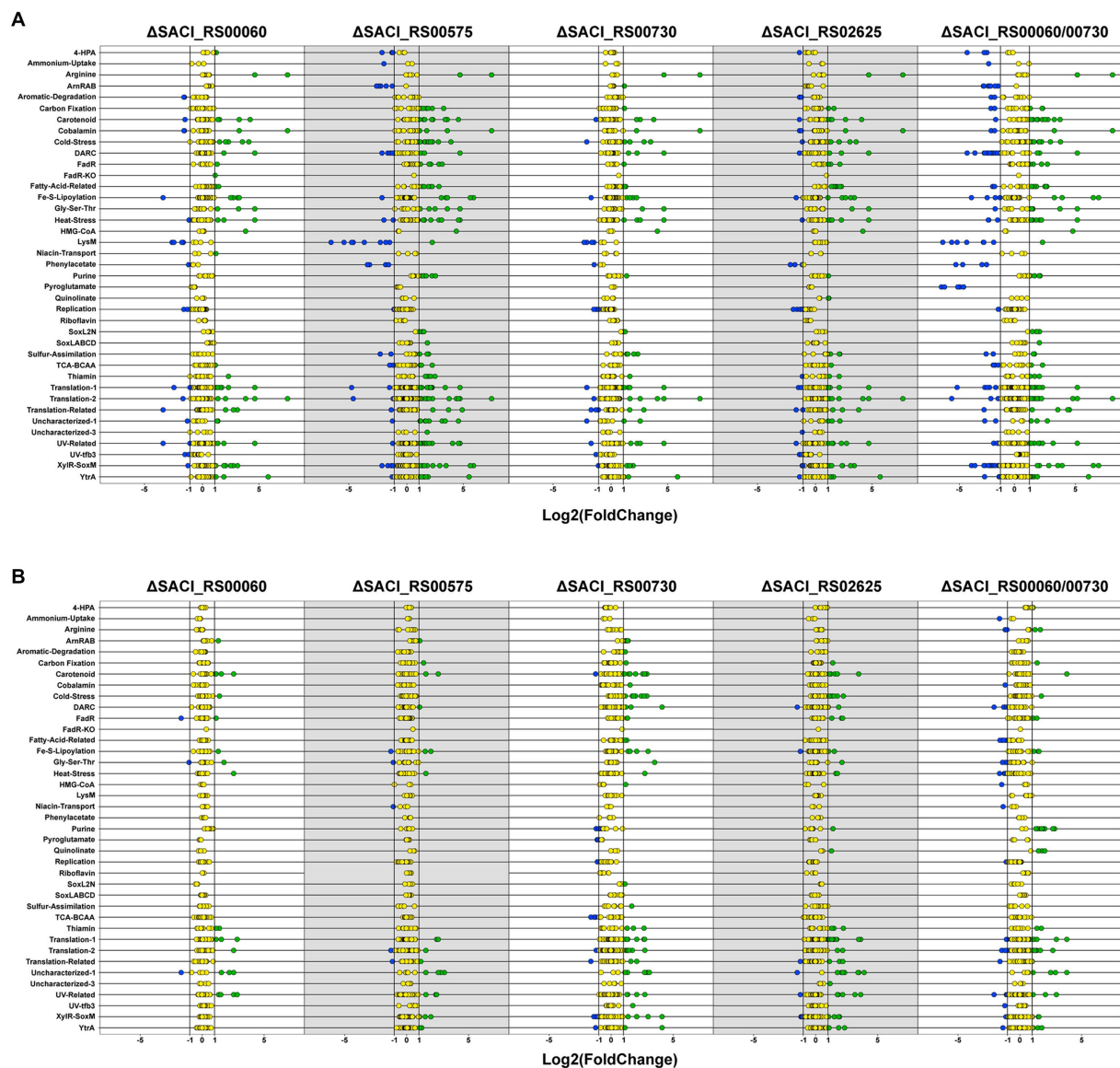


FIGURE 4

Differentially regulated genes in Nudix gene deletion strains. Dot plot of differentially regulated genes for all Nudix knockout strains in the early stationary (A) and mid-log phase (B). Genes were assembled into clusters of similarly regulated genes based on the iModulonDB database (Chauhan et al., 2021; Rychel et al., 2021). Green: upregulated genes, blue: downregulated genes, yellow: unaffected.

does not regulate known heat shock proteins but two putative membrane proteins of unknown function and though its connection to thermal stress remains unclear, a recent study links transcriptional regulators such as YtrA and Type II TA systems with thermal stress response in thermophilic archaea (Lemmens et al., 2019; Cooper et al., 2023). While the exact nature of the connection between Nudix hydrolases, specifically SACI_RS02625, and the heat stress response system remains unclear, our results suggest it to be more than coincidental. Notably, a definite transcription factor regulating heat shock response in the Sulfolobales remains to be identified.

The LysM iModulon represents lysine biosynthesis in *S. acidocaldarius* and contains specifically (though not exclusively) the genes of the lysWXJK operon (Chauhan et al., 2021). This operon encodes bi-functional enzymes involved in lysine and arginine biosynthesis and is activated by LysM. In turn, LysM is inactivated by

lysine and excess amounts of lysine were shown to inactivate the expression of the operon, leading to a shutdown of arginine synthesis (Brinkman et al., 2002; Ouchi et al., 2013). The LysM iModulon is downregulated in the early stationary phase in the Δ SACI_RS00575 strain and the double KO strain (Figure 4). However, downregulation of this operon was also observed under nutrient limitation (Bischof et al., 2019). Notably, the lysine biosynthesis operon is downregulated in the double KO strain but not in the respective single deletion strains Δ SACI_RS00730 and Δ SACI_RS00060 (Figure 4).

In *S. acidocaldarius*, the archaeum formation operon is controlled by the inducers ArnR and ArnR1 and the repressors ArnA and ArnB and represented in the datasets by the ArnRAB iModulon (Reimann et al., 2012; Lassak et al., 2013). ArnR (SACI_RS05625/saci_1180) and the archaeum formation iModulon are downregulated in Δ SACI_RS00575, in contrast to upregulation of

ArnR1 (*SACI_RS05580/saci_1171*) and the archaellum formation iModulon in Δ *SACI_RS00730* (Figure 4). The archaellum repressors ArnA and ArnB are not affected in any of the Nudix deletion strains. ArnR was observed to be induced under tryptone starvation conditions and is suggested to be controlled by a still unidentified superior transcription factor (Lassak et al., 2013). ArnR1 is upregulated in the double knock-out while the remaining archaellum formation iModulon is downregulated (Figure 4). ArnR and ArnR1 both promote motility to a different extent, as deletion mutants of *arnR* and *arnR1* exhibited a strongly and mildly diminished motility phenotype, respectively (Lassak et al., 2013). Several studies in different archaeal organisms report archaellum formation to be either repressed or stimulated under different nutrient limitation conditions (Mukhopadhyay et al., 2000; Szabó et al., 2007; Hendrickson et al., 2008; Xia et al., 2009). Arguably, intracellular nutrient limitations in Δ *SACI_RS00575* and the double knock-out strain may cause the cells to shut down energy-consuming archaellum production. Hence, downregulation of the archaellum formation iModulon may be indicative of the cell's energy state in consequence of Nudix gene deletion(s).

Furthermore, Δ *SACI_RS02625* and the double KO strain exhibit upregulation of *nadA* (a putative quinolinate synthase) and *nadB* (a putative L-aspartate oxidase). In the genome, *nadB* is located directly upstream of *SACI_RS02625*, however its upregulation in the double KO strain refutes a locational effect. Both gene products presumably catalyze the initial two steps in the NAD⁺ *de novo* biosynthesis pathway, as inferred from homology to *T. kodakarensis*. Upregulation of the *de novo* synthesis pathway is possibly elicited by a disturbance in the NAD⁺ salvage pathway, which involves recycling NAD⁺ from nicotinamide and ADPR following its thermal degradation (Hachisuka et al., 2018). Other genes involved in these pathways were not detected to be co-regulated in the respective datasets.

2.5. Concluding remarks

The genome of *S. acidocaldarius* encodes four Nudix domain-containing proteins, which were shown to be not essential to the organism under conditions tested. While preliminary phenotypical screenings did not exhibit a distinct deletion phenotype compared to the wild type strain, RNA-seq revealed unique transcriptome profiles for each Nudix deletion strain. Their transcriptomes were shown to be highly affected, especially regarding iModulons which assemble genes responsive to nutrient limitation and heat stress. Considering the high number of affected transcriptional regulators, the Nudix hydrolases might be involved in the pathways of metabolites that act as effector molecules to the transcriptional regulators, leading to an altered transcriptomic state as a consequence of altered metabolite composition. Indeed, the transcriptomes of the Nudix deletion strains exhibit remarkable similarity to the transcriptome in response to nutrient limitation stress (Bischof et al., 2019). Alternatively, Nudix hydrolases could also affect transcription *via* direct regulation of the levels of (capped) transcripts. We expect that these results stimulate the characterization of the Nudix hydrolases from *S. acidocaldarius*, which is proposed to include metabolomic profiling of the Nudix deletion strains.

3. Materials and methods

3.1. Generation of Nudix deletion strains

This work uses the uracil auxotrophic strain *Sulfolobus acidocaldarius* DSM639 MW001 (Wagner et al., 2012). Cultures were grown aerobically at 120 rpm and 75°C in Brock media at pH 3.5 (Brock et al., 1972). The media was supplied with 0.1% (w/v) NZ-amine, 0.2% (w/v) dextrin and 10 µg/ml uracil. Cell growth was determined by measuring the optical density (OD) at 600 nm with a cell density meter (Amersham Biosciences). The generation of markerless Nudix deletion strains was conducted using the deletion plasmid pSVA431, as described in Wagner et al. (2012). Briefly, this plasmid carries a dual marker system consisting of the uracil cassette *pyrEF* and the *lacS* gene from *Saccharolobus solfataricus* plus two multiple cloning sites, harboring part of the gene of interest and its upstream and downstream flanking regions, respectively. The entire cassette is transformed as a linear fragment and integrated into the genome *via* homologous recombination. Plasmids were constructed using Gibson Assembly (Gibson et al., 2009) with the primers listed in Supplementary Table S3. The genes encoding all four Nudix proteins *SACI_RS00730*, *Saci_NudT5* (*SACI_RS00060*), *SACI_RS02625* and *SACI_RS00575* were individually deleted from the genome of *S. acidocaldarius* DSM639 MW001 without interrupting their partially overlapping neighboring genes. To generate the double deletion strain, competent cells from the *Saci_NudT5* deletion strain were transformed with the linear marker cassette targeting *SACI_RS00730*. The successful removal of the Nudix genes was subsequently verified by Sanger sequencing of PCR products amplified from the deletion loci and RNA-sequencing. For PCR analysis of *S. acidocaldarius* cells, 20 µl cell culture was lysed in 20 µl 0.2 M NaOH for 5 min at RT, neutralized by addition of 80 µl 0.2 M Tris-HCl pH 6.5 and 5 µl suspension was used in a 20 µl reaction using DreamTaq DNA Polymerase (Thermo Scientific). Genomic DNA of *S. acidocaldarius* was isolated from 2 mL late logarithmic-phase cultures using the NucleoSpin® Tissue Kit (Macherey-Nagel), according to the manufacturer's instructions for cultured cells.

3.2. Preparation of electrocompetent *Sulfolobus acidocaldarius* cells

Cells were grown in 50 mL Brock medium supplied with 0.1% (w/v) NZ-Amine, 0.2% (w/v) dextrin and 10 µg/ml uracil, pH 3.5, at 75°C and 120 rpm to OD₆₀₀ = 0.3–0.7. A calculated amount of culture was subsequently transferred into 50 ml fresh medium, grown to OD₆₀₀ = 0.2–0.3 and then incubated on ice for 10–15 min. Cells were harvested by centrifugation for 15–20 min at 2500 × g and 4°C and the pellet was washed three times with each 30 ml of ice-cold 20 mM sucrose. Next, the pellet was resuspended in 1 ml of ice-cold 20 mM sucrose, transferred to a 1.5 ml aliquot and centrifuged for another 5 min at 2500 × g, 4°C. Finally, the pellet was resuspended in 20 mM ice-cold sucrose to a theoretical OD₆₀₀ = 20 and 50 µl aliquots were stored at –80°C without the use of liquid nitrogen until further use.

3.3. Transformation of *Sulfolobus acidocaldarius*

Prior to transformation into *S. acidocaldarius*, all plasmids or linearized DNA fragments were methylated to circumvent the activity of the restriction endonuclease SmaI (Berkner et al., 2007). To this end, plasmids were transformed into the strain *Escherichia coli* ER1821 (New England Biolabs) carrying the plasmid pM.ESABC4I. The methylated deletion plasmids were digested with NotI-HF (New England Biolabs) to yield linear fragments and electroporated in 1 mm Gene Pulser® electroporation cuvettes (Bio-Rad) with a constant time protocol using the input parameters 1.5 kV, 25 μ F and 600 Ω on a Gene Pulser® II electroporation system (Bio-Rad). Recovery was performed for 30 min at 75°C, 300 rpm, in Brock Recovery Medium (Brock medium supplied with 0.1% (w/v) NZ-Amine, no pH adjustment), before plating cells on uracil-lacking first selection plates. The plates were wrapped in wet paper towels, placed in plastic boxes to avoid drying out and incubated for 7 days at 75°C. For blue-white screening, plates were sprayed with 25 mg/mL X-gal in DMF diluted 1:5 in 20% (w/v) dextrin and incubated for 30 min at 75°C.

3.4. Isolation of total RNA

Sulfolobus acidocaldarius DSM639 and Nudix deletion strains were grown in duplicates in Brock media supplied with 0.1% (w/v) NZ-Amine, 0.2% (w/v) dextrin and 10 μ g/ml uracil at 75°C, 120 rpm, to an OD_{600} = 0.3 and 0.7, corresponding to mid-log and early stationary growth phases, respectively. 2 mL culture samples were pelleted by centrifugation for 15 min at max. Speed, RT, and total RNA was isolated using the mirVana™ miRNA Isolation Kit (Invitrogen) according to the manufacturer's instructions. Subsequently, total RNA extractions were digested with 1 U DNaseI/ μ g RNA (New England Biolabs) for 2 h at 37°C and cleaned up using the Monarch® RNA Cleanup Kit (50 μ g) (New England Biolabs). Quantitation of RNA samples was performed using the Qubit™ RNA High Sensitivity Assay (Agilent Technologies) and Qubit® 2.0 fluorometer (Thermo Fisher Scientific GmbH).

3.5. Library preparation for RNA-sequencing

Ribosomal RNAs were depleted using the Pan-Archaea riboPOOL probes (siTOOLS Biotech) and streptavidin-coated magnetic beads (siTOOLS Biotech) according to the manufacturer's instructions. Depleted RNA samples were cleaned up with the Monarch® RNA Cleanup Kit (10 μ g) (New England Biolabs) and successful rRNA depletion was verified with the RNA 6000 Pico Assay for the Agilent Bioanalyzer (Agilent Technologies) according to the manufacturer's instructions. The preparation of cDNA libraries from rRNA-depleted total RNA was performed using the NEBNext® Ultra™ II Directional RNA Library Prep Kit for Illumina® (New England Biolabs), AMPure XP SPRI beads (Beckman Coulter) and NEBNext® Multiplex Oligos for Illumina® (New England Biolabs), following the instructions of the NEBNext Library Prep Kit. Quality control

and size distribution of the cDNA libraries was assessed with the High Sensitivity DNA Assay for Bioanalyzer (Agilent Technologies). Sequencing was performed as 150 nt single reads on an Illumina® NextSeq550 at the Genomics Core Facility of the Philipps University Marburg.

3.6. Analysis of Illumina RNA-Seq data

Raw reads were adapter and quality trimmed using Cutadapt (v2.8) and checked with FASTQC (v0.11.9) (FastQC, 2015). Processed reads (≥ 18 nt) were mapped to the reference genome of *S. acidocaldarius* DSM639 (NC_007181.1) using Hisat2 (v2.2.1) (Kim et al., 2019). Multiple mapped reads with the exact match score were randomly distributed. After the strand-specific screening, HTSeq (v2.0.2) was used to count gene hits (Anders et al., 2015). Statistical and differential expression analyses were performed with DESeq2 (v1.36.0) and genes with a log2 fold-change (≤ -1 or ≥ 1), value of $p < 0.05$, and adjusted value of $p < 0.1$ were considered differentially expressed (Love et al., 2014). Next, genes were classified according to previously established iModulons (Chauhan et al., 2021) and their general expression profiles were analyzed. The Integrative Genomics Viewer (IGV, v2.13.2) was used for data inspection (Robinson et al., 2011).

3.7. RT-qPCR analyses

Total RNA samples (0.04 ng/ μ l) were used as template for RT-qPCR analysis using the KAPA SYBR fast one-step qRT-PCR kit (Merck) following the manufacturer's instructions. The reactions were carried out in a CFX384 Touch real-time PCR detection system (Bio-Rad). The gene SACI_RS06385 was used as an internal control. All primers used for RT-qPCR analysis are listed in the Supplementary Table S3. Each RNA sample was tested in triplicates, and data was analyzed using CFX Manager software (Bio-Rad).

3.8. Growth curves

Sulfolobus acidocaldarius DSM639 and Nudix deletion strains were grown as pre-cultures in 25 ml Brock media supplied with 0.1% (w/v) NZ-Amine, 0.2% (w/v) dextrin and 10 μ g/ml uracil at 75°C, 120 rpm. Upon reaching the stationary phase, a calculated volume of each strain was transferred into 50 ml fresh medium corresponding to a starting OD_{600} = 0.01. Each strain was grown in triplicates in Brock media + NZ-Amine + Dextrin + Uracil or Brock + Dextrin + Uracil or Brock + NZ-Amine + Uracil. At the indicated time points, 200 μ l from each culture were transferred into a 96 well plate and adsorption at 600 nm was measured using a CLARIOstar® Plus microplate reader (BMG Labtech).

3.9. Heat shock spotting assays

Heat shocking spotting assays were modified from Baes et al. (2020). *S. acidocaldarius* DSM639 and Nudix deletion strains were grown in triplicates in Brock media supplied with 0.1% (w/v)

NZ-Amine, 0.2% (w/v) dextrin and 10 µg/ml uracil at 75°C, 120 rpm, until reaching mid-logarithmic phase. Samples were transferred into pre-warmed aliquots and incubated for 15 min at 75°C, 300 rpm, in a thermomixer (StarLab GmbH) (“adaption period”). After removing a heat shock-control sample from each tube, tubes were covered with gas-permeable sealing membrane (Breathe-Easy, Diversified Biotech) and heat shock was administered for 30 min at 87°C, 300 rpm, on the thermomixer, using a digital pocket thermometer with a K-type probe (Traceable® Products) to monitor the temperature inside the liquid. Heat shock and control samples were diluted down to OD₆₀₀ = 0.1, followed by the preparation of a 10⁻¹ to 10⁻⁶ dilution series in Brock Recovery medium [Brock medium supplied with 0.1% (w/v) NZ-Amine, no pH adjustment]. Finally, 3 µl of each dilution of each sample were spotted onto a solid Brock Gelrite plate supplied with 0.1% (w/v) NZ-Amine, 0.2% (w/v) dextrin and 10 µg/ml uracil and plates were incubated in a plastic box lined with wet paper towels at 75°C. After 5 days, plates were photographed and cell viability was determined by measuring spot density from the 10⁻⁴ dilution step using the oval selection and area measurement tools from Fiji (Schindelin et al., 2012).

Data availability statement

The datasets presented in this study can be found in the online repository European Nucleotide Archive (ENA) under the accession number PRJEB60684.

Author contributions

RB, JG-F, and LR designed the experiments. RB, JY, and JG-F performed the experiments. RB, JY, and JG-F analyzed the data. RB wrote the manuscript with input from JG-F and LR. All authors reviewed and edited the manuscript.

References

- Abdelraheim, S. R., Spiller, D. G., and McLennan, A. G. (2003). Mammalian NADH diphosphatases of the Nudix family: cloning and characterization of the human peroxisomal NUDT12 protein. *Biochem. J.* 374, 329–335. doi: 10.1042/bj20030441
- Alm, E. J., Huang, K. H., Price, M. N., Koche, R. P., Keller, K., Dubchak, I. L., et al. (2005). The MicrobesOnline web site for comparative genomics. *Genome Res.* 15, 1015–1022. doi: 10.1101/gr.3844805
- 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
- Baes, R., Lemmens, L., Mignon, K., Carlier, M., and Peeters, E. (2020). Defining heat shock response for the thermoacidophilic model crenarchaeon *Sulfolobus acidocaldarius*. *Extremophiles* 24, 681–692. doi: 10.1007/s00792-020-01184-y
- Berkner, S., Grogan, D., Albers, S. V., and Lipps, G. (2007). Small multicopy, non-integrative shuttle vectors based on the plasmid pRN1 for *Sulfolobus acidocaldarius* and *Sulfolobus solfataricus*, model organisms of the (cren-)archaea. *Nucleic Acids Res.* 35, e88–e12. doi: 10.1093/nar/gkm449
- Bessman, M. J. (2019). A cryptic activity in the Nudix hydrolase superfamily. *Protein Sci.* 28, 1494–1500. doi: 10.1002/pro.3666
- Bessman, M. J., Frick, D. N., and O’Handley, S. F. (1996). The MutT proteins or “Nudix” hydrolases, a family of versatile, widely distributed, “housecleaning” enzymes. *J. Biol. Chem.* 271, 25059–25062. doi: 10.1074/jbc.271.41.25059
- Bischof, L. F., Haurat, M. F., Hoffmann, L., Albersmeier, A., Wolf, J., Neu, A., et al. (2019). Early response of *Sulfolobus acidocaldarius* to nutrient limitation. *Front. Microbiol.* 9, 1–17. doi: 10.3389/fmicb.2018.03201
- Brinkman, A. B., Bell, S. D., Lebbink, R. J., De Vos, W. M., and Der Van Oost, J. (2002). The *Sulfolobus solfataricus* Lrp-like protein LysM regulates lysine biosynthesis in response to lysine availability. *J. Biol. Chem.* 277, 29537–29549. doi: 10.1074/jbc.M203528200
- Brock, T. D., Brock, K. M., Belly, R. T., and Weiss, R. L. (1972). *Sulfolobus*: a new genus of sulfur-oxidizing bacteria living at low pH and high temperature. *Arch. Mikrobiol.* 84, 54–68. doi: 10.1007/BF00408082
- Cahová, H., Winz, M. L., Höfer, K., Nübel, G., and Jäschke, A. (2015). NAD captureSeq indicates NAD as a bacterial cap for a subset of regulatory RNAs. *Nature* 519, 374–377. doi: 10.1038/nature14020
- Chauhan, S. M., Poudel, S., Rychel, K., Lamoureux, C., Yoo, R., Al Bulushi, T., et al. (2021). Machine learning uncovers a data-driven transcriptional regulatory network for the crenarchaeal thermoacidophile *Sulfolobus acidocaldarius*. *Front. Microbiol.* 12:753521. doi: 10.3389/fmicb.2021.753521
- Chen, Y. G., Kowtoniuk, W. E., Agarwal, I., Shen, Y., and Liu, D. R. (2009). LC/MS analysis of cellular RNA reveals NAD-linked RNA. *Nat. Chem. Biol.* 5, 879–881. doi: 10.1038/nchembio.235
- Contursi, P., Fusco, S., Limauro, D., and Fiorentino, G. (2013). Host and viral transcriptional regulators in *Sulfolobus*: an overview. *Extremophiles* 17, 881–895. doi: 10.1007/s00792-013-0586-9
- Cooper, C. R., Lewis, A. M., Notey, J. S., Mukherjee, A., Willard, D. J., Blum, P. H., et al. (2023). Interplay between transcriptional regulators and VapBC toxin-antitoxin loci during thermal stress response in extremely Thermoacidophilic Archaea. *Environ. Microbiol.* 25, 1200–1215. doi: 10.1111/1462-2920.16350

Funding

This work was funded by the German Research Foundation (DFG) (Grant RA 2169/8–1 to LR). Open Access funding was provided by the Open Access Publishing Fund of Philipps-Universität Marburg with support of the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation).

Acknowledgments

We thank Sonja-Verena Albers and Marleen van Wolferen for providing strain *S. acidocaldarius* DSM639 MW001 and plasmid pSVA431 and for teaching us genetic manipulation of *S. acidocaldarius*.

Conflict of interest

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.

Publisher’s note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2023.1197877/full#supplementary-material>

- Deana, A., Celesnik, H., and Belasco, J. G. (2008). The bacterial enzyme RppH triggers messenger RNA degradation by 5' pyrophosphate removal. *Nature* 451, 355–358. doi: 10.1038/nature06475
- Dunn, C. A., O'Handley, S. F., Frick, D. N., and Bessman, M. J. (1999). Studies on the ADP-ribose pyrophosphatase subfamily of the Nudix hydrolases and tentative identification of *trgB*, a gene associated with tellurite resistance. *J. Biol. Chem.* 274, 32318–32324. doi: 10.1074/jbc.274.45.32318
- Frick, D. N., and Bessman, M. J. (1995). Cloning, purification, and properties of a novel NADH pyrophosphatase. Evidence for a nucleotide pyrophosphatase catalytic domain in MutT-like enzymes. *J. Biol. Chem.* 270, 1529–1534. doi: 10.1074/jbc.270.4.1529
- Frindert, J., Zhang, Y., Nübel, G., Kahloon, M., Kolmar, L., Hotz-Wagenblatt, A., et al. (2018). Identification, biosynthesis, and Decapping of NAD-capped RNAs in *B. subtilis*. *Cell Rep.* 24, 1890–1901.e8. doi: 10.1016/j.celrep.2018.07.047
- Gabelli, S. B., Bianchet, M. A., Xu, W. L., Dunn, C. A., Niu, Z. D., Amzel, L. M., et al. (2007). Structure and function of the *E. coli* Dihydroneopterin triphosphate pyrophosphatase: a Nudix enzyme involved in folate biosynthesis. *Structure* 15, 1014–1022. doi: 10.1016/j.str.2007.06.018
- Gibson, D. G., Young, L., Chuang, R., Venter, J. C., Hutchison, C., and Smith, H. O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* 6, 343–345. doi: 10.1038/nmeth.1318
- Gomes-Filho, J. V., Breuer, R., Morales-Fillo, H. G., Pozhydaeva, N., Börst, A., Paczia, N., et al. (2022). Identification of NAD-RNAs and ADPR-RNA decapping in the archaeal model organisms *Sulfolobus acidocaldarius* and *Haloferax volcanii*. *bioRxiv*. doi: 10.1101/2022.11.02.514978
- Grudzien-Nogalska, E., Wu, Y., Jiao, X., Cui, H., Mateyak, M. K., Hart, R. P., et al. (2019). Structural and mechanistic basis of mammalian NudT12 RNA deNADding. *Nat. Chem. Biol.* 15, 575–582. doi: 10.1038/s41589-019-0293-7
- Hachisuka, S., Sato, T., and Atomi, H. (2018). Hyperthermophilic archaeon *Thermococcus kodakarensis* utilizes a four-step pathway for NAD⁺ salvage through nicotinamide deamination. *J. Bacteriol.* 200:e00785–17. doi: 10.1128/JB.00785-17
- Hendrickson, E. L., Liu, Y., Rosas-Sandoval, G., Porat, I., Söll, D., Whitman, W. B., et al. (2008). Global responses of *Methanococcus maripaludis* to specific nutrient limitations and growth rate. *J. Bacteriol.* 190, 2198–2205. doi: 10.1128/JB.01805-07
- Höfer, K., Li, S., Abele, F., Frindert, J., Schlotthauer, J., Grawenhoff, J., et al. (2016). Structure and function of the bacterial decapping enzyme NudC. *Nat. Chem. Biol.* 12, 730–734. doi: 10.1038/nchembio.2132
- Hori, M., Fujikawa, K., Kasai, H., Harashima, H., and Kamiya, H. (2005). Dual hydrolysis of diphosphate and triphosphate derivatives of oxidized deoxyadenosine by Orf17 (NtpA), a MutT-type enzyme. *DNA Repair* 4, 33–39. doi: 10.1016/j.dnarep.2004.07.010
- Hudeček, O., Benoni, R., Reyes-Gutierrez, P. E., Culka, M., Šanderová, H., Hubálek, M., et al. (2020). Dinucleoside polyphosphates act as 5'-RNA caps in bacteria. *Nat. Commun.* 11, 1–11. doi: 10.1038/s41467-020-14896-8
- Jiao, X., Doamekpor, S. K., Bird, J. G., Nickels, B. E., Tong, L., Hart, R. P., et al. (2017). 5' end nicotinamide adenine dinucleotide cap in human cells promotes RNA decay through DXO-mediated deNADding. *Cells* 168, 1015–1027.e10. doi: 10.1016/j.cell.2017.02.019
- Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nat. Cell Biol.* 596, 583–589. doi: 10.1038/s41586-021-03819-2
- Kim, D., Paggi, J. M., Park, C., Bennett, C., and Salzberg, S. L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* 37, 907–915. doi: 10.1038/s41587-019-0201-4
- Komorowski, L., Verheyen, W., and Schäfer, G. (2002). The archaeal respiratory supercomplex SoxM from *S. acidocaldarius* combines features of quinone and cytochrome c oxidases. *Biol. Chem.* 383, 1791–1799. doi: 10.1515/BC.2002.200
- Lassak, K., Peeters, E., Wróbel, S., and Albers, S. (2013). The one-component system ArnR: a membrane-bound activator of the crenarchaeal archaeellum. *Mol. Microbiol.* 88, 125–139. doi: 10.1111/mmi.12173
- Lemmens, L., Tillemann, L., De Koning, E., Valegård, K., Lindås, A. C., Van Nieuwerburgh, F., et al. (2019). YtrA_{Ssa}, a GntR-family transcription factor, represses two genetic loci encoding membrane proteins in *Sulfolobus acidocaldarius*. *Front. Microbiol.* 10:2084. doi: 10.3389/fmicb.2019.02084/BIBTEX
- Li, Y., Song, M., and Kiledjian, M. (2011). Differential utilization of decapping enzymes in mammalian mRNA decay pathways. *RNA* 17, 419–428. doi: 10.1261/rna.2439811
- Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15, 550–521. doi: 10.1186/s13059-014-0550-8
- Maezato, Y., Daugherty, A., Dana, K., Soo, E., Cooper, C., Tachdjian, S., et al. (2011). VapC6, a ribonucleolytic toxin regulates thermophilicity in the crenarchaeote *Sulfolobus solfataricus*. *RNA* 17:1381. doi: 10.1261/rna.2679911
- McLennan, A. G. (2006). The Nudix hydrolase superfamily. *Cell. Mol. Life Sci.* 63, 123–143. doi: 10.1007/s00018-005-5386-7
- McLennan, A. G. (2013). Substrate ambiguity among the nudix hydrolases: biologically significant, evolutionary remnant, or both? *Cell. Mol. Life Sci.* 70, 373–385. doi: 10.1007/s00018-012-1210-3
- Mukhopadhyay, B., Johnson, E. F., and Wolfe, R. S. (2000). A novel P(H2) control on the expression of flagella in the hyperthermophilic strictly hydrogenotrophic methanarchaeon *Methanococcus jannaschii*. *Proc. Natl. Acad. Sci. U. S. A.* 97, 11522–11527. doi: 10.1073/pnas.97.21.11522
- Ouchi, T., Tomita, T., Horie, A., Yoshida, A., Takahashi, K., Nishida, H., et al. (2013). Lysine and arginine biosyntheses mediated by a common carrier protein in *Sulfolobus*. *Chem. Biol.* 9, 277–283. doi: 10.1038/nchembio.1200
- Perera, I. C., and Grove, A. (2010). Molecular mechanisms of ligand-mediated attenuation of DNA binding by MarR family transcriptional regulators. *J. Mol. Cell Biol.* 2, 243–254. doi: 10.1093/jmcb/mjq021
- Pérez-Rueda, E., and Collado-Vides, J. (2001). Common history at the origin of the position–function correlation in transcriptional regulators in Archaea and Bacteria. *J. Molec. Evol.* 53, 172–179. doi: 10.1007/s002390010207
- Reimann, J., Lassak, K., Khadouma, S., Ettema, T. J. G., Yang, N., Driessen, A. J. M., et al. (2012). Regulation of archaeal expression by the FHA and von Willebrand domain-containing proteins ArnA and ArnB in *Sulfolobus acidocaldarius*. *Mol. Microbiol.* 86, 24–36. doi: 10.1111/j.1365-2958.2012.08186.x
- Richards, J., Liu, Q., Pellegrini, O., Celesnik, H., Yao, S., Bechhofer, D. H., et al. (2011). An RNA Pyrophosphohydrolase triggers 5'-exonucleolytic degradation of mRNA in *Bacillus subtilis*. *Mol. Cell* 43, 940–949. doi: 10.1016/j.molcel.2011.07.023
- Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., et al. (2011). Integrative genomics viewer. *Nat. Biotechnol.* 29, 24–26. doi: 10.1038/nbt.1754
- Rodionov, D. A., De Ingeniis, J., Mancini, C., Cimadamore, F., Zhang, H., Osterman, A. L., et al. (2008). Transcriptional regulation of NAD metabolism in bacteria: NtrR family of Nudix-related regulators. *Nucleic Acids Res.* 36, 2047–2059. doi: 10.1093/nar/gkn047
- Ruiz-Larrabeiti, O., Benoni, R., Zemlianski, V., Hanišáková, N., Schwarz, M., Brezovská, B., et al. (2021). NAD⁺ capping of RNA in Archaea and mycobacteria. *BioRxiv*. doi: 10.1101/2021.12.14.472595
- Rychel, K., Decker, K., Sastry, A. V., Phaneuf, P. V., Poudel, S., and Palsson, B. O. (2021). iModulonDB: a knowledgebase of microbial transcriptional regulation derived from machine learning. *Nucleic Acids Res.* 49, D112–D120. doi: 10.1093/nar/gkaa810
- Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., et al. (2012). Fiji: an open-source platform for biological-image analysis. *Nat. Methods* 9, 676–682. doi: 10.1038/nmeth.2019
- Song, M.-G., Bail, S., and Kiledjian, M. (2013). Multiple Nudix family proteins possess mRNA decapping activity. *RNA (New York, N.Y.)* 19, 390–399. doi: 10.1261/rna.037309.112
- Song, M.-G., Li, Y., and Kiledjian, M. (2010). Multiple mRNA decapping enzymes in mammalian cells. *Mol. Cell* 40, 423–432. doi: 10.1016/j.molcel.2010.10.010
- Srouji, J. R., Xu, A., Park, A., Kirsch, J. F., and Brenner, S. E. (2017). The evolution of function within the Nudix homology clan. *Proteins* 85, 775–811. doi: 10.1002/prot.25223
- Szabó, Z., Sani, M., Groeneveld, M., Zolghadr, B., Schelert, J., Albers, S. V., et al. (2007). Flagellar motility and structure in the hyperthermoacidophilic archaeon *Sulfolobus solfataricus*. *J. Bacteriol.* 189, 4305–4309. doi: 10.1128/JB.00042-07
- Tachdjian, S., and Kelly, R. M. (2006). Dynamic metabolic adjustments and genome plasticity are implicated in the heat shock response of the extremely thermoacidophilic archaeon *Sulfolobus solfataricus*. *J. Bacteriol.* 188, 4553–4559. doi: 10.1128/JB.00080-06
- van der Kolk, N., Wagner, A., Wagner, M., Waßmer, B., Siebers, B., and Albers, S. V. (2020). Identification of XylR, the activator of arabinose/xylose inducible regulon in *Sulfolobus acidocaldarius* and its application for homologous protein expression. *Front. Microbiol.* 11:1066. doi: 10.3389/fmicb.2020.01066
- Varadi, M., Anyango, S., Deshpande, M., Nair, S., Natassia, C., Yordanova, G., et al. (2022). AlphaFold protein structure database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.* 50, D439–D444. doi: 10.1093/nar/gkab1061
- Wagner, M., van Wolferen, M., Wagner, A., Lassak, K., Meyer, B. H., Reimann, J., et al. (2012). Versatile genetic tool box for the crenarchaeote *Sulfolobus acidocaldarius*. *Front. Microbiol.* 3, 1–12. doi: 10.3389/fmicb.2012.00214
- Walters, R. W., Matheny, T., Mizoue, L. S., Rao, B. S., Muhlrud, D., and Parker, R. (2017). Identification of NAD⁺ capped mRNAs in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. U. S. A.* 114, 480–485. doi: 10.1073/pnas.1619369114
- Weixler, L., Feijs, K. L. H., and Zaja, R. (2022). ADP-ribosylation of RNA in mammalian cells is mediated by TRPT1 and multiple PARPs. *Nucleic Acids Res.* 50, 9426–9441. doi: 10.1093/nar/gkac711
- Xia, Q., Wang, T., Hendrickson, E. L., Lie, T. J., Hackett, M., and Leigh, J. A. (2009). Quantitative proteomics of nutrient limitation in the hydrogenotrophic methanogen *Methanococcus maripaludis*. *BMC Microbiol.* 9, 1–10. doi: 10.1186/1471-2180-9-149



OPEN ACCESS

EDITED BY

Marleen van Wolferen,
University of Freiburg, Germany

REVIEWED BY

Lennart Randau,
University of Marburg, Germany
Jose Luis Llacer,
Spanish National Research Council (CSIC),
Spain

*CORRESPONDENCE

Tamara Basta
✉ tamara.basta@i2bc.paris-saclay.fr

PRESENT ADDRESS

Violette Da Cunha,
Génomique Métabolique, Genoscope,
Institut François Jacob,
CEA, CNRS, Univ Evry,
Université Paris-Saclay,
Evry, France

RECEIVED 11 April 2023

ACCEPTED 02 June 2023

PUBLISHED 21 June 2023

CITATION

Pichard-Kostuch A, Da Cunha V, Oberto J,
Sauguet L and Basta T (2023) The universal
Sua5/TsaC family evolved different
mechanisms for the synthesis of a key tRNA
modification.

Front. Microbiol. 14:1204045.

doi: 10.3389/fmicb.2023.1204045

COPYRIGHT

© 2023 Pichard-Kostuch, Da Cunha, Oberto,
Sauguet and Basta. This is an open-access
article distributed under the terms of the
[Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/).
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 universal Sua5/TsaC family evolved different mechanisms for the synthesis of a key tRNA modification

Adeline Pichard-Kostuch¹, Violette Da Cunha^{1†}, Jacques Oberto¹,
Ludovic Sauguet² and Tamara Basta^{1*}

¹CEA, CNRS, Institute for Integrative Biology of the Cell (I2BC), Université Paris-Saclay, Gif-sur-Yvette, France, ²Architecture and Dynamics of Biological Macromolecules, Institut Pasteur, Université Paris Cité, CNRS, UMR 3528, Paris, France

TsaC/Sua5 family of enzymes catalyzes the first step in the synthesis of N⁶-threonyl-carbamoyl adenosine (t⁶A) one of few truly ubiquitous tRNA modifications important for translation accuracy. TsaC is a single domain protein while Sua5 proteins contains a TsaC-like domain and an additional SUA5 domain of unknown function. The emergence of these two proteins and their respective mechanisms for t⁶A synthesis remain poorly understood. Here, we performed phylogenetic and comparative sequence and structure analysis of TsaC and Sua5 proteins. We confirm that this family is ubiquitous but the co-occurrence of both variants in the same organism is rare and unstable. We further find that obligate symbionts are the only organisms lacking *sua5* or *tsaC* genes. The data suggest that Sua5 was the ancestral version of the enzyme while TsaC arose via loss of the SUA5 domain that occurred multiple times in course of evolution. Multiple losses of one of the two variants in combination with horizontal gene transfers along a large range of phylogenetic distances explains the present day patchy distribution of Sua5 and TsaC. The loss of the SUA5 domain triggered adaptive mutations affecting the substrate binding in TsaC proteins. Finally, we identified atypical Sua5 proteins in Archaeoglobi archaea that seem to be in the process of losing the SUA5 domain through progressive gene erosion. Together, our study uncovers the evolutionary path for emergence of these homologous isofunctional enzymes and lays the groundwork for future experimental studies on the function of TsaC/Sua5 proteins in maintaining faithful translation.

KEYWORDS

universal proteins, enzyme, Sua5, TsaC, t⁶A, tRNA, evolution

Introduction

tRNA requires post-transcriptional maturation to be functional during the translation process. This involves *inter alia* the enzymatic modification of the canonical A, U, C and G bases to form modified nucleosides by addition of a variety of chemical groups (Väre et al., 2017). About hundred different modified nucleosides have been identified so far in tRNA, 20 of which are universal and were likely inherited from the Last Universal Common Ancestor (LUCA; Cantara et al., 2011; Machnicka et al., 2014). Among those, N⁶-threonyl-carbamoyl-adenosine (t⁶A) is found on the adenosine in position 37, next to the anticodon of almost all tRNAs that recognize ANN codons (where N = G, A, C or U; Figure 1A; Ishikura et al., 1969; Powers and Peterkofsky, 1972).

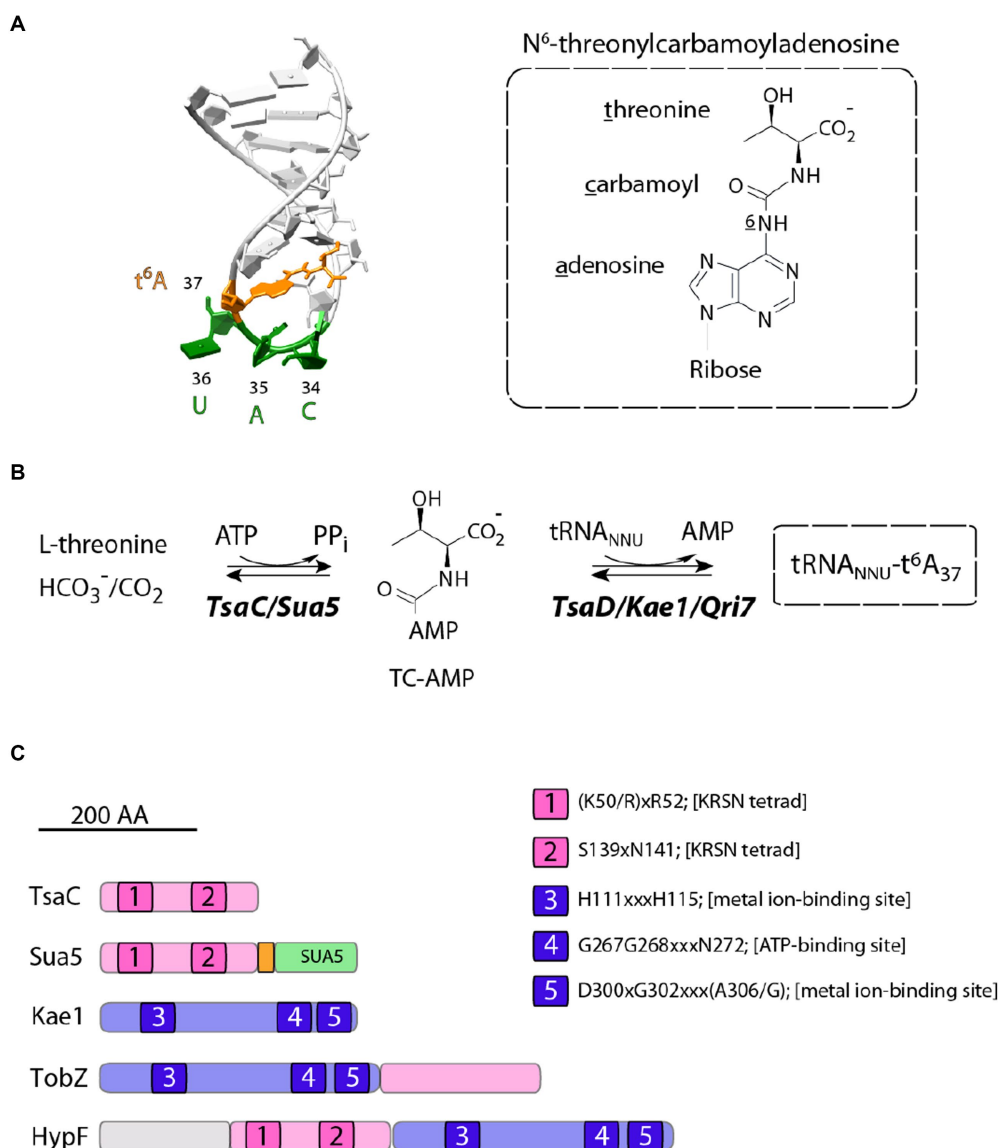


FIGURE 1

Structure and biosynthesis of t^6A_{37} nucleoside. (A) On the left is shown the structure of the anticodon loop of *Schizosaccharomyces pombe* $tRNA^{Met}$ (PDB 2G1G). The anticodon residues and t^6A_{37} nucleoside are colored in green and orange, respectively. The chemical formula of t^6A_{37} is depicted on the right. (B) Proposed mechanism for t^6A_{37} biosynthesis. (C) Primary structures of proteins containing TsaC orthologous domains. All proteins are drawn to scale and the approximate position of conserved motifs (1–5) is indicated. The numbering corresponds to that of *E. coli* TsaC and TsaD proteins. Sua5 proteins contain TsaC-orthologous domain with a loop of about 20 residues (orange) fused to SUA5 domain of about 100 residues (green). TobZ proteins catalyze the formation of the antibiotic nebramycin using the hydrolysis of carbamoyl phosphate and its subsequent adenylation by ATP to yield O-carbamoyladenylate. HypF is a carbamoyl transferase involved in the maturation of [NiFe] hydrogenases. HypF uses carbamoylphosphate as a substrate and transfers the carboxamido moiety in an ATP-dependent reaction to the thiolate of the C-terminal cysteine of HypE yielding a protein-S-carboxamide.

t^6A is required for translation accuracy (Högenauer et al., 1972; Weissenbach and Grosjean, 1981; Stuart et al., 2000; Murphy et al., 2004; El Yacoubi et al., 2011), aminoacylation of some tRNA species (Nureki et al., 1994), translation initiation (Thiaville et al., 2016; Llácer et al., 2018) and translocation of the ribosome on mRNA (Phelps et al., 2004). The mutations in t^6A synthetic genes were linked with a variety of phenotypes in eukaryotes and prokaryotes ranging from cell death or slow growth to short telomeres and defects in transcription probably reflecting the far-reaching indirect consequences of the unfaithful translation (El Yacoubi et al., 2009; Naor et al., 2012;

Beenstock and Sicheri, 2021). In humans, mutations in all t^6A -synthetic genes were linked with Galloway-Mowat syndrome a severe genetic disease causing brain malformation and renal dysfunction (Braun et al., 2017; Edvardson et al., 2017; Arrondel et al., 2019).

t^6A -tRNA biosynthesis proceeds in two steps and requires multiple proteins. The universal enzyme family Sua5/TsaC uses L-threonine, ATP and bicarbonate/carbon dioxide (HCO_3^-/CO_2) to synthesize the activated intermediate threonyl-carbamoyl-adenylate (TC-AMP) via a complex and poorly understood mechanism (Figure 1B; El Yacoubi et al., 2009; Lauhon, 2012; Perrochia et al.,

2013a,b; Harris et al., 2015). In the second step, the threonyl-carbamoyl moiety is transferred from the TC-AMP onto the substrate tRNA. This reaction is catalyzed by Kae1/TsaD/Qri7 universal protein family (Figure 1B) which associates with accessory proteins to form the DEZ complex in bacteria and the KEOPS complex in archaea and eukaryotes [reviewed in Thiaville et al. (2014b)].

TsaC/Sua5 family of proteins is a part of a restricted set of about 60 proteins shared by all cellular organisms (Koonin, 2003; Galperin, 2004). Remarkably, this family is composed of two distinct variants, TsaC and Sua5, which share a homologous catalytic domain but differ in the presence of a second domain, named SUA5 (Figure 1C), found only in Sua5 proteins. Initial bioinformatics analysis reported no clear phyletic pattern underpinning the distribution of TsaC and Sua5 proteins and additionally showed that the simultaneous occurrence of both genes in a genome seems to be rare (Thiaville et al., 2014b). TsaC-like domain is also found fused to Kae1-like domain in HypF proteins involved in the maturation of [NiFe] hydrogenases and in TobZ proteins that synthesize the antibiotic nebramycin (Figure 1C). In these proteins, the TsaC-like domains use carbamoyl phosphate to catalyze a carbamoylation reaction similar to a putative step in TC-AMP synthesis (Petkun et al., 2011; Parthier et al., 2012).

The structures of TsaC from *E. coli* (Ec-TsaC; Teplova et al., 2000; Harris et al., 2015) and Sua5 from the *Sulfolobus tokodaii* (St-Sua5; Agari et al., 2008; Kuratani et al., 2011; Parthier et al., 2012) and *Pyrococcus abyssi* (Pa-Sua5; Pichard-Kostuch et al., 2018) show that the catalytic domain adopts a globular twisted fold made of 7 parallel and anti-parallel β -strands bordered by 7 α -helices. The active site is in the cavity formed by this domain and contains highly conserved residues K⁵⁶XR⁵⁸/S¹⁴³XN¹⁴⁵ and T³³/S¹⁸¹/R¹⁹⁵ (Pa-Sua5 numbering) involved in ATP and threonine binding, respectively. The mutational analyses showed that these residues were all required for TC-AMP synthesis by Sua5 from *S. cerevisiae* (Wan et al., 2013).

The TsaC-like catalytic domain of Sua5 proteins carries a C-terminal extension of about 20 residues forming a flexible loop that contains a set of highly conserved residues, Pro²²⁸-Gly²²⁹-Met²³⁰ and His²³⁴-Tyr²³⁵ (Pa-Sua5 numbering; Pichard-Kostuch et al., 2018). The loop folds into the active site gorge and was suggested to act as a gate regulating the binding and release of ligands (Pichard-Kostuch et al., 2018).

The C-terminal extremity of the loop is anchored in the SUA5 domain which is composed of about 100 residues (Agari et al., 2008; Kuratani et al., 2011; Pichard-Kostuch et al., 2018). The domain exhibits an atypical Rossmann fold composed of five β -strands (β 12- β 16) and three α -helices (α 9- α 11). SUA5 domain and the catalytic TsaC-like domain form a tight interface that was shown to be important for the activity (Pichard-Kostuch et al., 2018).

How TsaC and Sua5 emerged in course of evolution and what are the functional differences between these two different protein scaffolds remains elusive. To address these questions, we performed a comprehensive phylogenetic, structural, and sequence analysis of Sua5/TsaC proteins. The data suggest pre-LUCA origin of this family and a complex evolution including gene erosion, multiple gene losses and horizontal gene transfers. We identified conserved residues specific for TsaC or Sua5 proteins and tentatively assigned them a role in substrate binding. Finally, we identified atypical Sua5 proteins which seem to be in the process of losing the SUA5 domain thus potentially ongoing the transition toward TsaC homologs. Together,

this work provides testable hypothesis for uncovering the functional differences between TsaC and Sua5 proteins and, more generally, provides insights into the evolutionary mechanisms driving the emergence of isofunctional enzymes.

Materials and methods

Extraction of Sua5 and TsaC sequences and distribution analysis

Sua5 and TsaC sequences were extracted from UniProt (version 06/2021 or version 11/2022) and UniRef (version 11/2022). TsaC/Sua5 sequences in DPANN archaea were retrieved using BLASTp searches with Pa-Sua5 protein sequence as query against all genome assemblies available (nr database, version 11/2022). The hits were filtered for the presence of KxR/SxN tetrad and one sequence per species was chosen to reduce redundancy. Following this pipeline, 5,889 bacterial, 761 archaeal and 1,318 eukaryotic sequences were recovered. The identified Sua5 or TsaC orthologs were mapped on the RNA polymerase-based universal phylogenetic tree (Da Cunha et al., 2017) or on the previously established phylogeny of DPANN archaea (Moody et al., 2022). The organisms were ranked according to the NCBI taxonomy database (as of 11/2022) or, for DPANN organisms, based on (Moody et al., 2022). The accession numbers for all sequences are available at figshare repository (doi: 10.6084/m9.figshare.22283929).

Phylogenetic analysis

For computing the universal phylogenetic tree we chose a set of species that has been previously vetted for its ability to produce a robust phylogenetic signal using universal proteins (Da Cunha et al., 2017). The corresponding Sua5/TsaC sequences (Supplementary Tables 2, 3) were extracted from Uniprot database, aligned using MAFFT v7 with auto settings (Yamada et al., 2016). The alignment was trimmed using BMGE (Criscuolo and Gribaldo, 2010) with BLOSUM30 matrix leaving 153 positions for tree construction, corresponding only to the catalytic domain residues (the SUA5 domain was removed by the trimming software). In addition, we generated an alignment of full length Sua5 sequences that contained 470 positions. The maximum likelihood (ML) trees were inferred using IQ-TREE v1.4.3 (Trifinopoulos et al., 2016) with the TESTNEW option for model selection. The branch support was obtained using nonparametric bootstrap (100 replicates), SH-aLRT test and ultrafast bootstrap approximation (1,000 replicates; Minh et al., 2013; Kalyaanamoorthy et al., 2017) or booster (Lemoine et al., 2018).

Archaeoglobi Sua5 sequences were extracted from Uniprot (doi: 10.6084/m9.figshare.22283929) and the Maximum Likelihood tree was inferred using Phylogeny.fr web page (Dereeper et al., 2008).

Horizontal gene transfer analysis

To screen for putative horizontally transferred genes in Archaea we extracted all sequences annotated as TsaC/Sua5 from

Uniprot database (9,553 sequences, database release 2019_06). We filtered this dataset to remove sequences containing less than 100 amino acids or more than 400 amino acids as well as those without the KRSN tetrad. The remaining set of 1,275 archaeal TsaC/Sua5 sequences were classified according to their taxonomic affiliation to 13 different archaeal taxa (doi: 10.6084/m9.figshare.22283929). This allowed to identify candidate proteins in “mixed” groups containing both Sua5 and TsaC users. To reduce the impact of convergent evolution on small proteins, we focused our analyses on transfer of *sua5* genes. We used BLASTp searches against the nr protein database to identify the most similar sequences to the candidate query protein. If the best-hit species was only distantly related to the query species, the query protein was selected as candidate for HGT. The HGT candidate were subjected to phylogenetic analysis to determine their origin. The phylogenetic inferences were done as indicated above (phylogenetic analysis section).

Protein sequence alignment

To identify Sua5- or TsaC-specific residues we aligned four representative sequences for TsaC and for TsaC-like domain from each domain of life (doi: 10.6084/m9.figshare.22283929) using MAFFT v7 with auto settings (Yamada et al., 2016).

To examine the distribution of the signature residue Pro¹⁴³/Thr¹³⁸ we retrieved 26,036 TsaC and Sua5 sequences from the Complete Genome Data Bank of the NCBI (database release 2015_05) using BLASTN, with Sua5 from *Pyrococcus abyssi* or TsaC from *Escherichia coli* as query. Using hmmsearch (Finn et al., 2011) we further retrieved 1,088 eukaryotic TsaC and Sua5 sequences from the Uniprot-Proteomes database (database release 2018_11). The hit sequences missing the KRSN catalytic tetrad or partial sequences were eliminated from the analysis. One UniRef90 sequence per genus was kept for sequence alignment, 654 for prokaryotes and 564 for eukaryotes in total (accession numbers were deposited to doi: 10.6084/m9.figshare.22283929). Sequence alignment depiction was done using ESPript3 with default parameters (Robert and Gouet, 2014).

Structure prediction and analysis

AlphaFold2 (Jumper et al., 2021) batch colab¹ was used to predict *de novo* structures of atypical Sua5 proteins from *Archaeoglobus* archaea: *Archaeoglobus profundus* (Ap-Sua5, UniProt D2RFV3), *A. veneficus* (Av-Sua5, UniProt F2KMZ1) and *A. fulgidus* (Af-Sua5, UniProt O29477). Crystal structures of several TsaC and Sua5 proteins were used for structural comparison. The PDB accession codes of those structures are 1HRU and 2MX1 for Ec-TsaC, 2EQA, 3AJE and 4E1B for St-Sua5, 6F87 for Pa-Sua5, 3VEZ for St-TobZ and 3TTC for Ec-HypF. Structures were visualized and edited with UCSF ChimeraX v. 1.4 (Pettersen et al., 2021).

Results

The distribution of Sua5 and TsaC proteins does not follow a clear phyletic pattern

Previous work established that TsaC/Sua5 family of proteins (COG0009) is ubiquitous however the specific distribution of TsaC or Sua5 orthologs was only examined for a small number of representative isolated species.

Here, we exploited the recent massive increase of (meta)genome sequencing data to perform an up to date comprehensive analysis of the distribution of Sua5 and TsaC proteins among living organisms. To this end, we retrieved 5,889 bacterial, 761 archaeal and 1,318 eukaryotic sequences from public databases and mapped their presence on reference phylogenetic trees (Figure 2; Supplementary Table 1; Supplementary Figure 1).

The data confirm the ubiquitous distribution of the Sua5/TsaC family among organisms. We found that the obligate ectosymbiotic archaea *Nanoarchaeum equitans*, and *Nanobdella aerobiophila*, were the only cultured organisms that lack the genes encoding TsaC/Sua5 proteins. In addition, we failed to detect Sua5/TsaC proteins in Undinarchaeota and Huberarchaea lineages from DPANN archaeal superphylum (Supplementary Figure 1) but it is unclear whether they are genuinely missing or were missed during metagenomic assemblies. Notably, in the entire tree we detected few species encoding both TsaC and Sua5: in the *Leotiomyceta* fungi clade (e.g., *Aspergillus bombycis* – Uniprot ID A0A1F7ZM40 and A0A1F8A8Q8) and in the *Vibrio* bacteria (e.g., *Vibrio aquaticus* – Uniprot A0A3S0QCR7 and A0A3S0V185). The data further show that out of 41 examined taxonomic groups depicted in Figure 2; Supplementary Figure 1, 10 groups contain species that are exclusively Sua5 users, 8 groups contain species that encode the TsaC orthologs only, 3 groups contain at least 98% of species that encode only the TsaC orthologs and the remaining 20 taxons contain species encoding Sua5 or TsaC. Intriguingly, however, the distribution of TsaC and Sua5 users across the tree shows no clear phyletic pattern suggesting a complex evolutionary history for the Sua5/TsaC family of proteins.

The phylogeny suggests a pre-LUCA origin of TsaC/Sua5 family

To gain insight into the evolutionary history of Sua5/TsaC family we constructed a phylogenetic tree using balanced taxonomic sampling across the three domains of life (Supplementary Tables 2, 3). The tree was inferred from the alignment of TsaC proteins and the catalytic TsaC-like domain of Sua5 proteins. It exhibits bipartite topology whereby TsaC proteins and TsaC-like domains robustly (bootstrap values 100%) segregated into two monophyletic clades (Figure 3), each clade containing sequences from the three domains of life. The branches in the TsaC part of the tree are in general longer suggesting that these proteins evolved at a higher rate. While in the Sua5 part of the tree no clear phyletic pattern can be observed, the TsaC tree shows that eukaryotic sequences cluster together and seem to be more related to archaeal TsaC sequences thus recapitulating the established phylogeny of organisms. Similarly, the tree built from the alignment of full-length Sua5 sequences (Supplementary Figure 2) revealed a clear phyletic pattern whereby the bacterial and archaeal sequences formed two distinct groups while the eukaryotic sequences were split into two

1 https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/batch/AlphaFold2_batch.ipynb

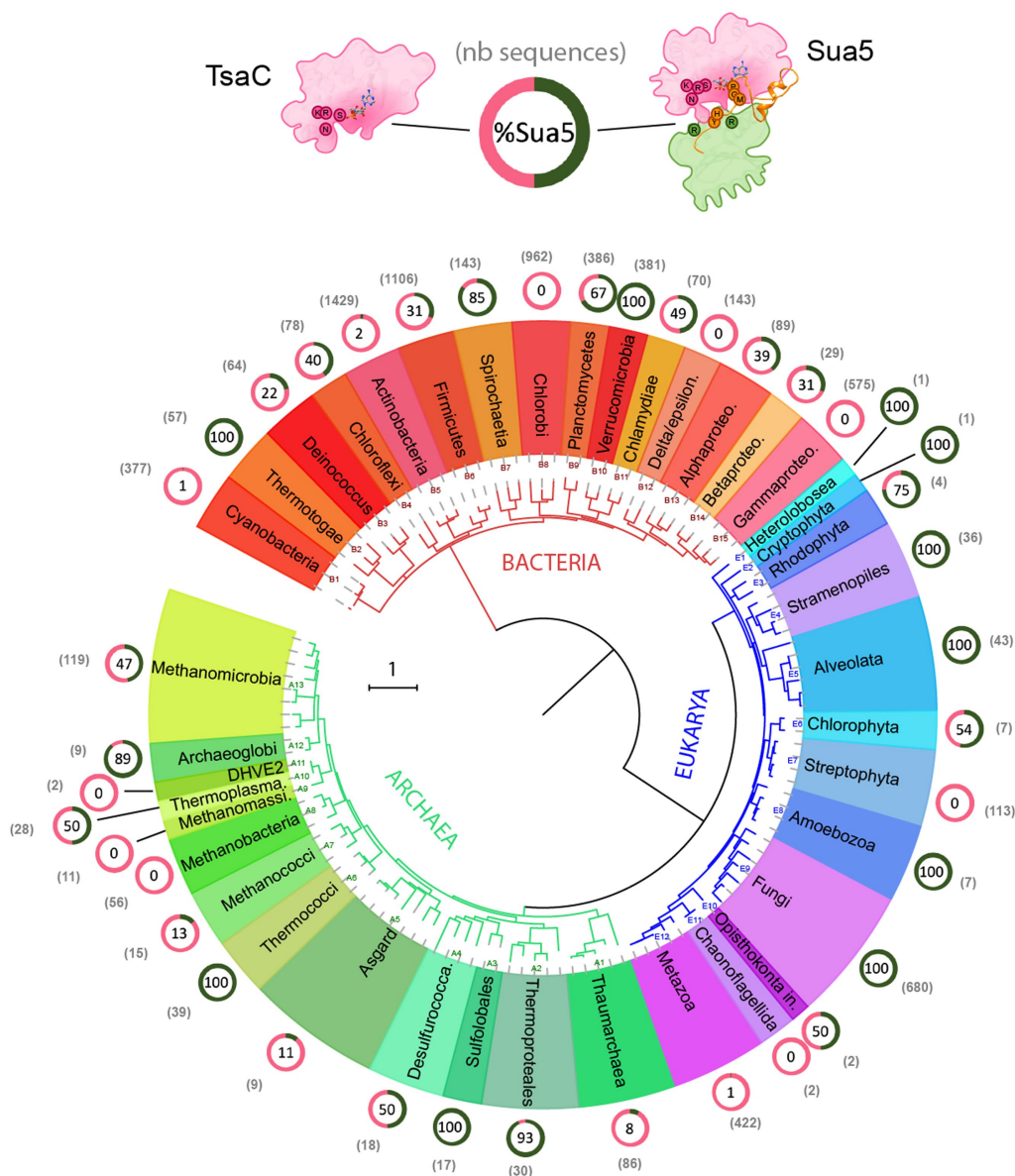


FIGURE 2

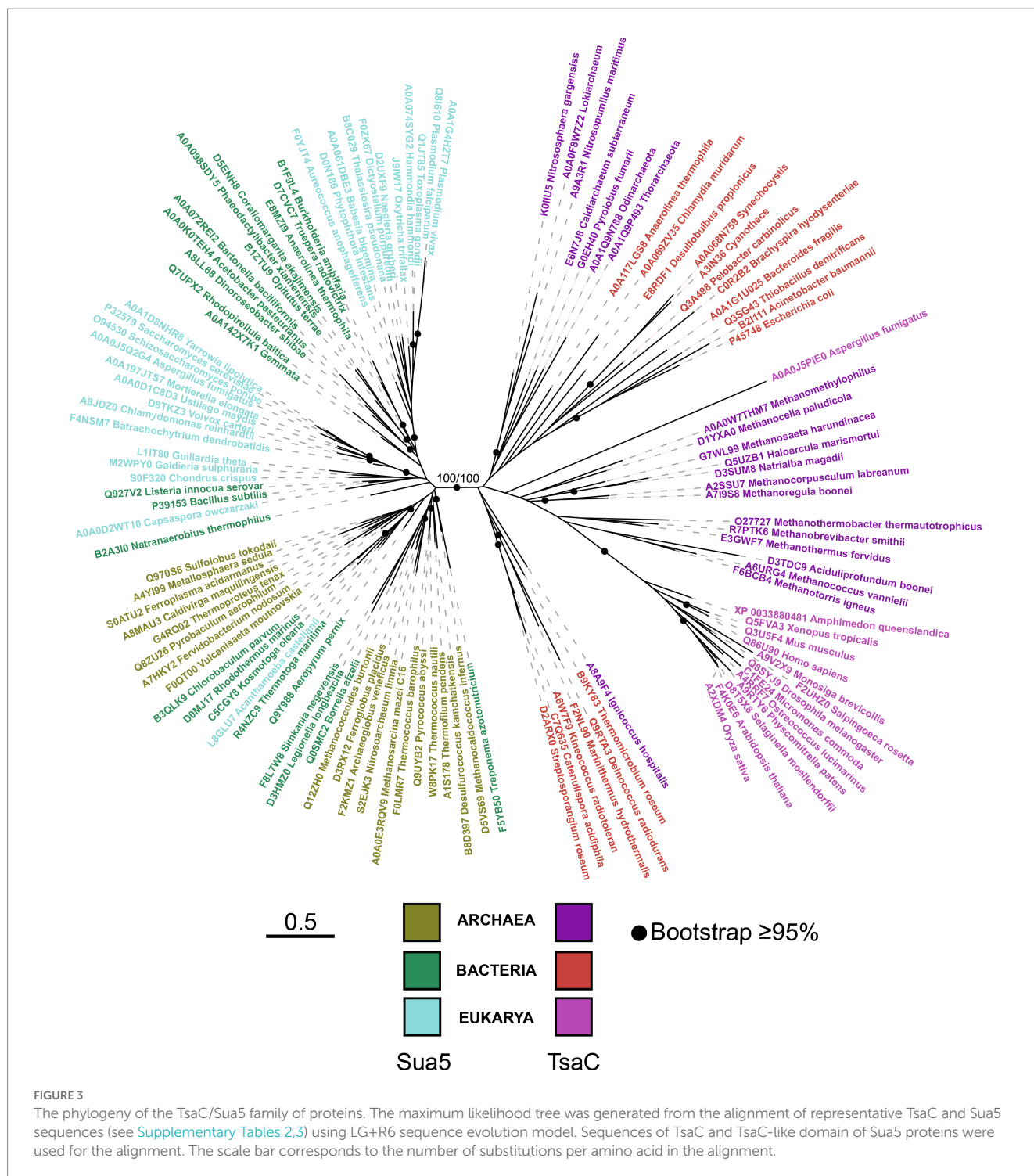
Distribution of TsaC and Sua5 proteins across the tree of life. The universal phylogeny based on RNA polymerase sequences is depicted. The number of TsaC and Sua5 sequences is indicated in the brackets for each taxon. The ring graph indicates the ratio of TsaC (pink) and Sua5 (green) orthologs for a given taxon. The numbers in the ring correspond to the percentage of Sua5 sequences. The scale bar corresponds to the number of substitutions per position in the alignment.

clades. The bipartite tree topology suggests that Sua5 and TsaC emerged from an ancient duplication event that occurred before LUCA, followed by either SUA5 domain acquisition or loss to yield the two variants.

Horizontal transfer of sua5 and tsaC genes occurs across a wide range of phylogenetic distances

The patchy distribution of Sua5 and TsaC proteins could, at least in part, be explained by horizontal gene transfers (HGT) of *sua5* and *tsaC* genes. To identify such events, we performed an initial screen using BLASTp searches whereby we focused on “mixed” taxonomic groups containing Sua5 users and TsaC users. For example, within

Chlorophyta phylum (eukaryotes, green algae, E6 in Figure 2), the Chlorophyceae family (e.g., *Volvox carteri f. nagariensis* – Uniprot D8TKZ3) is the only one carrying Sua5 orthologs (Supplementary Table 1). This protein is most similar to the Sua5 from *Apophysomyces ossiformis* (Uniprot A0A162WET7, 57% sequence identity) which is a Mucoromycota fungus. Similarly, Rhodophyta (E3), are Sua5 users except for one species (*Rhodospirillum rubrum* – Uniprot A0A7S0G1B8) that encodes TsaC. This protein is most similar (49% identity) to the TsaC from Firmicutes bacteria (*Brochothrix campestris*). Leotiomyces fungi such as *Aspergillus* are a peculiar case because several encode both Sua5 and TsaC proteins with a complete KxR/SxN motif suggesting that both orthologs are functional. However, the latter protein shows highest sequence similarity with Actinobacteria (*Amycolatopsis bartoniae*, 47.5% identity) and



Proteobacteria (*Verticillium sediminum*, 41.9% identity) suggesting that this may be a case of inter-domain HGT. This initial screen identified several other candidate proteins in Bacteria and Archaea. Among those, we observed that all *Methanothermococcus* species encoded TsaC ortholog except *M. thermolithotrophicus* which carries a Sua5 ortholog (WP_018153339.1). The sequence of this ortholog is identical to Sua5 from *Methanococcus maripaludis* KA1 and the gene is encoded on a ~14 kbp fragment that is 99% identical to *M. maripaludis* KA1 genome ([Supplementary Figure 3](#)). This suggests a recent acquisition of the whole fragment by horizontal gene transfer.

We next tested the robustness of the BLASTp results using phylogenetic analysis. As an example, we selected outlier Sua5 proteins from groups which are predominantly TsaC-users (the Methanococci archaea (A7), Thaumarchaea (A1) and Thermotogae bacteria (B2); [Supplementary Table 1](#)) and aligned those with representative Sua5 sequences from Bacteria and Archaea. The tree was resolved into two distinct clades containing bacterial or archaeal sequences allowing to identify the closest homologs for the outlier Sua5 proteins ([Figure 4A](#)). The tree topology showed that (i) the Sua5 proteins from *Methanococcus* species and Thaumarchaeota branched within Euryarchaeota; (ii) the

Sua5 protein from *Methanocaldococcus* archaeon branched within bacterial Sua5 orthologs; (iii) Sua5 proteins from Thermotoga bacteria branched within Crenarchaeal Sua5 orthologs. This suggested that the outlier Sua5 proteins were acquired by HGT from closely related (*Methanococcus*) or distantly related (Thaumarchaeota, *Methanocaldococcus*, and Thermotoga) organisms (Figure 4B).

Together, these analyses show that TsaC and Sua5 encoding genes have been submitted to horizontal transfers across different phylogenetic distances including at the highest taxonomic level.

The identification of variant-specific and conserved residues indicates distinct modes of substrate recognition in TsaC and Sua5 proteins

The segregation of TsaC and TsaC-like sequences into two distinct phylogenetic clades indicated that each variant must contain a specific set of residues in addition to the residues conserved across the whole TsaC/Sua5 family. In search for such residues, we aligned TsaC or Sua5 sequences representative for each domain of life (Supplementary Figure 4). This showed that almost half out of approximately 200 positions in the alignment are similar, in line with a strong conservation of structure and function of the TsaC domain across the tree of life. In addition, we identified 15 conserved residues specific for Sua5 proteins (Supplementary Figure 4), 14 of which are variable among TsaC proteins. Notably, three of them, Pro⁵⁹, Asn⁶² and His⁶⁷ (*Pa*-Sua5 numbering) are found in the loop that forms the deep part of the catalytic cavity surface, with their side chains pointing towards the cavity (Figures 5A,C). His⁶⁷ side chain forms a H-bond (2.7 Å) with the hydroxyl function of L-threonine in the structures of *St*-Sua5 (Supplementary Figure 5A). In the structure of *Pa*-Sua5 the side chain nitrogen of Asn⁶² acts as a donor to form a H-bond (2.4 Å) with PPI (Figure 5C) while its side chain oxygen acts as an acceptor to form a H-bond with the 3'-OH function of the ribosyl moiety of TC-AMP in the structure of *St*-Sua5 (2.6 Å; Supplementary Figure 5A). In several TsaC variants, Asn⁶² is replaced by a lysine (Supplementary Figure 4). However, unlike asparagine, lysine can only act as a H-bond donor suggesting their respective role in TC-AMP synthesis may not be identical.

Most of the other Sua5 specific conserved residues form two motifs at the inter-domain interface: the KPSPT motif is composed of (Lys/Arg)¹⁴⁹-Pro¹⁵⁰-Ser¹⁵¹-Pro¹⁵²-Thr¹⁵³ and the GVE motif is composed of Gly¹⁷⁸-Hyd¹⁷⁹-Glu¹⁸⁰ (where Hyd corresponds to the hydrophobic residue Val, Ile or Leu; Figure 5D). The motifs interact with one another via a H-bond (2.7 Å) formed by the side chains of Ser¹⁵¹ and Glu¹⁸⁰. The backbone chain of the KPSPT motif forms several H-bonds with *inter alia* the side chain of His¹⁵⁷, another Sua5-specific residue, and the conserved Arg³⁰¹ present in the SUA5 domain. In addition, the Sua5-specific amino acids Asp¹⁶¹ and Glu¹⁸⁰ are forming salt bridges with Arg³⁰¹ and Arg³²⁸, respectively (Figure 5D), thus likely playing a key role in the stabilization of the inter-domain interaction.

Remarkably, only one TsaC-specific conserved residue, Thr¹³⁸ (*Ec*-TsaC numbering), could be clearly identified (Supplementary Figure 4). This residue is found in the catalytic cavity of TsaC proteins and could potentially interact with substrate molecules (Figure 5B). To investigate this further we analyzed the structures of TsaC-like domains of TobZ and HypF proteins that co-crystallized with substrate molecules. The corresponding residue Thr⁵²⁹ in TsaC-like domain of

TobZ protein forms a H-bond (2.8 Å) with the phosphor-moiety of the carbamoyl-phosphate a precursor for the formation of carbamoyl-adenylate (Figure 5F). This phosphor-moiety occupies the same position as the phosphor-moiety of TC-AMP and as the β phosphate of AMPPNP in the structures of *St*-Sua5 (Supplementary Figure 5B). In the TsaC-like domain of HypF from *E. coli*, the equivalent residue Thr³²¹ interacts via a H-bond (2.6 Å) with the β phosphates of the bound ADP (Supplementary Figure 5C).

Overall, the comparative sequence and structural analyses identified variant-specific conserved residues with putative key roles in substrate binding thus highlighting previously unrecognized functional differences between TsaC and Sua5 proteins.

A single conserved residue constitutes a distinguishing feature between Sua5 and TsaC proteins

Structural and sequence analyses identified Thr¹³⁸ (*Ec*-TsaC numbering) as the only conserved residues specific for TsaC proteins. Intriguingly, the corresponding residue in Sua5 proteins is a conserved Proline (Pro¹⁴³ in *Pa*-Sua5; Supplementary Figure 4; Figure 5C) suggesting that this single and evolutionary conserved residue is a key determinant of specific mechanisms by which Sua5 and TsaC proteins catalyze the synthesis of TC-AMP. To test the relevance of this observation, we extracted 653 prokaryotic and 566 eukaryotic sequences from complete bacterial and archaeal genomes (one sequence per genus) and plotted the nature of the signature residue as a function of the size of the protein (Figure 6; Supplementary Figure 6). This showed that Sua5 proteins have exclusively Pro¹⁴³ at this position. In TsaC sequences, Thr¹³⁸ is predominant with up to 99% of occurrences in eukaryotes but is replaced by Ser¹³⁸ in 16% of bacterial sequences. The only exception to this rule are TsaC sequences from the bacterial genera *Brachyspira*, *Lactococcus*, *Erysipelothrix* as well as the entire *Aquificae* phylum where Thr¹³⁸ is replaced, intriguingly, by a Sua5-typical proline. These TsaC proteins lack the Sua5 specific motifs important for the interdomain interaction (Supplementary Figure 7) suggesting that these are genuine exceptions.

Together, these data confirm that the highly conserved Thr¹³⁸/Ser¹³⁸ and the corresponding Pro¹⁴³ accurately identify TsaC and Sua5 proteins, respectively, in the vast majority of cases.

Atypical Sua5 proteins of Archaeoglobi archaea carry functionally relevant deletions in the SUA5 domain

The identification of TsaC and Sua5 specific residues allowed us to spot atypical sequences such as those of Pro¹⁴³-containing bacterial TsaC proteins. In addition to these outliers, our attention was drawn to the Sua5 sequence from the archaeon *Archaeoglobus profundus* (*Ap*-Sua5) that carried Thr¹³⁸ instead of Pro¹⁴³ signature residue and was significantly shorter (289 AA) than a classical Sua5 protein (average size 331 AA). Extending the analysis to six other isolated Archaeoglobi species showed that, in addition to *Ap*-Sua5, the Sua5 proteins from *A. veneficus* (*Av*-Sua5) and *A. fulgidus* (*Af*-Sua5) carry deletions in the interdomain loop and in the SUA5 domain (Supplementary Figure 8). Around 50% of the interdomain loop residues are missing in these

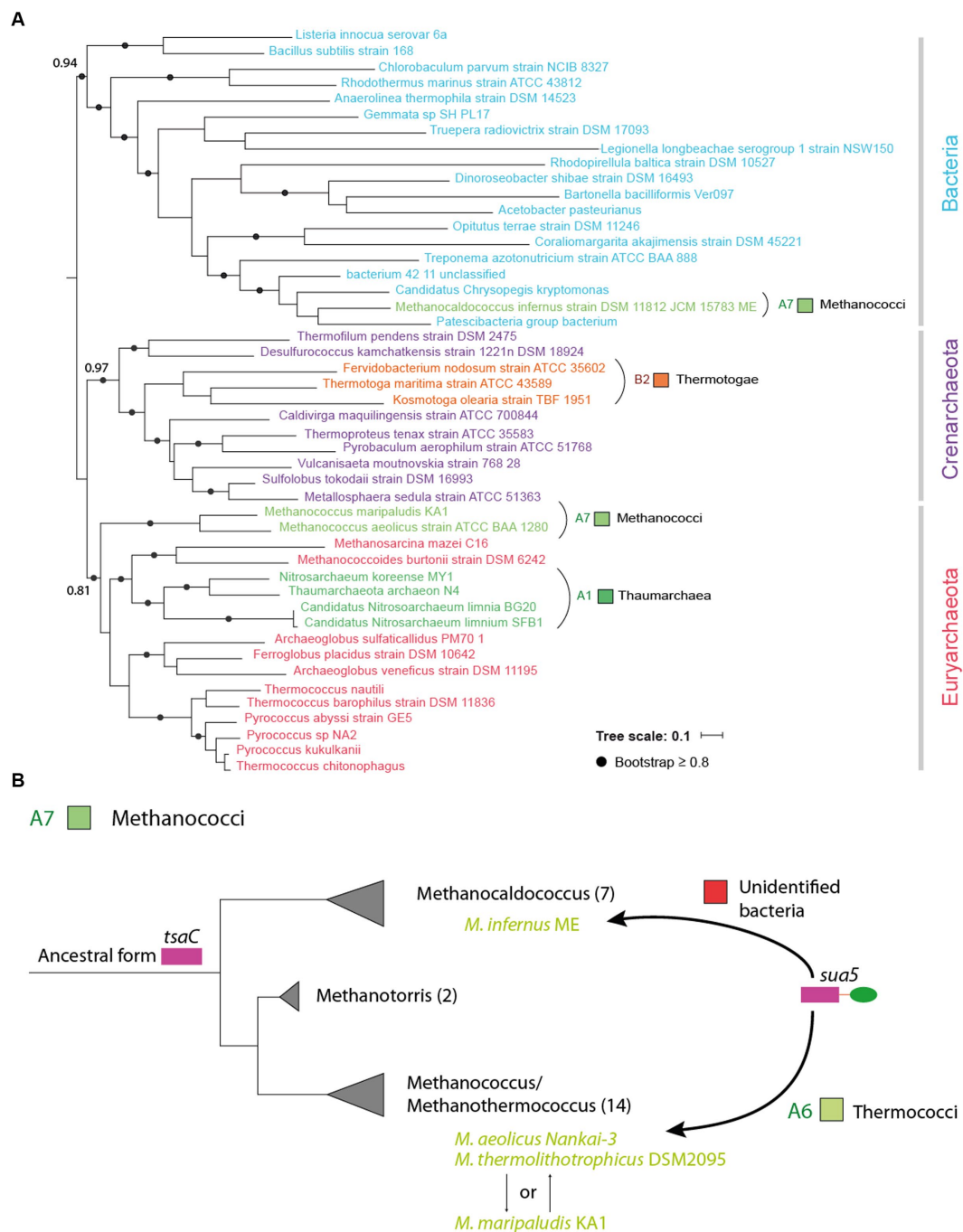


FIGURE 4

Horizontal transfer of *sua5* genes occurs across different phylogenetic distances. (A) Maximum likelihood phylogenetic tree of *Sua5* sequences from a wide range of bacterial taxa and from representative sequences of Crenarchaeota and Euryarchaeota. The tree was arbitrarily rooted between Bacteria and Archaea. The *Sua5* sequences from Methanococci, Thaumarchaea and Thermotogae are indicated in colors corresponding to their taxonomic group (see Supplementary Table 1). The scale bar corresponds to the number of substitutions per amino acid. (B) Evolutionary scenario for the acquisition of *sua5* genes in Methanococci archaea. The established phylogeny of the four main genera of Methanococci is shown. The number of species in each genus is indicated in brackets. All species carry *TsaC* orthologs except for the four species (indicated in green color) that carry *Sua5* orthologs. This suggests that the common ancestor of Methanococci was a *TsaC*-user and that the four species acquired the *sua5* gene by HGT. The putative donor of these *sua5* genes is indicated above the arrows.

proteins and among those the highly conserved and functionally important motif Pro²²⁸-Gly²²⁹-Met²³⁰. Moreover, when comparing with the archetypal *Sua5* from *A. sulfaticallidus* these three proteins lack 12%, (*Av-Sua5*), 13% (*Af-Sua5*), and 36% (*Ap-Sua5*) of residues in the SUA5 domain.

To get insight into the history of the acquisition of the detected mutations and their potential consequences for the function of these atypical *Sua5* proteins, we performed phylogenetic analyses and AlphaFold2 modelling. The sequences that accumulated the highest number of mutations clustered within a single clade (Figure 7A). Of note,

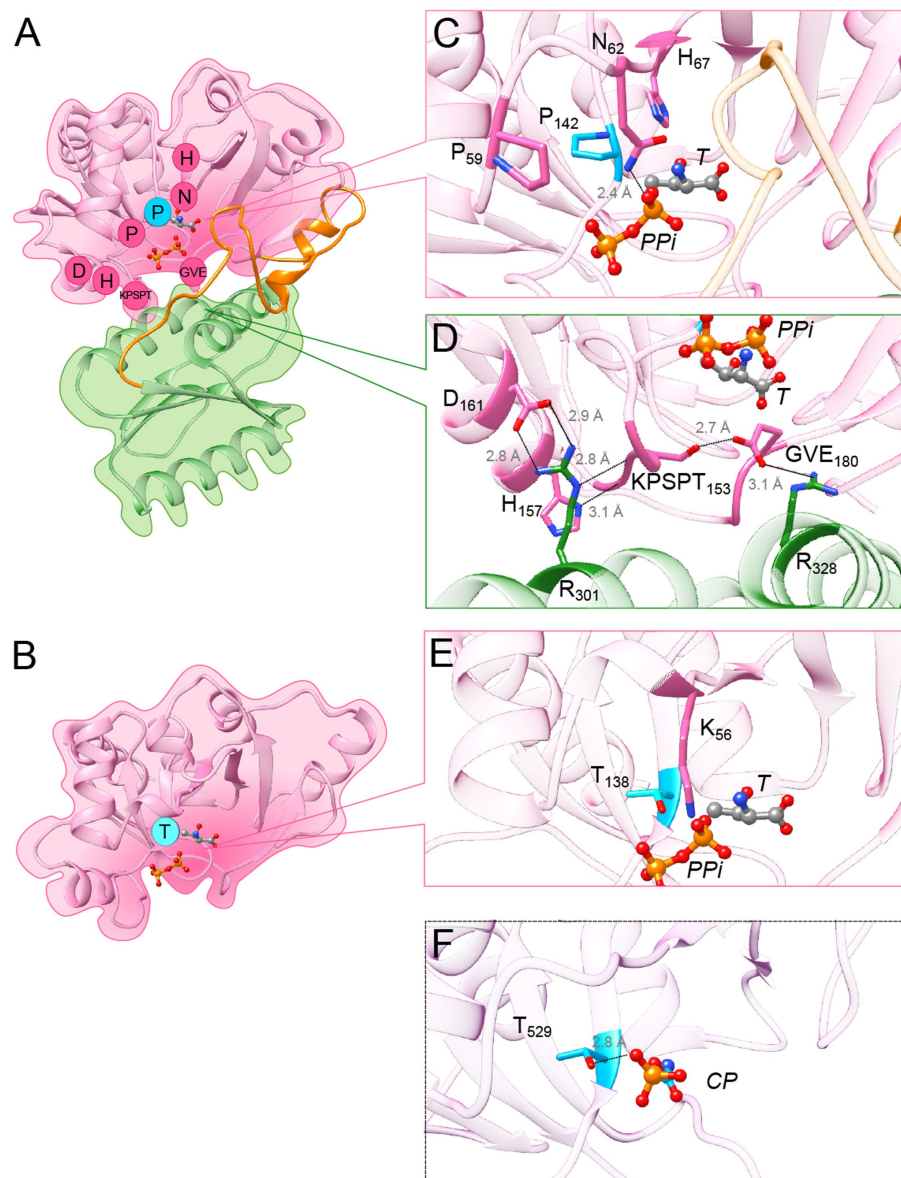


FIGURE 5

Variant-specific conserved residues interact with substrates and stabilize interdomain interface. **(A)** Cartoon showing the structure of Pa-Sua5 in complex with threonine and PPI shown as stick presentation. The TsaC-like domain is in pink, SUA5 domain is in green and the flexible loop is in orange. The Sua5-specific conserved residues are indicated with one letter code in pink circles or in cyan circle for the signature residue. **(B)** Cartoon showing the crystal structure of Ec-TsaC. Threonine and PPI molecules were modelled in the active site by superposing the Ec-TsaC with Pa-Sua5. The signature residue Thr is indicated with one code letter in a cyan-colored circle. **(C)** Zoom in the active site of Pa-Sua5. Sua5-specific conserved residues Pro⁵⁹, Asn⁶², His⁶⁷, and Pro¹⁴² are highlighted. Asn⁶² forms a hydrogen bond (indicated as dotted line) with the PPI molecule. **(D)** Zoom at the interface between the TsaC-like and SUA5 domains. The KPSPT motif and the GVE motif are shown as well as the network of H-bonds formed between these residues. Two salt bridges formed by Asp¹⁶¹ and Glu¹⁸⁰ with Arg³⁰¹ and Arg³²⁸, respectively, likely stabilize the two domains. **(E)** Zoom in the active site cavity of Ec-TsaC. Threonine and PPI molecules were modelled in the active site by superposing the Ec-TsaC and Pa-Sua5 that co-crystallized with these ligands. Side chain of the signature residue Thr¹³⁸ is shown as stick model. **(F)** Zoom in the active site of TsaC-like domain of TobZ protein from *Streptoalloteichus tenebrarius* bound to its substrate carbamoyl-phosphate (CP). The side chain of Thr⁵²⁹ (the equivalent of Thr¹³⁸ in Ec-TsaC) is indicated as stick model. The H-bond between Thr⁵²⁹ and the phosphate moiety is indicated by a dotted line.

they all carry a deletion in the interdomain loop leading to the loss of the PGM motif and, concomitantly, the substitution of Asn⁶² to Lys and His⁶⁷ to Gly/Val, the residues that we have tentatively identified as being involved in interaction with substrates (Figures 5C, 7A). We next modelled the structure of Ap-Sua5 and compared it to the crystal structure of Pa-Sua5. This showed that the TsaC-like domain adopts a typical fold whereby Lys and Val residues superpose well with Asn⁶² and His⁶⁷, respectively, suggesting that these mutations affect the binding of

substrates (Figure 7B). Despite a significant shortening, the SUA5 domain of Ap-Sua5 was modelled as a SUA5-like globular fold with the two arginines (Arg²⁵⁷ and Arg²⁷⁹ in Ap-Sua5) being correctly positioned to build salt bridges with the catalytic domain (Figure 7B). However, the interdomain loop seem to adopt a much more linear conformation leaving the entrance to the catalytic cavity open to the solvent (Figure 7C).

Together, the data suggest a progressive accumulation of mutations in interdomain loop and SUA5 domain of Archaeoglobi

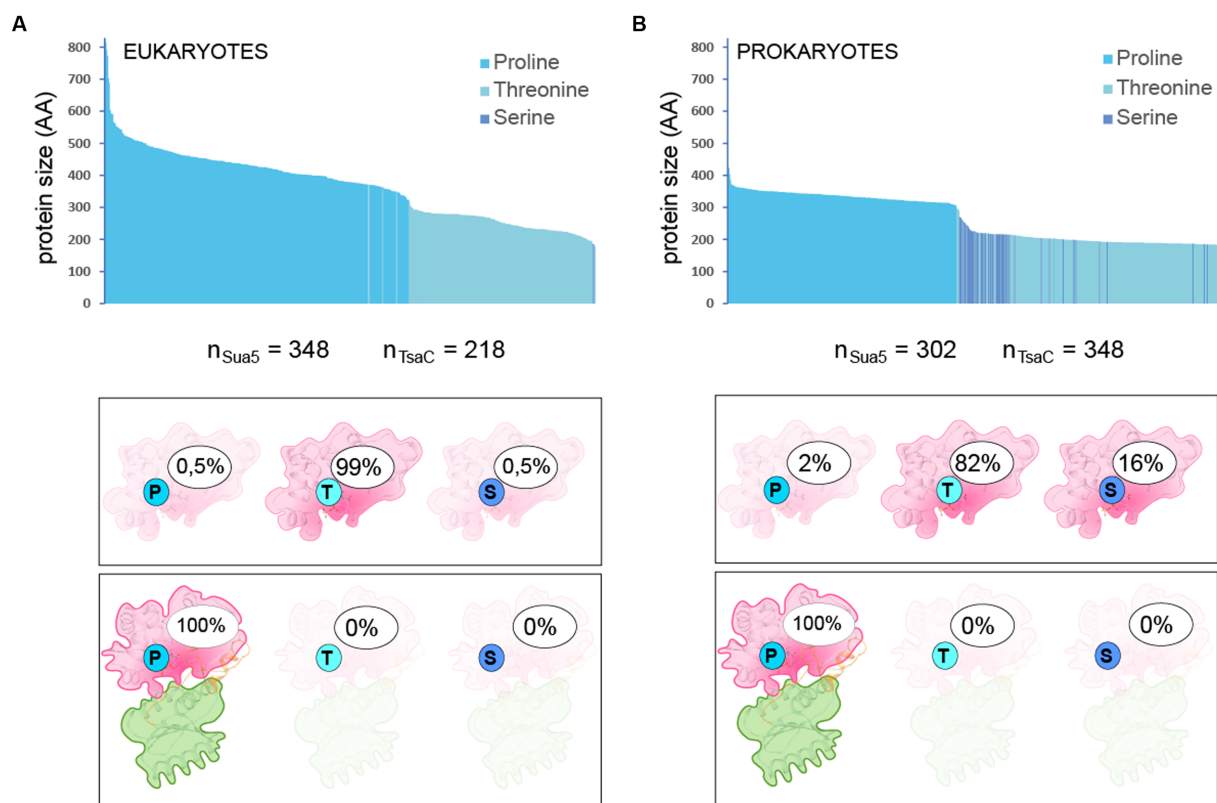


FIGURE 6

Nature and frequency of the signature residue in TsaC and Sua5 sequences Nature of the signature residue Pro¹⁴³/Thr¹³⁸/Ser¹³⁸ in TsaC and Sua5 sequences from Eukaryotes (A) and Prokaryotes (B). The top graph shows the protein length distribution whereby each sequence is represented as a vertical bar. The number of analyzed sequences is indicated under the graph. The legend gives the correspondence between the color of the bars and the presence of one of the signature residues. Bottom cartoons depict the percentage of TsaC and Sua5 having either Pro¹⁴³, Thr¹³⁸, or Ser¹³⁸ as the signature residue.

Sua5 with *Ap*-Sua5 being the extreme case. Consequently, these atypical Sua5 proteins seem to have acquired compensatory mutations in the catalytic domain that probably affect the binding to substrates and/or TC-AMP molecules.

Discussion

In the present work, we delineated the evolutionary pathway that led to the existence of two partially homologous proteins performing the same essential function in all extant organisms. We show that the evolution resulted in two different substrate binding modalities with potential consequences for the catalytic mechanism and/or efficiency of Sua5 and TsaC enzymes.

Early comparative genomics analyses identified TsaC/Sua5 family of proteins as part of a small set of about 60 truly ubiquitous proteins that were probably present in the Last Universal Common Ancestor (LUCA) of all extant organisms (Galperin, 2004). Our up to date distribution analysis shows that this is still true. The only fully sequenced organisms lacking *tsaC* or *sua5* genes are obligate ectosymbionts such as the DPANN archaea *N. equitans* and *N. aerobiophila*. These organisms carry greatly reduced genomes (0.49 Mbp and 0.67 Mbp, respectively) and depend on their hosts for the uptake of essential metabolites such as nucleotides, lipids and ATP (Huber et al., 2002; Waters et al., 2003; Kato et al., 2022). This suggests

that the absence of *tsaC* and *sua5* genes in a genome could be a marker of symbiotic or parasitic lifestyle. It remains to be investigated whether these organisms adapted to translate their genetic information faithfully in the absence of t⁶A-modified tRNA or if this function is supplied by the host. The latter may be the case for *N. equitans* which encodes the KEOPS complex (Waters et al., 2003) suggesting that it synthesizes t⁶A-tRNA but needs the host to supply TC-AMP or the TsaC/Sua5 enzyme by some yet unknown mechanism. The transport of TC-AMP from the host *Acidianus hospitalis* to *N. equitans* seems unlikely since this molecule is highly unstable at 90°C, the optimal growth temperatures of these organisms (Lauhon, 2012). The export of Sua5 enzyme from the cytoplasm to mitochondria via a signal sequence has, however, been reported, in *S. cerevisiae* (Thiaville et al., 2014a) suggesting that uptake of Sua5/TsaC by *N. equitans* could be a possible mechanism.

Our data further show that the concomitant presence of the *tsaC* and *sua5* genes in a genome is rare in accordance with previous work reporting the co-occurrence of these genes only in *Acetobacterium woodii* and *Vibrio cholerae* strains among 9,200 analyzed genomes (Thiaville et al., 2014b). In the additional cases we detected, we found that the second gene was always of exogenous origin suggesting that most of the co-occurrence cases are explained by recent horizontal gene transfers (HGT). These transfers can occur even over highest phylogenetic distances, between two different domains, suggesting that TsaC/Sua5 proteins are not dependent upon a particular cellular context and/or partners to function. This is further corroborated by

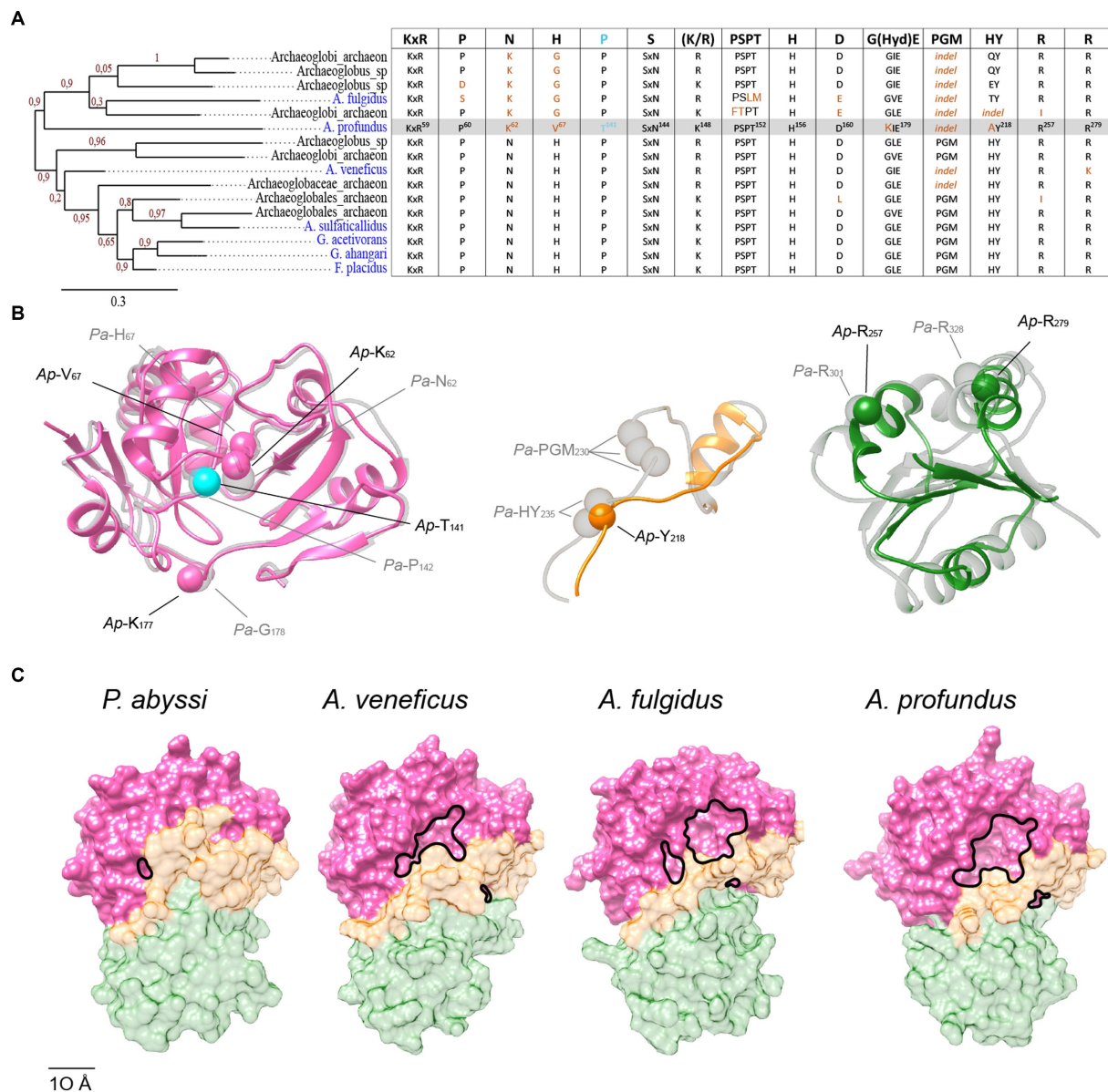


FIGURE 7

Progressive erosion of the SUA5 domain in Sua5 proteins from Archaeoglobi archaea. (A) Maximum likelihood tree of Sua5 proteins from Archaeoglobi species. The isolated species are highlighted in blue color. Bootstrap values for branch support are indicated. The table on the right shows the occurrence of the Sua5-specific conserved residues in the analyzed sequences. The canonical residues are indicated on the top of the table. The residues found in Ap-Sua5 are highlighted in grey and signature residue Pro¹⁴³/Thr¹³⁸ is indicated in cyan. Residues diverging from the consensus sequence and indels are in orange. (B) AlphaFold2 model of Ap-Sua5 (in color) is superimposed onto the crystal structure of Pa-Sua5 shown in light gray. The consensus Sua5-specific residues and the corresponding residues found in Ap-Sua5 are indicated as spheres. (C) Structure of Sua5 proteins shown as molecular surface. The crystal structure of Pa-Sua5 was retrieved from the PDB database while the structures of Av-Sua5, Af-Sua5, and Ap-Sua5 were modeled using AlphaFold2. TsaC-like domain, the interdomain loop, and SUA5 domain are depicted in pink, orange, and green color, respectively.

functional complementation experiments showing that *tsaC* and *sua5* genes originating from very distantly related organisms (yeast, bacteria and archaea) are all interchangeable *in vitro* and *in vivo* (El Yacoubi et al., 2009; Perrochia et al., 2013a). Such “independency” would probably facilitate horizontal gene transfers of *sua5* and *tsaC* genes and this may explain, at least in part, the patchy distribution of these genes across the universal tree.

Present day distribution of the TsaC/Sua5 family is difficult to reconcile with a simple evolutionary scenario. The bipartite tree topology of this family resembles that of translation elongation factors

EF-Tu and EF-G which are paralogs that duplicated before the divergence of all extant organismal lineages (Iwabe et al., 1989; Baldauf et al., 1996). Therefore, the phylogeny indicates that TsaC/Sua5 family emerged in a pre-LUCA ancestor and that at the time of LUCA both *tsaC* and *sua5* existed. However, several observations we made contradict this scenario. First, TsaC-like domain of Sua5 proteins contains 15 conserved residues (10% of the alignment positions used for tree construction) that are variable in the TsaC proteins. We therefore cannot exclude the possibility that the bipartite tree topology we observe is artificial and simply reflects the fact that

TsaC and TsaC-like sequences are “mechanically” segregated. Second, we identified several clear cases of HGT of *sua5* or *tsaC* genes whereby the endogenous gene was lost suggesting that the concomitant presence of the two genes carrying out the same function is not advantageous for an organism. The presence in many bacterial lineages of an inactive paralog of TsaC called YciO (Thiaville et al., 2014b) further supports this idea. Finally, the concomitant presence of both enzymes in LUCA would imply numerous independent gene loss events throughout the tree of life in order to explain the presence of only one of the two genes in extant organisms. In the light of these observations we therefore favour a more parsimonious scenario whereby LUCA encoded one of the two enzymes.

Which of the two variants is then the version that was present in LUCA? One possibility is the emergence of *tsaC* first and *sua5* post-LUCA by acquisition of an additional SUA5 domain. Domain fusion is a frequent method for new proteins to arise (Bornberg-Bauer et al., 2010), with evolution famously being described as a tinkerer rather than an inventor (Jacob, 1977). The SUA5 domain contains an atypical Rossmann fold which is an ancient and widely distributed protein fold (Ma et al., 2008) and therefore we cannot exclude a reassignment of this domain into a SUA5 domain. Of note, the presence of highly conserved motifs at the interface between the two domains supports the idea that all current Sua5 proteins originated from one common ancestral protein suggesting that the SUA5 domain acquisition, if it occurred, was a unique event.

The emergence of *tsaC* from *sua5* could have happened through the introduction of a premature stop codon. However, such drastic event may disrupt stability and/or activity of a protein (Weiner et al., 2006) and we observed this for truncated version of *Pa-Sua5* whereby the whole SUA5 domain was removed (Pichard-Kostuch et al., 2018). Rather, a progressive accumulation of deletions in SUA5 domain such as observed in atypical Archaeoglobi Sua5 could be a mechanism that initially led to emergence of TsaC. This process would need to be accompanied by mutations compensating for the loss of the loop and the SUA5 domain. Archaeoglobi Sua5 indeed presents non-neutral mutations of highly conserved Sua5-specific residues (Asn⁶²/Lys⁶² and His⁶⁷/Gly or Val⁶⁷). Given their position in the active site, we speculate that these mutations affect the substrates and/or product binding.

The gene erosion to yield shorter TsaC variant may have occurred several times in the course of evolution and each time a different compensatory set of mutations could have arisen thus explaining the absence of highly conserved TsaC-specific residues. This makes the TsaC-specific and conserved Thr/Ser¹³⁸, the sole exception to this rule, particularly interesting as it suggests that this residue is a key prerequisite for becoming the shorter version of the enzyme. The comparative structural analysis performed here suggests that Thr/Ser¹³⁸ contacts the alpha phosphate moiety of ATP/TC-AMP. Intriguingly, the corresponding Sua5 residue Pro143 would not be able to establish such a contact. However, the highly conserved histidine in the HY motif found in the interdomain loop was shown to be important for the activity of *Pa-Sua5*, likely by being involved in the binding of PPI/ATP (Pichard-Kostuch et al., 2018). Notably, none among the thousands of Sua5/TsaC sequences we screened in our analysis displayed both of these residues. It is therefore tempting to suggest that simultaneous presence of Thr/Ser¹³⁸ and HY motifs, both of which may interact with PPI/ATP, could have a negative impact on catalysis. Consequently, this combination was counterselected, resulting in the two variant-specific signature residues. Some evidence that this may be a plausible scenario comes from the atypical Sua5 protein of

Archaeoglobus profundus, the most advanced case of SUA5 domain erosion, where the Pro¹⁴³ signature residue was replaced by a threonine.

The Sua5 protein of *A. profundus* would be a good candidate to test whether a Sua5 protein could be evolved to become TsaC. Using directed mutagenesis it could be possible to generate even shorter active variants and ultimately remove the SUA5 domain completely. If our hypothesis is true than the complete loss of the SUA5 domain would only be possible in combination with further mutations of substrate binding residues to their TsaC-like counterparts. As a further line of experimental studies, it would be interesting both from the evolutionary and mechanistic standpoint, to compare the specific activities of TsaC and Sua5 proteins from closely related organisms.

In conclusion, we suggest that Sua5 was the ancestral version of the extant TC-AMP synthetic enzyme family. By articulating our previous experimental observations (Pichard-Kostuch et al., 2018) with our new in-depth sequence-structure analysis of this family, we suggest that the SUA5 domain is still essential for the activity of Sua5 variant by *inter alia* ensuring the right positioning of the catalytically important linker. However, key mutations in the TsaC-like domain can reduce this dependency and lead to the emergence of TsaC from Sua5 through SUA5 domain loss. Although it remains to be tested how these mutations affect the activity, this scenario combined with HGT events would account for the present day broad but inconsistent distribution of *tsaC* and *sua5* genes among different lineages. We suggest that SUA5 domain loss occurred at several independent occasions and that the resulting TsaC proteins adapted by evolving different interfaces for binding to substrates and/or products. Thus, TsaC proteins could be a more “advanced” version of the TC-AMP producing enzyme suggesting that, to quote Leonardo da Vinci, “simplicity is the ultimate sophistication” when it comes to the evolution of Sua5/TsaC family.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found at: <https://figshare.com/>, doi: 10.6084/m9.figshare.22283929.

Author contributions

APK, VDC and TBL conceived the study and all authors analysed the data. APK, VDC and TBL prepared the figures, wrote the draft and finalized the manuscript. All authors proofread the manuscript and approved the submitted version.

Funding

This work was funded by Agence Nationale de la Recherche, grant number ANR-18-CE11-0018 to TB.

Acknowledgments

The authors wish to thank Ryan Catchpole and Patrick Forterre for stimulating discussions during the early stages of this work.

Conflict of interest

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.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated

organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2023.1204045/full#supplementary-material>

References

- Agari, Y., Sato, S., Wakamatsu, T., Bessho, Y., Ebihara, A., Yokoyama, S., et al. (2008). X-ray crystal structure of a hypothetical Sua5 protein from *Sulfolobus tokodaii* strain 7. *Proteins Struct. Funct. Bioinform.* 70, 1108–1111. doi: 10.1002/prot.21794
- Arrondel, C., Missoury, S., Snoek, R., Patat, J., Menara, G., Collinet, B., et al. (2019). Defects in t(6a) tRNA modification due to GON7 and YRDC mutations lead to Galloway-Mowat syndrome. *Nat. Commun.* 10:3967. doi: 10.1038/s41467-019-11951-x
- Baldauf, S. L., Palmer, J. D., and Doolittle, W. F. (1996). The root of the universal tree and the origin of eukaryotes based on elongation factor phylogeny. *Proc. Natl. Acad. Sci. U.S.A.* 93, 7749–7754. doi: 10.1073/pnas.93.15.7749
- Beenstock, J., and Sicheri, F. (2021). The structural and functional workings of KEOPS. *Nucleic Acids Res.* 49, 10818–10834. doi: 10.1093/nar/gkab865
- Bornberg-Bauer, E., Huylmans, A.-K., and Sikosek, T. (2010). How do new proteins arise? *Curr. Opin. Struct. Biol.* 20, 390–396. doi: 10.1016/j.sbi.2010.02.005
- Braun, D. A., Rao, J., Mollet, G., Schapiro, D., Dageron, M.-C., Tan, W., et al. (2017). Mutations in KEOPS-complex genes cause nephrotic syndrome with primary microcephaly. *Nat. Genet.* 49, 1529–1538. doi: 10.1038/ng.3933
- Cantara, W. A., Crain, P. F., Rozenski, J., McCloskey, J. A., Harris, K. A., Zhang, X., et al. (2011). The RNA modification database, RNAMDB: 2011 update. *Nucleic Acids Res.* 39, D195–D201. doi: 10.1093/nar/gkq1028
- Criscuolo, A., and Gribaldo, S. (2010). BMGE (block mapping and gathering with entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. *BMC Evol. Biol.* 10:210. doi: 10.1186/1471-2148-10-210
- Da Cunha, V., Gaia, M., Gadelle, D., Nasir, A., and Forterre, P. (2017). Lokiarchaea are close relatives of Euryarchaeota, not bridging the gap between prokaryotes and eukaryotes. *PLoS Genet.* 13:e1006810. doi: 10.1371/journal.pgen.1006810
- Dereeper, A., Guignon, V., Blanc, G., Audic, S., Buffet, S., Chevenet, F., et al. (2008). *Phylogeny.fr*: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res.* 36, W465–W469. doi: 10.1093/nar/gkn180
- Edvardson, S., Prunetti, L., Arraf, A., Haas, D., Bacusmo, J. M., Hu, J. F., et al. (2017). tRNA N6-adenosine threonylcarbamoyltransferase defect due to KAE1/TCS3 (OSGEP) mutation manifest by neurodegeneration and renal tubulopathy. *Eur. J. Hum. Genet.* 25, 545–551. doi: 10.1038/ejhg.2017.30
- EL Yacoubi, B., Lyons, B., Cruz, Y., Reddy, R., Nordin, B., Agnelli, F., et al. (2009). The universal YrdC/Sua5 family is required for the formation of threonylcarbamoyladenine in tRNA. *Nucleic Acids Res.* 37, 2894–2909. doi: 10.1093/nar/gkp152
- EL Yacoubi, B., Hatin, I., Deutsch, C., Kahveci, T., Rousset, J.-P., Iwata-Reuyl, D., et al. (2011). A role for the universal Kae1/Qri7/YgdD (COG0533) family in tRNA modification: t⁶ a biosynthesis. *EMBO J.* 30, 882–893. doi: 10.1038/emboj.2010.363
- Finn, R. D., Clements, J., and Eddy, S. R. (2011). HMMER web server: interactive sequence similarity searching. *Nucleic Acids Research.* 39, W29–W37. doi: 10.1093/nar/gkr367
- Galperin, M. Y. (2004). “Conserved hypothetical” proteins: prioritization of targets for experimental study. *Nucleic Acids Res.* 32, 5452–5463. doi: 10.1093/nar/gkh885
- Harris, K. A., Bobay, B. G., Sarachan, K. L., Sims, A. F., Bilbille, Y., Deutsch, C., et al. (2015). NMR-based structural analysis of Threonylcarbamoyl-AMP synthase and its substrate interactions. *J. Biol. Chem.* 290, 20032–20043. doi: 10.1074/jbc.M114.631242
- Högenauer, G., Turnowsky, F., and Unger, F. M. (1972). Codon-anticodon interaction of methionine specific tRNAs. *Biochem. Biophys. Res. Commun.* 46, 2100–2106. doi: 10.1016/0006-291X(72)90765-6
- Huber, H., Hohn, M. J., Rachel, R., Fuchs, T., Wimmer, V. C., and Stetter, K. O. (2002). A new phylum of Archaea represented by a nanosized hyperthermophilic symbiont. *Nature* 417, 63–67. doi: 10.1038/417063a
- Ishikura, H., Yamada, Y., Murao, K., Saneyoshi, M., and Nishimura, S. (1969). The presence of N-[9-(β-D-ribofuranosyl)purin-6-ylcarbamoyl]threonine in serine, methionine and lysine transfer RNAs from *Escherichia coli*. *Biochem. Biophys. Res. Commun.* 37, 990–995. doi: 10.1016/0006-291X(69)90229-0
- Iwabe, N., Kuma, K., Hasegawa, M., Osawa, S., and Miyata, T. (1989). Evolutionary relationship of archaeobacteria, eubacteria, and eukaryotes inferred from phylogenetic trees of duplicated genes. *Proc. Natl. Acad. Sci. U.S.A.* 86, 9355–9359. doi: 10.1073/pnas.86.23.9355
- Jacob, F. (1977). Evolution and tinkering. *Science* 196, 1161–1166. doi: 10.1126/science.860134
- Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature* 596, 583–589. doi: 10.1038/s41586-021-03819-2
- Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., and Jermini, L. S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* 14, 587–589. doi: 10.1038/nmeth.4285
- Kato, S., Ogasawara, A., Itoh, T., Sakai, H. D., Shimizu, M., Yuki, M., et al. (2022). Nanobdella aerobiofilia gen. nov., sp. nov., a thermoacidophilic, obligate ectosymbiotic archaeon, and proposal of Nanobdellaceae fam. nov., Nanobdellales Ord. nov. and Nanobdellia class. nov. *Int. J. Syst. Evol. Microbiol.* 72, 005489. doi: 10.1099/ijsem.0.005489
- Koonin, E. V. (2003). Comparative genomics, minimal gene-sets and the last universal common ancestor. *Nat. Rev. Microbiol.* 1, 127–136. doi: 10.1038/nrmicro751
- Kurata, M., Kasai, T., Akasaka, R., Higashijima, K., Terada, T., Kigawa, T., et al. (2011). Crystal structure of *Sulfolobus tokodaii* Sua5 complexed with L-threonine and AMPNP. *Proteins* 79, 2065–2075. doi: 10.1002/prot.23026
- Laumon, C. T. (2012). Mechanism of N6-Threonylcarbamoyladenine (t⁶ a) biosynthesis: isolation and characterization of the intermediate Threonylcarbamoyl-AMP. *Biochemistry* 51, 8950–8963. doi: 10.1021/bi301233d
- Lemoine, F., Domelevo Entfellner, J.-B., Wilkinson, E., Correia, D., Dávila Felipe, M., De Oliveira, T., et al. (2018). Renewing Felsenstein's phylogenetic bootstrap in the era of big data. *Nature* 556, 452–456. doi: 10.1038/s41586-018-0043-0
- Llaser, J. L., Hussain, T., Saini, A. K., Nanda, J. S., Kaur, S., Gordiyenko, Y., et al. (2018). Translational initiation factor eIF5 replaces eIF1 on the 40S ribosomal subunit to promote start-codon recognition. *Life* 7:e39273. doi: 10.7554/eLife.39273
- Ma, B.-G., Chen, L., Ji, H.-F., Chen, Z.-H., Yang, F.-R., Wang, L., et al. (2008). Characters of very ancient proteins. *Biochem. Biophys. Res. Commun.* 366, 607–611. doi: 10.1016/j.bbrc.2007.12.014
- Machnicka, M. A., Olchowik, A., Grosjean, H., and Bujnicki, J. M. (2014). Distribution and frequencies of post-transcriptional modifications in tRNAs. *RNA Biol.* 11, 1619–1629. doi: 10.4161/15476286.2014.992273
- 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
- Moody, E. R., Mahendrarajah, T. A., Dombrowski, N., Clark, J. W., Petitjean, C., Offie, P., et al. (2022). An estimate of the deepest branches of the tree of life from ancient vertically evolving genes. *Life* 11:e66695. doi: 10.7554/eLife.66695
- Murphy, F. V., Ramakrishnan, V., Malkiewicz, A., and Agris, P. F. (2004). The role of modifications in codon discrimination by tRNA^{Lys}UUU. *Nat. Struct. Mol. Biol.* 11, 1186–1191. doi: 10.1038/nsmb861
- Naor, A., Thiaville, P. C., Altman-Price, N., Cohen-Or, I., Allers, T., de Crécy-Lagard, V., et al. (2012). A genetic investigation of the KEOPS complex in Halophilic Archaea. *PLoS One* 7:e43013. doi: 10.1371/journal.pone.0043013
- Nureki, O., Niimi, T., Muramatsu, T., Kanno, H., Kohno, T., Florentz, C., et al. (1994). Molecular recognition of the identity-determinant set of isoleucine transfer RNA from *Escherichia coli*. *J. Mol. Biol.* 236, 710–724. doi: 10.1006/jmbi.1994.1184
- Parthier, C., Görlich, S., Jaenecke, F., Breithaupt, C., Bräuer, U., Fandrich, U., et al. (2012). The O-Carbamoyltransferase TobZ catalyzes an ancient enzymatic reaction. *Angew. Chem. Int. Ed.* 51, 4046–4052. doi: 10.1002/anie.201108896
- Petrochia, L., Crozat, E., Hecker, A., Zhang, W., Bareille, J., Collinet, B., et al. (2013a). In vitro biosynthesis of a universal t6A tRNA modification in Archaea and Eukarya. *Nucleic Acids Res.* 41, 1953–1964. doi: 10.1093/nar/gks1287

- Perrochia, L., Guetta, D., Hecker, A., Forterre, P., and Basta, T. (2013b). Functional assignment of KEOPS/EKC complex subunits in the biosynthesis of the universal t(6)a tRNA modification. *Nucleic Acids Res.* 41, 9484–9499. doi: 10.1093/nar/gkt720
- Petkun, S., Shi, R., Li, Y., Asinas, A., Munger, C., Zhang, L., et al. (2011). Structure of Hydrogenase maturation protein HypF with reaction intermediates shows two active sites. *Structure* 19, 1773–1783. doi: 10.1016/j.str.2011.09.023
- Petersen, E. F., Goddard, T. D., Huang, C. C., Meng, E. C., Couch, G. S., Croll, T. I., et al. (2021). UCSF ChimeraX: structure visualization for researchers, educators, and developers. *Protein Sci.* 30, 70–82. doi: 10.1002/pro.3943
- Phelps, S. S., Malkiewicz, A., Agris, P. F., and Joseph, S. (2004). Modified nucleotides in tRNA^{Lys} and tRNA^{Val} are important for translocation. *J. Mol. Biol.* 338, 439–444. doi: 10.1016/j.jmb.2004.02.070
- Pichard-Kostuch, A., Zhang, W., Liger, D., Dageron, M.-C., Létouart, J., Li de la Sierra-Gallay, I., et al. (2018). Structure–function analysis of Sua5 protein reveals novel functional motifs required for the biosynthesis of the universal t6A tRNA modification. *RNA* 24, 926–938. doi: 10.1261/rna.066092.118
- Powers, D. M., and Peterkofsky, A. (1972). The presence of N-(Purin-6-ylcarbamoyl) threonine in transfer ribonucleic acid species whose codons begin with adenine. *J. Biol. Chem.* 247, 6394–6401. doi: 10.1016/S0021-9258(19)44706-6
- Robert, X., and Gouet, P. (2014). Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* 42, W320–W324. doi: 10.1093/nar/gku316
- Stuart, J. W., Gdaniec, Z., Guenther, R., Marszalek, M., Sochacka, E., Malkiewicz, A., et al. (2000). Functional anticodon architecture of human tRNA^{Lys} includes disruption of Intraloop hydrogen bonding by the naturally occurring amino acid modification, t⁶a¹. *Biochemistry* 39, 13396–13404. doi: 10.1021/bi0013039
- Teplova, M., Tereshko, V., Egli, M., Sanishvili, R., Joachimiak, A., Bushueva, T., et al. (2000). The structure of the yrdC gene product from *Escherichia coli* reveals a new fold and suggests a role in RNA binding. *Protein Sci.* 9, 2557–2566. doi: 10.1110/ps.9.12.2557
- Thiaville, P. C., el Yacoubi, B., Perrochia, L., Hecker, A., Prigent, M., Thiaville, J. J., et al. (2014a). Cross kingdom functional conservation of the Core universally conserved Threonylcarbamoyladenine tRNA synthetase enzymes. *Eukaryot. Cell* 13, 1222–1231. doi: 10.1128/EC.00147-14
- Thiaville, P. C., Iwata-Reuyl, D., and de Crécy-Lagard, V. (2014b). Diversity of the biosynthesis pathway for threonylcarbamoyladenine t(6)a, a universal modification of tRNA. *RNA Biol.* 11, 1529–1539. doi: 10.4161/15476286.2014.992277
- Thiaville, P. C., Legendre, R., Rojas-Benítez, D., Baudin-Baillieu, A., Hatin, I., Chalancon, G., et al. (2016). Global translational impacts of the loss of the tRNA modification t(6)a in yeast. *Microb Cell* 3, 29–45. doi: 10.15698/mic2016.01.473
- Trifinopoulos, J., Nguyen, L.-T., von Haeseler, A., and Minh, B. Q. (2016). W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Res.* 44, W232–W235. doi: 10.1093/nar/gkw256
- Väre, V. Y. P., Erusyal, E. R., Narendran, A., Sarachan, K. L., and Agris, P. F. (2017). Chemical and conformational diversity of modified nucleosides affects tRNA structure and function. *Biomol. Ther.* 7:29. doi: 10.3390/biom7010029
- Wan, L. C. K., Mao, D. Y. L., Neculai, D., Strecker, J., Chiovitti, D., Kurinov, I., et al. (2013). Reconstitution and characterization of eukaryotic N6-threonylcarbamoylation of tRNA using a minimal enzyme system. *Nucleic Acids Res.* 41, 6332–6346. doi: 10.1093/nar/gkt322
- Waters, E., Hohn, M. J., Ahel, I., Graham, D. E., Adams, M. D., Barnstead, M., et al. (2003). The genome of Nanoarchaeum equitans: insights into early archaeal evolution and derived parasitism. *Proc. Natl. Acad. Sci. U. S. A.* 100, 12984–12988. doi: 10.1073/pnas.1735403100
- Weiner, J., Beausart, F., and Bornberg-Bauer, E. (2006). Domain deletions and substitutions in the modular protein evolution. *FEBS J.* 273, 2037–2047. doi: 10.1111/j.1742-4658.2006.05220.x
- Weissenbach, J., and Grosjean, H. (1981). Effect of Threonylcarbamoyl modification (t6A) in yeast tRNA^{Arg}III on codon-anticodon and anticodon-anticodon interactions. A thermodynamic and kinetic evaluation. *Eur. J. Biochem.* 116, 207–213. doi: 10.1111/j.1432-1033.1981.tb05320.x
- Yamada, K. D., Tomii, K., and Katoh, K. (2016). Application of the MAFFT sequence alignment program to large data—reexamination of the usefulness of chained guide trees. *Bioinformatics* 32, 3246–3251. doi: 10.1093/bioinformatics/btw412



OPEN ACCESS

EDITED BY

Arthur Charles-Orszag,
University of California, San Francisco,
United States

REVIEWED BY

Horia Todor,
University of California, San Francisco,
United States
Iain G. Duggin,
University of Technology Sydney, Australia
Jörg Soppa,
Goethe University Frankfurt, Germany

*CORRESPONDENCE

Micaela Cerletti
✉ mcerletti@gmail.com
Alex Bisson
✉ bisson@brandeis.edu

[†]These authors have contributed equally to this work and share first authorship

RECEIVED 12 April 2023

ACCEPTED 27 July 2023

PUBLISHED 10 August 2023

CITATION

Rados T, Andre K, Cerletti M and
Bisson A (2023) A sweet new set of inducible
and constitutive promoters in *Haloferax*
volcanii.
Front. Microbiol. 14:1204876.
doi: 10.3389/fmicb.2023.1204876

COPYRIGHT

© 2023 Rados, Andre, Cerletti and Bisson. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). 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 sweet new set of inducible and constitutive promoters in *Haloferax volcanii*

Theopi Rados^{1†}, Katherine Andre^{1†}, Micaela Cerletti^{1,2*} and Alex Bisson^{1*}

¹Department of Biology, Brandeis University, Waltham, MA, United States, ²Instituto de Investigaciones Biológicas, Facultad de Ciencias Exactas y Naturales, Universidad Nacional de Mar del Plata, Mar del Plata, Argentina

Inducible promoters are one of cellular and molecular biology's most important technical tools. The ability to deplete, replete, and overexpress genes on demand is the foundation of most functional studies. Here, we developed and characterized a new xylose-responsive promoter (Pxyl), the second inducible promoter system for the model haloarchaeon *Haloferax volcanii*. Generating RNA-seq datasets from cultures in the presence of four historically used inducers (arabinose, xylose, maltose, and IPTG), we mapped upregulated genomic regions primarily repressed in the absence of the above inducers. We found a highly upregulated promoter that controls the expression of the *xacEA* (*HVO_B0027-28*) operon in the pHV3 chromosome. To characterize this promoter region, we cloned msfGFP (monomeric superfold green fluorescent protein) under the control of two upstream regions into a modified pTA962 vector: the first 250 bp (P250) and the whole 750 bp intergenic fragments (P750). The P250 sequence drove the expression of msfGFP constitutively, and its expression did not respond to the presence or absence of xylose. However, the P750 promoter showed not only to be repressed in the absence of xylose but also expressed higher levels of msfGFP than the previously described inducible promoter PtnaA in the presence of the inducer. Finally, we validated the inducible Pxyl promoter by reproducing morphological phenotypes already described in the literature. By overexpressing the tubulin-like FtsZ1 and FtsZ2, we observed similar but slightly more pronounced morphological defects than the tryptophan-inducible promoter PtnaA. FtsZ1 overexpression created larger, deformed cells, whereas cells overexpressing FtsZ2 were smaller but mostly retained their shape. In summary, this work contributes a new xylose-inducible promoter that could be used simultaneously with the well-established PtnaA in functional studies in *H. volcanii* in the future.

KEYWORDS

archaea, haloarchaea, *Haloferax volcanii*, inducible promoters, constitutive promoters, xylose-inducible promoter

Introduction

Inducible promoters have been essential tools for molecular and cell biology studies in bacteria and eukaryotes, allowing for control over the expression of genes of interest in terms of both timing and strength of expression. Among the multiple uses of inducible promoters are dynamic studies of gene expression *in vivo*, timed expression of tagged protein fusions, and depletion assays. On the other hand, constitutive promoters have a wide range of applications,

from expressing selective marker genes to consistent overexpression of proteins.

Like most model organisms, the budding yeast *Saccharomyces cerevisiae* has a breadth of inducible promoters, most relying on ethanol and sugars (Weinhandl et al., 2014). These have later been replaced by variations of bacterial promoters due to tighter control of the latter. Additionally, there are two kinds of light-induced promoters used in mammalian cell studies, based on a light-oxygen-voltage (LOV) domain from the fungus *Neurospora crassa* and the *Arabidopsis* photoreceptor (Kallunki et al., 2019), which have been used in the research and identification of essential signaling pathways such as mTOR and Hippo (Kallunki et al., 2019). In plants, there is also a variety of available inducible promoters that can be used in *Arabidopsis thaliana* and other model systems, both chemically induced by compounds distinct from those used in mammalian systems (ethanol, dexamethasone, and β -estradiol) (Borghi, 2010) as well as the promoter of the *A. thaliana* heat-shock protein *HSP18.2* gene, which is a strong inducible promoter that is activated by heat shock (Takahashi and Komeda, 1989; Matsuhara et al., 2000). In 2020, a xylose-inducible promoter was first cloned and used to express heterologous proteins in the thermoacidophilic archaeon *Sulfolobus acidocaldarius* (van der Kolk et al., 2020). Currently, the only inducible promoter available in the model haloarchaeon *Haloferax volcanii* is the tryptophan-inducible PtnaA (Allers et al., 2010).

Proteins isolated from halophiles have been increasingly used in biotechnology, from cosmetics manufacturing to bioremediation (Haque et al., 2020). Methods for expressing and purifying halophilic proteins, which commonly misfold and aggregate when expressed in *Escherichia coli* (Allers, 2010) have been described in *Haloferax volcanii* (Large et al., 2007) using the PtnaA promoter. Overexpression of such proteins for purification could benefit from stronger inducible or strong constitutive promoters being available in *H. volcanii*. Likewise, as *H. volcanii* emerges as a well-studied archaeal model due to its relative ease of cultivation and established genetics (Pohlschroder and Schulze, 2019), the development of molecular biology tools that allow the expression of multiple genes simultaneously becomes essential.

Recently, Nußbaum and coworkers (Nußbaum et al., 2021) reported that experiments on the hierarchy of recruitment of SepF by FtsZ1 or FtsZ2 were not feasible due to a lack of a second inducible promoter in *H. volcanii*. Despite the considerable characterization of the metabolism and catabolism of sugars in *H. volcanii* (Johnsen et al., 2009, 2013, 2015), little is known about how these promoters behave under ectopic expression during live-cell imaging experiments. To address the needs of the *Haloferax* community, we have here characterized two versions of the same promoter region: a strong constitutive and a new xylose-inducible promoter to be used as a tool for genetic studies in *H. volcanii*.

Methods

Haloferax volcanii cultures

Cells were grown in 16×25 mm glass tubes with 3 mL of Hv-Cab, a semi-defined medium based on casamino acids that improve growth and cell shape maintenance (de Silva et al., 2021). Cultures were incubated at 42°C under constant agitation with the inducers D-xylose

TABLE 1 Plasmids used in this work.

Alias	Plasmid (Promoter)	Reference
pTA962	pTA962 (PtnaA)	Allers et al. (2010)
eAD8	pTA962::PfdX-msfGFP (PfdX)	This work
eTR8	pTA962::msfGFP (PtnaA)	This work
eTR34	pAL750 (Pxyl)	This work
eTR36	pAL750::msfGFP (Pxyl)	This work
eTR38	pAL750::ftsZ1 (Pxyl)	This work
eTR40	pAL750::ftsZ2 (Pxyl)	This work
eBL82	pTA962::ftsZ1 (PtnaA)	This work
eBL86	pTA962::ftsZ2 (PtnaA)	This work
eBL227	pAL250::msfGFP (P250)	This work
eTR42	pAL250 (P250)	This work

TABLE 2 *Haloferax volcanii* strains used in this work.

Alias	Genotype	Reference
DS2	Wild-type	Hartman et al. (2010)
H26	Δ pyrE2	Allers et al. (2004)
aAD9	pTA962::PfdX-msfGFP	This work
aBL296	Δ pyrE2 pAL250::msfGFP	This work
aTR24	Δ pyrE2 pTA962::msfGFP	This work
aTR94	Δ pyrE2 pAL750::msfGFP	This work
aTR102	Δ pyrE2 pAL750	This work
aTR103	Δ pyrE2 pAL750::ftsZ1	This work
aTR104	Δ pyrE2 pAL750::ftsZ2	This work
aTR105	Δ pyrE2 pTA962::ftsZ1	This work
aTR106	Δ pyrE2 pTA962::ftsZ2	This work

(Thermo Scientific Chemicals), L-arabinose (Thermo Scientific Chemicals), IPTG (Fisher Bioreagents), D-maltose (Fisher Bioreagents), or L-tryptophan (Thermo Scientific Chemicals) in the concentrations indicated. Plasmid and strain list can be found in Tables 1, 2.

Cloning and transformations

All oligos used to make our constructs can be found in Table 3. The eTR8 vector was constructed by Gibson assembly (Gibson et al., 2009) from two PCR fragments using the oligos oTR26 and oTR27 (msfGFP) and a linearized pTA962 digested with NdeI.

The eAD08 vector was constructed by Gibson assembly from two PCR fragments using the oligos oBL340 and oJM96 (to amplify the msfGFP fragment) and the linearized pTA962 digested with NcoI.

The pAL250::msfGFP and pAL750::msfGFP vectors were constructed by Gibson assembly from three PCR fragments using the oligos oBL169 and oBL345 (msfGFP, eTR8 as a template), oBL343 and oBL344 (the first 250 bp upstream to the *HVO_B0027* start codon using the *H. volcanii* strain DS2 gDNA as a template) or oBL343 and oBL345 (the first 750 bp upstream to the *HVO_B0027* start codon

TABLE 3 Oligos used in this work.

Alias	Sequence (5' → 3')
oBL97	AGGTGGCACTTTTCGG
oBL105	TGAGCAAAAGGCCAGC
oBL169	ATGCGAAAAGGGGAAGAATTGTTTAC
oBL340	GTGCTGCGTTCGCCATCTAGATCATTGTAAAGTTCATCCATTCCAT
oBL343	CTATAGGGCGAATTGGGTACACCCGCCGACTCGGCGT
oBL344	AATTCTTCCCCTTTTCGCATTGCAGTATCCTCATTACCAGC
oBL345	GCTCTAGAACTAGTGGATCCTCATTGTAAAGTTCATCCATTCCATGC
oBL365	ACGCCGAGTCGCGGGTCCGGTACCGGGTCGA
oJM96	CCGAAGTCTGCAGCCATGCGAAAAGGGGAAGAATT
oSB33	GCACATTTCCCCGAA
oTR26	TTCGCGGACCTATTGCGCATATGCGAAAAGGGGAAGAATTGTTTA
oTR27	ATCAAGCTTATCGATTTTCATTCATTGTAAAGTTCATCCATTCC
oTR150	GGATACTGCATATGAGGATCCACTAGTT
oTR151	TGCAGTATCCTCATTACC
oTR152	AGAACTAGTGGATCCTCATACTACTCGACGTAGTCGATGTCTT
oTR153	GCTGGTAATGAGGATACTGCAATGGACTCTATCGTCGGC
oTR154	TCTAGAACTAGTGGATCCTCATTACCGGATGACGT
oTR155	GGTAATGAGGATACTGCATATGCAGGATATCGTTCCG
oZC23	GCTGGCCTTTTGCTCACATGAGCTTCTTTGATTTCGAGC

using the *H. volcanii* strain DS2 gDNA as a template), and a linearized pTA962 previously digested with KpnI and BamHI.

The pAL750 vector was created by Gibson assembly from three PCR fragments using the pAL::msfGFP vector as a template and the oligos oZC23 and oTR150 (*pHV2 ori*), oTR151 and oSB33 (*pyrE2-Pxyl₇₅₀* region), and oBL97 and oBL105 (*E. coli oriC* and Amp^R cassette).

The pAL750::ftsZ1 and pAL750::ftsZ2 vectors were cloned by Gibson assembly of two fragments using the oligos oTR151 and oTR152 (*ftsZ1*) or oTR153 and oTR154 (*ftsZ2*) using the *H. volcanii* strain DS2 gDNA as a template, and a linear fragment of the pAL vector digested with NdeI.

All Gibson reactions were transformed into competent *E. coli* DH5α cells and clones confirmed by whole-plasmid sequencing (Plasmidsaurus). Plasmid preps were then transformed into *H. volcanii* using the method previously described (Dyall-Smith, 2009), with 0.5 M EDTA (Thermo Scientific Catalog #J15694-AE) and PEG600 (Sigma, catalog # 87333-250G-F).

Growth curves

Cells were grown to an OD₆₀₀ of ~0.5 and diluted to an OD₆₀₀ of 0.05. Then, 200 μL of culture was placed in a 96-well flat-bottom plate (Corning Inc.). All wells surrounding the plate's edge (A and H rows, 1 and 12 columns) were filled with 200 μL of water to prevent media evaporation. Growth curves of three biological triplicates were performed using an EPOCH 2 microplate spectrophotometer (Agilent) with constant orbital shaking at 42°C. Data points were collected every 30 min.

Microscopy and image analysis

Cells were grown in Hv-Cab (de Silva et al., 2021) or YPC medium (Allers et al., 2004) to mid-exponential or stationary OD₆₀₀ without or with inducers as indicated. Cultures were then concentrated 10-fold by centrifuging (3,000×g for 2–5 min), and a 3 μL droplet of culture was placed on a 60×24 mm coverslip and gently covered with a 1.5×1.5 cm 0.25% Hv-Cab agarose pad (SeaKem LE Agarose, Lonza). Staining with Ethidium Bromide (Invitrogen, catalog #15585–011) was performed by adding 3 μg/mL of the dye into the cell culture and incubating at 42°C for 5 min. Cells were then imaged at 42°C using a Nikon TI-2 Nikon Inverted Microscope within an Okolab H201 enclosure. Phase-contrast and GFP-fluorescence images were acquired with a Hamamatsu ORCA Flash 4.0 v3 sCMOS Camera (6.5 μm/pixel), a CFI PlanApo Lambda 100x DM Ph3 Objective, and a Lumencor Sola II Fluorescent LED (380–760 nm). Image analysis was performed using FIJI (Schindelin et al., 2012). Cells were segmented using the rolling ball background subtraction (value = 20) on phase-contrast images, followed by thresholding (default) and using the analyze particles function with a minimum particle size of 0.2 μm². The mask was then used to acquire shape descriptor data and applied to the GFP channel (after background subtraction) for fluorescence quantification.

RNA extraction and sequencing

All steps were performed at room temperature unless otherwise indicated. Cells were grown to an OD₆₀₀ of ~0.5 in the presence of xylose (10 mM), arabinose (10 mM), IPTG (1 mM), and maltose

(1 mM) and harvested by centrifugation (4,500xg for 10 min). RNA extraction was performed using 1 mL of TriZol (Invitrogen) per sample, followed by vigorous pipetting to lyse all cells in the sample. 200 μ L of chloroform was added to the sample mixture, and samples were vortexed for 90 s. Samples were then centrifuged at 12,000xg at 4°C for 15 min, and the supernatant was collected and mixed with 2 μ L of glycogen (Sigma Aldrich) and 400 μ L of isopropanol (Thermo Scientific Chemicals). Samples were incubated overnight at –20°C. Samples were centrifuged at 21,000xg, 4°C for 30 min and washed twice with 75% ethanol (Zaffagni et al., 2022). Samples were treated with DNase I (NEB) for 12 min at 37°C, followed by a second ethanol precipitation. Purified RNA was sent to SeqCenter LLC (Pittsburgh, United States, seqcenter.com) for ribosome depletion using *Haloferax*-specific probes (Supplementary Table S1) and sequencing. Results were mapped to the *H. volcanii* DS2 genome and analyzed using Geneious 2022.2. Transcripts per million separated by ORF and inducer can be found in Supplementary Table S2. The complete raw RNA-seq datasets presented in this study can be found in NCBI GEO online repositories: PRJNA953041 (no induction), PRJNA953037 (xylose), PRJNA953035 (IPTG), PRJNA953033 (Arabinose), PRJNA953034 (Maltose).

Results

RNA-seq screening to identify new sugar-responsive promoters

To find native inducible promoters, we investigated four different sugars frequently used as inducers in other microbial models: arabinose (Guzman et al., 1995), xylose (Kim et al., 1996), maltose (Ming-Ming et al., 2006), and the sugar analog IPTG (Dubendorff and Studier, 1991). To determine the highest sugar concentrations we could use in *H. volcanii* cultures without compromising growth rates, cell size, and morphology, we titrated each sugar from 1 mM to 100 mM. Based on concentrations used to induce promoters in bacteria, we expected that concentrations higher than 1 mM could yield slow-growing, smaller cells. Surprisingly, we could only observe this outcome from cultures above 10 mM of xylose (Figures 1A,B). Therefore, we focused on concentrations at 10 mM.

Next, we employed RNA-seq from mid-exponential cultures with or without each of the four inducers to map candidates for new inducible promoters. Using a set of specific probes for *H. volcanii* (Supplementary Table S1) (Pastor et al., 2022), the total ribosomal RNA detected in our samples was between 0.21 to 0.37% of the total number of reads from each sample. By comparing the relative fold change of transcripts per million, we obtained 58 genes from which mRNA levels were significantly upregulated (\log_2 fold change ≥ 1 and value of $p \leq 0.05$) for xylose, 42 genes upregulated by maltose, 0 for IPTG, and 45 genes upregulated for arabinose. We identified 15 downregulated genes (\log_2 fold change ≤ -1 and value of $p \leq 0.05$) for xylose, 44 genes for maltose, 3 for IPTG, and 10 for arabinose (Supplementary Table S2).

To create a shortlist of promoter candidates, we could further validate experimentally, we arbitrarily selected genes above 5-fold or higher ratio change upon the addition of the inducer (Figure 1C). Above this threshold, we assigned 1 promoter for maltose

(HVO_0562–64); 4 for xylose (HVO_B0027–29, HVO_B0030–32, HVO_B0035, and HVO_B0036–38); 3 for arabinose (HVO_B0027–29, HVO_B0030–32, and HVO_B0035; and none for IPTG (Figures 1C,D). In the first two cases, candidate gene expression was upregulated upon adding xylose or arabinose (regions 1 and 2, Figure 1D). In contrast, the genomic region comprising gene HVO_B106 did not pass our requirements; it called our attention for being arabinose-specific and not being significantly upregulated under xylose addition (Figure 1C, orange arrowhead). Likewise, genes HVO_0562–HVO_0566 are specifically upregulated under maltose induction (Figure 1C, green arrowhead), and three regions (HVO_A0173, HVO_B0032, and HVO_B0303) respond to IPTG (Figure 1C, pink arrowheads).

Despite observing an approximately 10-fold increase in mRNA levels upon inducer addition, we wanted to ensure these genes were tightly controlled by an inducible promoter with precise linear titration power and not already constitutively expressed at a high basal expression level. To understand the basal expression levels of genes, including regions 1 and 2, we plotted the raw transcription profile (Figure 1E, left panel). We compared it to the transcriptional levels of all other genes (Figure 1E, right panel). As expected, after ribosomal RNA depletion, the transcripts with higher relative values in our datasets were the ones mapping to the S-layer glycoprotein (*csg*, over 26,000 TPM - Transcripts Per Million - across samples) and the translational factor *EF1a* transcripts (above 10,000 TPM across samples). On the other hand, both promoter regions 1 and 2 transcriptional levels are placed in the low quartile range of our dataset in the uninduced sample (1.44 TPM and 23.65 TPM, respectively).

Altogether, genomic regions 1 and 2 are promising candidates for new xylose- and arabinose-inducible promoters for *H. volcanii*. For the context of this work, and based on the dynamic range suggested by our RNA-seq dataset above, we focused on the characterization of the promoter regulating the gene *xacE* (HVO_B0027), or genomic region 1, under the induction of xylose.

A *xacE* 5' long upstream region is required for *xacE* repression in the absence of xylose

To test the *xacE* promoter region, we sub-cloned the fluorescent protein msfGFP under the control of two different putative promoter regions: 250bp (P250) and 750bp (P750) upstream to *xacE* (Figure 2A). Plasmids were then transformed into *H. volcanii* H26 (Δ pyrE2), and selected on plates without uracil. As a negative control, we used the H26 strain transformed with the empty pAL750 vector (labeled from now on as wild type). As a comparison, we analyzed the expression of msfGFP under the control of the popular inducible promoter PtnaA, which activates upon tryptophan addition (Allers et al., 2010). As a benchmark, we also inserted msfGFP under the control of PfdX, a strong and constitutive promoter used to express the *pyrE2* gene in our vectors as a selective transformation marker (PfdX-msfGFP-pyrE2 operon).

Imaging of live cells by phase-contrast and epifluorescence microscopy showed cytoplasmic signal emitted by the msfGFP fluorescent protein from single cells (Figure 2B). Cells carrying empty plasmids (Figure 2B, first column) showed low auto-fluorescence at 488nm excitation compared to cells carrying vectors inducing

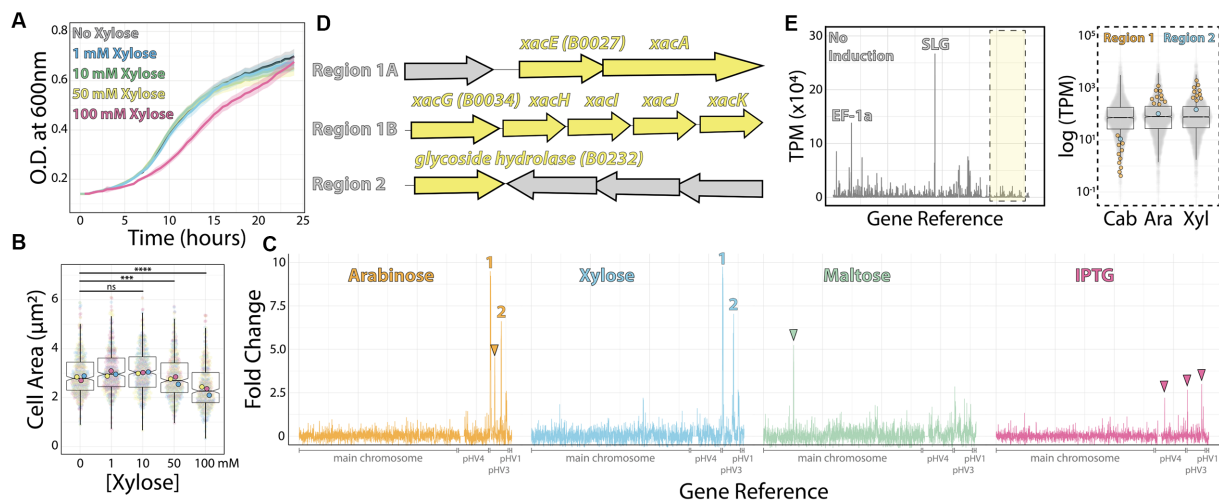


FIGURE 1

RNA-seq map of inducer-responsive promoters in *H. volcanii*. (A) Growth curves of *H. volcanii* DS2 cells under different xylose concentrations. (B) Cell area measurements at different xylose concentrations by phase contrast microscopy from mid-exponential cultures. Datapoints (and means) were colored (pink, yellow, and blue), indicating different biological replicates. (C) Gene expression ratios mapped across the *H. volcanii* genome from RNA-seq datasets of mid-exponential cultures with and without 10 mM arabinose (orange), xylose (blue), maltose (green), and IPTG (pink). Numbers 1 and 2 indicate genomic regions where gene expression increased above 5-fold. Arrowheads indicate promising genomic regions that did not satisfy our arbitrary 5-fold cutoff. (D) Locus organization of regions 1 and 2. (E) Expression map (transcripts per million) of genomic regions from RNA-seq datasets without inducers (left) and the relative increase in expression (right) from uninduced (Cab), arabinose (Ara), and xylose (Xyl).

msfGFP under the control of PfdX and Pxyl promoters. For a quantitative picture of the induction power of each promoter, we segmented individual cells and measured the mean fluorescence per cell from three biological triplicates and graphed using SuperPlots (Lord et al., 2020; Goedhart, 2021). Surprisingly, the P250 promoter was insensitive to xylose and constitutively expressed 2-fold above the PfdX control ($10,048 \pm 2,256$ and $4,967 \pm 929$ AU, respectively) (Figure 2C).

In contrast, the P750 promoter harboring the whole upstream intergenic sequence showed approximately a 5.5-fold repression ($1,827 \pm 848$) without induction compared to P250. However, msfGFP levels of non-induced P750 cells were still relatively high, 5.7-fold higher than uninduced PtnaA cells (320 ± 14) (Figures 2B,C). To minimize the transcriptional leakage from P750, we tested whether *H. volcanii* cells would present catabolite repression upon adding glucose. This strategy has been successful in various bacterial and yeast systems (Gancedo, 1998; Deutscher, 2008). Adding 20 mM glucose to cultures decreased leakage of the Pxyl promoter, but msfGFP intensity is still 2.7-fold higher than PtnaA cells (Figure 2D).

However, different from the P250 promoter, adding 10 mM xylose to P750 cells resulted in a 10.3-fold increase ($18,792 \pm 4,129$) in fluorescence intensity with a wider heterogeneity across the population compared to P250. Nevertheless, the unusual decrease in msfGFP expression observed between P250 and P750 suggests that *xacE*'s upstream region between 250 bp and 750 bp might have important regulatory elements.

We also inspected the dynamic range of our P750 promoter compared to PtnaA. By titrating xylose (0 to 25 μM) and tryptophan (0 to 2 mM), we observed a significant improvement from an 11.2-fold linear range for the PtnaA promoter to a 23-fold for the P750 promoter (Figure 2E). Providing the relatively higher leakage of the P750 promoter levels compared to PtnaA, we concluded that this new

construct is a valuable tool for titration experiments and overexpression at high protein levels for functional studies and protein purification directly from *H. volcanii*. Notably, P250 could be used simultaneously with the above inducible promoters as another constitutive promoter in the *Haloferax* community.

Furthermore, we tested the ability to induce the P750 promoter in cells growing in Hv-YPC, a rich medium based on yeast extract instead of casamino acids (Allers et al., 2004). In contrast with Hv-Cab, cells at mid-exponential growth phase showed lower levels of promoter leakage (222 ± 188), 3.6-fold higher than in Hv-Cab (Figure 3). However, together with the tighter expression of P750, induced cells in Hv-YPC expressed msfGFP at 7.6 times lower than in Hv-Cab ($1,636 \pm 1,258$), possibly due to catabolic repression in response to a component in yeast extract.

In addition to Hv-YPC, we also measured the expression profile of P750 at stationary growth phase and upon the addition of arabinose instead of xylose (Figure 3). While we observed a similar induction profile from cells in stationary phase and under arabinose induction, both cases showed a wider signal distribution ($9,851 \pm 617$ and $11,699 \pm 797$, respectively) compared to xylose-based induction during exponential growth ($12,395 \pm 800$).

Overexpression of the tubulins FtsZ1 and FtsZ2 confirms reported morphological phenotypes

Recently, Liao and colleagues reported the role of two tubulin paralogs (FtsZ1 and FtsZ2) in *H. volcanii*'s cell division (Liao et al., 2021). The authors used PtnaA-controlled overexpression of FtsZ1 and FtsZ2 independently and observed distinct, specific morphological phenotypes related to the overexpression of each paralog. Cells under PtnaA-FtsZ1 overexpression were slightly larger but significantly

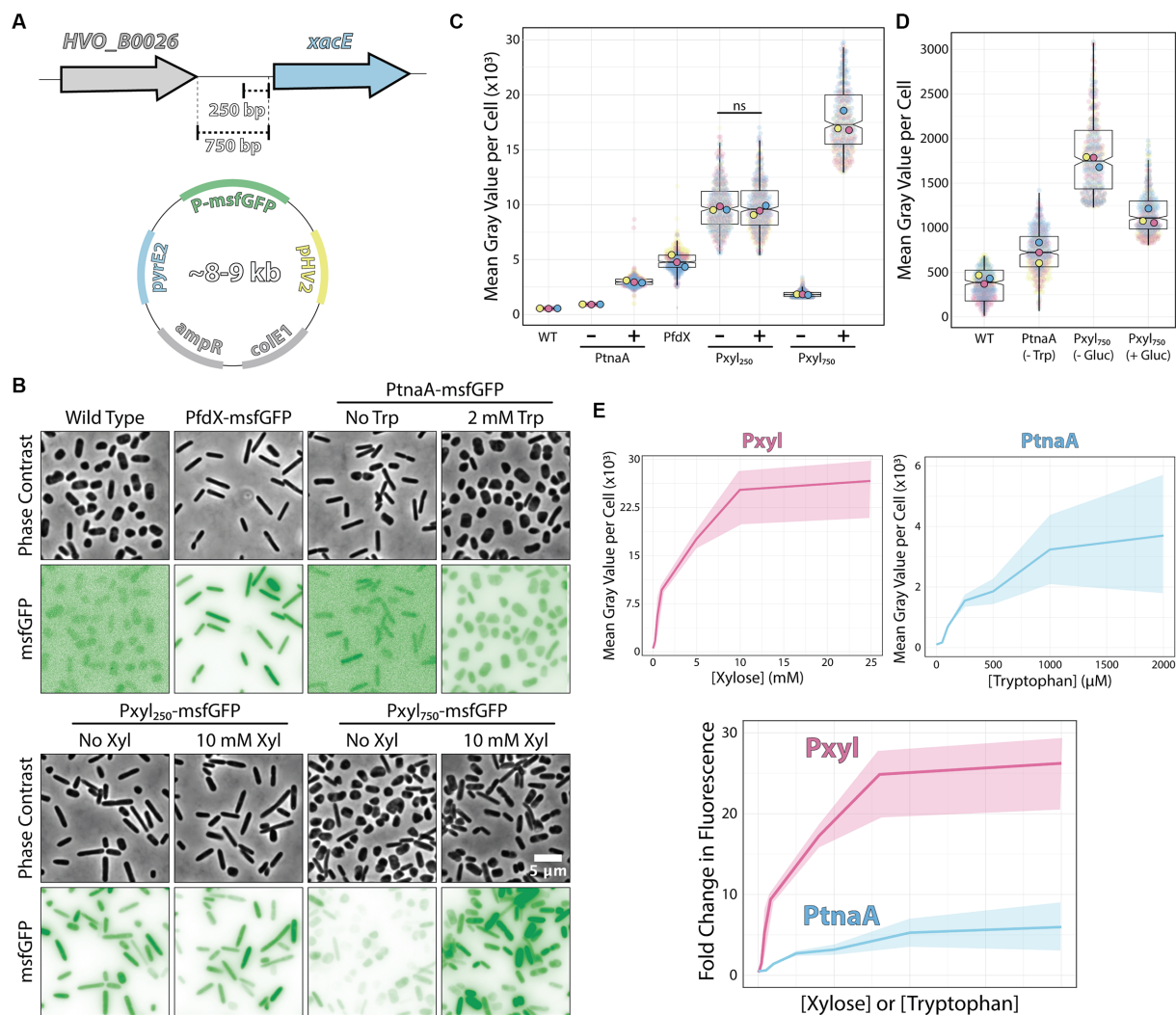


FIGURE 2

Pxyz constructs can be either strong inducible or constitutive promoters. (A) Fragments of the 5' intergenic region of *xacE* were used to clone into pAL vectors and map the P250 and P750 promoters tested in this work. (B) Phase-contrast and epifluorescence images of different constructs expressing the msfGFP fluorescent protein. (C) Mean msfGFP fluorescence measurements per cell from images shown in panel B. Each replicate means, and data points are independently labeled with different colors (pink, yellow, and blue). Datapoints (and means) were colored (pink, yellow, and blue), indicating different biological replicates. (D) Comparative background expression from constructs with and without glucose repression. Datapoints (and means) were colored (pink, yellow, and blue), indicating different biological replicates. (E) The dynamic range of PtnaA and Pxyz promoters across different inducer concentrations in raw numbers (top) and normalized (bottom) fold-changes. Shades indicate the 95% confidence interval from triplicate datasets.

misshaped compared to the control, whereas PtnaA-FtsZ2 cells looked significantly smaller but showed a more consistent morphology.

To confirm if those phenotypes are reproducible or even enhanced in our new Pxyz system, we cloned *ftsZ1* and *ftsZ2* in the pAL vector and analyzed the cell size and circularity compared to PtnaA-induced cells. Cells overexpressing *ftsZ2* under the Pxyz promoter are slightly smaller (1.2-fold decrease in average cell area) than cells overexpressing *ftsZ2* under the tryptophan-inducible PtnaA (Figure 4A). Meanwhile, cells overexpressing *ftsZ1* under the Pxyz promoter seemed to have more drastic phenotypes than those overexpressing *ftsZ1* using PtnaA, with deformed and enlarged cells (Figures 4A,C). Curiously, a fraction of the population exhibits narrow “cell bridges” connecting two enlarged cells (Figure 4B), similar to those previously described

(Rosenshine et al., 1989; Sivabalasarma et al., 2021; Von Kügelgen et al., 2021). This indicates that, in our samples, the cell-to-cell bridges observed in phase contrast could be formed by a product of incomplete cell division.

Discussion

Inducible promoters have been an invaluable resource in basic molecular biology research for the past 60 years, since the early days of the “PaJaMa Experiments” (Lewis, 2011). The addition of xylose to the culture medium was first shown to induce the expression of genes in *E. coli* (Batt et al., 1985) and *B. subtilis* in 1988 (Gärtner et al., 1988). This report describes the characterization of a new xylose-inducible

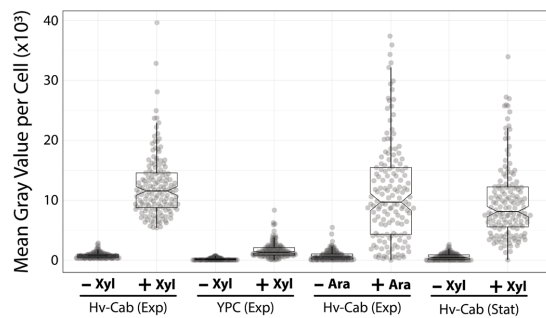


FIGURE 3

P750 shows different induction profiles in different media, at stationary phase, and upon adding arabinose. Epifluorescence microscopy quantification from exponential growing cells in Hv-Cab medium and Hv-YPC medium induced with 10 mM xylose (first and second plots). Expression of msfGFP was also recorded in cells grown to exponential phase in Hv-Cab medium with 10 mM arabinose (third plot). Cells grown in Hv-Cab to stationary phase and induced with 10 mM xylose (fourth plot).

promoter for the halophilic archaeon *H. volcanii*, a well-studied archaeal model (Pohlschroder and Schulze, 2019).

Using RNA-seq in cultures growing with and without four different inducers, we identified multiple promoter regions in which transcript levels were upregulated upon the addition of arabinose (total of 23 promoters, 3 above the arbitrary cut-off), xylose (total of 18 promoters, 4 above the arbitrary cut-off), maltose (total of 6 promoters, none above the arbitrary cut-off) and IPTG (total of 2 promoters, none above the arbitrary cut-off). Interestingly, our most promising promoter chosen to be characterized was already noted in past studies using DNA microarrays and C13 isotope tracking (Johnsen et al., 2009) and further characterized *in vitro* and *in vivo* (Johnsen et al., 2013, 2015). Interestingly, their measurements ranged from 100- to 400-fold increase upon arabinose addition compared to our observations from RNA-seq (9.8-fold, Figure 1C) and live-cell microscopy (10.3-fold increase, Figure 2C). This apparent discrepancy possibly originated from the differences in the experimental design, as we used rich undefined media and Johnson and colleagues grew the cells in synthetic medium with glucose. Yet on another note, against the anecdotal knowledge among researchers in the field, the PtnaA promoter showed a relatively low leakage in cultures without tryptophan (Figure 2D). This is a conundrum that shall be addressed by the community in the future.

One interesting previously unmentioned feature of the transcriptional regulation of the *xacEA* operon is that the P750 promoter is not only significantly repressed in the absence of xylose, but P750 shows a 1.9-fold increase in msfGFP signal in comparison to the constitutive P250 promoter (Figure 2C). The mechanistic details are still elusive, but it is possible that the extreme upstream fragment is the target of transcriptional factors competing to repress and activate the expression of *xacEA*. A good candidate for an activator is XacR (HVO_B0040), an IclR transcriptional factor family shown to work both as a repressor and activator in the same cells (Krell et al., 2006; Pan et al., 2011). In agreement with previous observations in bacteria, XacR in *H. volcanii* was shown to be required to activate *xacE* expression *in vivo* (Johnsen et al., 2015).

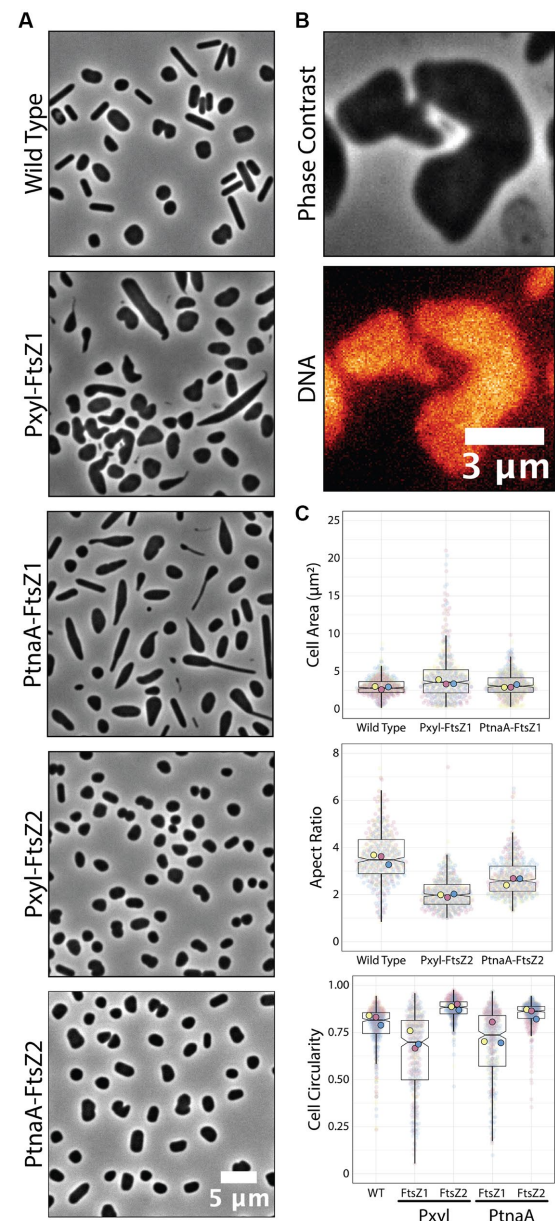


FIGURE 4

Overexpression of *ftsZ1* and *ftsZ2* under the Pxyl promoter. (A) Phase-contrast microscopy showing wild-type cells and cells overexpressing *ftsZ1* or *ftsZ2* under the Pxyl (5 mM xylose) or PtnaA (2 mM tryptophan) promoters. (B) Representative cell bridge phase-contrast and epifluorescence images. DNA was stained with ethidium bromide. (C) Cell area and circularity measurements from cells overexpressing *ftsZ1* and *ftsZ2* with the Pxyl (5 mM xylose) and the PtnaA (2 mM tryptophan) promoters. Datapoints (and means) were colored (pink, yellow, and blue), indicating different biological replicates.

We have shown that, through higher levels of expression than with the available promoter PtnaA, cells overexpressing the tubulin-like FtsZ1 had subtle but clear morphology defects (Figure 4) beyond those previously described (Liao et al., 2021). Interestingly, FtsZ2 overexpression under our Pxyl system did not result in a convincing change to PtnaA-FtsZ2 cells, in agreement with data suggesting that FtsZ2 proteins are more unstable compared to FtsZ1 (Liao et al., 2021). Alternatively, the FtsZ2 function could be coupled with other

factors that are more limited in the cell, and therefore a higher concentration of FtsZ2 would not linearly scale with cell size.

On the other hand, our work falls short of covering further characterization of the new Pxyl promoter. The region characterized as necessary for repression without xylose (750 bp) is notably longer than expected. Future experiments generating truncates and point mutations will be important to locate regulatory elements such as the B recognition element (BRE) and TATA box, but also create improved versions driving higher or more controlled expression. Likewise, the characterization of other promoter regions that showed to be independently upregulated upon the addition of maltose and IPTG (Figure 1C) may expand the toolbox of inducible promoters in *H. volcanii*.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found at: NCBI BioProject [<https://www.ncbi.nlm.nih.gov/bioproject/>], PRJNA953041 (no induction); PRJNA953037 (xylose); PRJNA953035 (IPTG); PRJNA953033 (Arabinose); PRJNA953034 (Maltose).

Author contributions

AB and TR conceived the study and wrote the first draft. AB, MC, TR, and KA designed the experiments. TR and KA performed the experiments and analyzed the data. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the Human Frontiers Science Program funding (RGY0074/2021) and Life Sciences-Moore-Simons Project on the Origin of the Eukaryotic Cell (doi: 10.46714/735929LPI) awarded to AB. AB is a Pew Scholar in the Biomedical Sciences,

supported by The Pew Charitable Trusts. MC was supported by the CONICET Partial Financing Program for Stays Abroad for Assistant Researchers.

Acknowledgments

The authors thank the Bisson Lab, Kiwi Shaw-Dodge, TB Manju, and Rosana De Castro (Universidad Nacional de Mar del Plata, Mar del Plata) for their insightful comments on our manuscript. The authors appreciate Sebastian Kadener and Sinead Nguyen (Brandeis University) for their help in optimizing the total RNA extraction from *H. volcanii*. The authors also thank Amy Schmid and Mar Martinez-Pastor (Duke University) for sharing data on their ribodepletion probes prior to publication.

Conflict of interest

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.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2023.1204876/full#supplementary-material>

References

- Allers, T. (2010). Overexpression and purification of halophilic proteins in *Haloferax volcanii*. *Bioeng. Bugs* 1, 290–292. doi: 10.4161/bbug.1.4.11794
- Allers, T., Barak, S., Liddell, S., Wardell, K., and Mevarech, M. (2010). Improved strains and plasmid vectors for conditional overexpression of his-tagged proteins in *Haloferax volcanii*. *Appl. Environ. Microbiol.* 76, 1759–1769. doi: 10.1128/AEM.02670-09
- Allers, T., Ngo, H.-P., Mevarech, M., and Lloyd, R. G. (2004). Development of additional selectable markers for the halophilic archaeon *Haloferax volcanii* based on the leuB and trpA genes. *Appl. Environ. Microbiol.* 70, 943–953. doi: 10.1128/aem.70.2.943-953.2004
- Batt, C. A., Bodis, M. S., Picataggio, S. K., Claps, M. C., Jamas, S., and Sinskey, A. J. (1985). Analysis of xylose operon regulation by mud (Apr, lac) fusion: trans effect of plasmid coded xylose operon. *Can. J. Microbiol.* 31, 930–933. doi: 10.1139/m85-174
- Borghli, L. (2010). “Inducible gene expression Systems for Plants” in *Plant developmental biology: Methods and protocols, methods in molecular biology*. eds. L. Hennig and C. Köhler (Totowa, NJ: Humana Press), 65–75.
- de Silva, R. T., Abdul-Halim, M. F., Pittrich, D. A., Brown, H. J., Pohlschroder, M., and Duggin, I. G. (2021). Improved growth and morphological plasticity of *Haloferax volcanii*. *Microbiology* 167:001012. doi: 10.1099/mic.0.001012
- Deutscher, J. (2008). The mechanisms of carbon catabolite repression in bacteria. *Curr. Opin. Microbiol.* 11, 87–93. doi: 10.1016/j.mib.2008.02.007
- Dubendorff, J. W., and Studier, F. W. (1991). Controlling basal expression in an inducible T7 expression system by blocking the target T7 promoter with lac repressor. *J. Mol. Biol.* 219, 45–59. doi: 10.1016/0022-2836(91)90856-2
- Dyall-Smith, M. (ed.) (2009). *The halo handbook: Protocols for halobacterial genetics*. Available at: https://haloarchaea.com/wp-content/uploads/2018/10/Halohandbook_2009_v7.3mids.pdf
- Gancedo, J. M. (1998). Yeast carbon catabolite repression. *Microbiol. Mol. Biol. Rev.* 62, 334–361. doi: 10.1128/MMBR.62.2.334-361.1998
- Gärtner, D., Geissendörfer, M., and Hillen, W. (1988). Expression of the *Bacillus subtilis* xyl operon is repressed at the level of transcription and is induced by xylose. *J. Bacteriol.* 170, 3102–3109. doi: 10.1128/jb.170.7.3102-3109.1988
- Gibson, D. G., Young, L., Chuang, R.-Y., Venter, J. C., Hutchison, C. A., and Smith, H. O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* 6, 343–345. doi: 10.1038/nmeth.1318
- Goedhart, J. (2021). SuperPlotsOfData—a web app for the transparent display and quantitative comparison of continuous data from different conditions. *Mol. Biol. Cell* 32, 470–474. doi: 10.1091/mbc.E20-09-0583

- Guzman, L. M., Belin, D., Carson, M. J., and Beckwith, J. (1995). Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *J. Bacteriol.* 177, 4121–4130. doi: 10.1128/jb.177.14.4121-4130.1995
- Haque, R. U., Paradisi, F., and Allers, T. (2020). *Haloferax volcanii* for biotechnology applications: challenges, current state and perspectives. *Appl. Microbiol. Biotechnol.* 104, 1371–1382. doi: 10.1007/s00253-019-10314-2
- Hartman, A. L., Norais, C., Badger, J. H., Delmas, S., Haldenby, S., Madupu, R., et al. (2010). The complete genome sequence of *Haloferax volcanii* DS2, a model archaeon. *PLoS One* 5:e9605. doi: 10.1371/journal.pone.0009605
- Johnsen, U., Dambeck, M., Zaiss, H., Fuhrer, T., Soppa, J., Sauer, U., et al. (2009). D-xylose degradation pathway in the halophilic archaeon *Haloferax volcanii*. *J. Biol. Chem.* 284, 27290–27303. doi: 10.1074/jbc.M109.003814
- Johnsen, U., Sutter, J.-M., Schulz, A.-C., Tästensen, J.-B., and Schönheit, P. (2015). XacR – a novel transcriptional regulator of D-xylose and L-arabinose catabolism in the haloarchaeon *Haloferax volcanii*. *Environ. Microbiol.* 17, 1663–1676. doi: 10.1111/1462-2920.12603
- Johnsen, U., Sutter, J.-M., Zaiß, H., and Schönheit, P. (2013). L-arabinose degradation pathway in the haloarchaeon *Haloferax volcanii* involves a novel type of L-arabinose dehydrogenase. *Extremophiles* 17, 897–909. doi: 10.1007/s00792-013-0572-2
- Kallunki, T., Barisic, M., Jäätelä, M., and Liu, B. (2019). How to choose the right inducible gene expression system for mammalian studies? *Cells* 8:796. doi: 10.3390/cells8080796
- Kim, L., Mogk, A., and Schumann, W. (1996). A xylose-inducible *Bacillus subtilis* integration vector and its application. *Gene* 181, 71–76. doi: 10.1016/s0378-1119(96)00466-0
- Krell, T., Molina-Henares, A. J., and Ramos, J. L. (2006). The IclR family of transcriptional activators and repressors can be defined by a single profile. *Protein Soc.* 15, 1207–1213. doi: 10.1110/ps.051857206
- Large, A., Stamme, C., Lange, C., Duan, Z., Allers, T., Soppa, J., et al. (2007). Characterization of a tightly controlled promoter of the halophilic archaeon *Haloferax volcanii* and its use in the analysis of the essential *ctt1* gene. *Mol. Microbiol.* 66, 1092–1106. doi: 10.1111/j.1365-2958.2007.05980.x
- Lewis, M. (2011). A tale of two repressors – a historical perspective. *J. Mol. Biol.* 409, 14–27. doi: 10.1016/j.jmb.2011.02.023
- Liao, Y., Ithurbide, S., Evenhuis, C., Löwe, J., and Duggin, I. G. (2021). Cell division in the archaeon *Haloferax volcanii* relies on two FtsZ proteins with distinct functions in division ring assembly and constriction. *Nat. Microbiol.* 6, 594–605. doi: 10.1038/s41564-021-00894-z
- Lord, S. J., Velle, K. B., Mullins, R. D., and Fritz-Laylin, L. K. (2020). SuperPlots: communicating reproducibility and variability in cell biology. *J. Cell Biol.* 219:e202001064. doi: 10.1083/jcb.202001064
- Matsuhara, S., Jingu, F., Takahashi, T., and Komeda, Y. (2000). Heat-shock tagging: a simple method for expression and isolation of plant genome DNA flanked by T-DNA insertions. *Plant J.* 22, 79–86. doi: 10.1046/j.1365-313x.2000.00716.x
- Ming-Ming, Y., Wei-Wei, Z., Xi-Feng, Z., and Pei-Lin, C. (2006). Construction and characterization of a novel maltose inducible expression vector in *Bacillus subtilis*. *Biotechnol. Lett.* 28, 1713–1718. doi: 10.1007/s10529-006-9146-z
- Nußbaum, P., Gerstner, M., Dingethal, M., Erb, C., and Albers, S.-V. (2021). The archaeal protein SepF is essential for cell division in *Haloferax volcanii*. *Nat. Commun.* 12:3469. doi: 10.1038/s41467-021-23686-9
- Pan, Y., Fiscus, V., Meng, W., Zheng, Z., Zhang, L.-H., Fuqua, C., et al. (2011). The *Agrobacterium tumefaciens* transcription factor BtCR is regulated via oligomerization. *J. Biol. Chem.* 286, 20431–20440. doi: 10.1074/jbc.M110.196154
- Pastor, M. M., Sakrikar, S., Rodriguez, D. N., and Schmid, A. K. (2022). Comparative analysis of rRNA removal methods for RNA-Seq differential expression in halophilic Archaea. *Biomol. Ther.* 12:682. doi: 10.3390/biom12050682
- Pohlschroder, M., and Schulze, S. (2019). *Haloferax volcanii*. *Trends Microbiol.* 27, 86–87. doi: 10.1016/j.tim.2018.10.004
- Rosenshine, I., Tchelet, R., and Mevarech, M. (1989). The mechanism of DNA transfer in the mating system of an archaeobacterium. *Science* 245, 1387–1389. doi: 10.1126/science.2818746
- Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., et al. (2012). Fiji: an open source platform for biological image analysis. *Nat. Methods* 9, 676–682. doi: 10.1038/nmeth.2019.FijiA., C
- Sivabalasarma, S., Wetzel, H., Nußbaum, P., van der Does, C., Beeby, M., and Albers, S.-V. (2021). Analysis of cell–cell bridges in *Haloferax volcanii* using Electron Cryo-tomography reveal a continuous cytoplasm and S-layer. *Front. Microbiol.* 11:612239. doi: 10.3389/fmicb.2020.612239
- Takahashi, T., and Komeda, Y. (1989). Characterization of two genes encoding small heat-shock proteins in *Arabidopsis thaliana*. *Mol. Gen. Genet. MGG* 219, 365–372. doi: 10.1007/BF00259608
- van der Kolk, N., Wagner, A., Wagner, M., Waßmer, B., Siebers, B., and Albers, S.-V. (2020). Identification of XylR, the activator of arabinose/xylose inducible regulon in *Sulfolobus acidocaldarius* and its application for homologous protein expression. *Front. Microbiol.* 11:1066. doi: 10.3389/fmicb.2020.01066
- Von Kügelgen, A., Alva, V., and Bharat, T. A. M. (2021). Complete atomic structure of a native archaeal cell surface. *Cell Rep.* 37:110052. doi: 10.1016/j.celrep.2021.110052
- Weinhandl, K., Winkler, M., Glieder, A., and Camattari, A. (2014). Carbon source dependent promoters in yeasts. *Microb. Cell Factories* 13:5. doi: 10.1186/1475-2859-13-5
- Zaffagni, M., Harris, J. M., Patop, I. L., Pamudurti, N. R., Nguyen, S., and Kadener, S. (2022). SARS-CoV-2 Nsp14 mediates the effects of viral infection on the host cell transcriptome. *elife* 11:e71945. doi: 10.7554/eLife.71945



OPEN ACCESS

EDITED BY

Solenne Ithurbide,
University of Freiburg, Germany

REVIEWED BY

Julian Quehenberger,
Vienna University of Technology, Austria
Changyi Zhang,
University of Illinois Urbana-Champaign,
United States
Maxime Tourte,
University of Freiburg, Germany

*CORRESPONDENCE

William D. Leavitt
✉ wil.leavitt@dartmouth.edu
Beverly K. Chiu
✉ beverly.k.chiu@dartmouth.edu

†PRESENT ADDRESS

Beverly K. Chiu,
C16 Biosciences, Inc., New York, NY,
United States

RECEIVED 09 May 2023

ACCEPTED 24 July 2023

PUBLISHED 15 August 2023

CITATION

Chiu BK, Waldbauer J, Elling FJ, Mete ÖZ,
Zhang L, Pearson A, Eggleston EM and
Leavitt WD (2023) Membrane lipid
and expression responses of *Saccharolobus*
islandicus REY15A to acid and cold stress.
Front. Microbiol. 14:1219779.
doi: 10.3389/fmicb.2023.1219779

COPYRIGHT

© 2023 Chiu, Waldbauer, Elling, Mete, Zhang,
Pearson, Eggleston and Leavitt. This is an
open-access article distributed under the terms
of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/)
(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.

Membrane lipid and expression responses of *Saccharolobus islandicus* REY15A to acid and cold stress

Beverly K. Chiu^{1*†}, Jacob Waldbauer², Felix J. Elling^{3,4},
Öykü Z. Mete^{1,3}, Lichun Zhang², Ann Pearson³,
Erin M. Eggleston⁵ and William D. Leavitt^{1,6*}

¹Department of Earth Sciences, Dartmouth College, Hanover, NH, United States, ²Department of the Geophysical Sciences, The University of Chicago, Chicago, IL, United States, ³Department of Earth and Planetary Sciences, Harvard University, Cambridge, MA, United States, ⁴Leibniz-Laboratory for Radiometric Dating and Isotope Research, Kiel University, Kiel, Germany, ⁵Department of Biology, Middlebury College, Middlebury, VT, United States, ⁶Department of Chemistry, Dartmouth College, Hanover, NH, United States

Archaea adjust the number of cyclopentane rings in their glycerol dibiphytanyl glycerol tetraether (GDGT) membrane lipids as a homeostatic response to environmental stressors such as temperature, pH, and energy availability shifts. However, archaeal expression patterns that correspond with changes in GDGT composition are less understood. Here we characterize the acid and cold stress responses of the thermoacidophilic crenarchaeon *Saccharolobus islandicus* REY15A using growth rates, core GDGT lipid profiles, transcriptomics and proteomics. We show that both stressors result in impaired growth, lower average GDGT cyclization, and differences in gene and protein expression. Transcription data revealed differential expression of the GDGT ring synthase *grsB* in response to both acid stress and cold stress. Although the GDGT ring synthase encoded by *grsB* forms highly cyclized GDGTs with ≥ 5 ring moieties, *S. islandicus* *grsB* upregulation under acidic pH conditions did not correspond with increased abundances of highly cyclized GDGTs. Our observations highlight the inability to predict GDGT changes from transcription data alone. Broader analysis of transcriptomic data revealed that *S. islandicus* differentially expresses many of the same transcripts in response to both acid and cold stress. These included upregulation of several biosynthetic pathways and downregulation of oxidative phosphorylation and motility. Transcript responses specific to either of the two stressors tested here included upregulation of genes related to proton pumping and molecular turnover in acid stress conditions and upregulation of transposases in cold stress conditions. Overall, our study provides a comprehensive understanding of the GDGT modifications and differential expression characteristic of the acid stress and cold stress responses in *S. islandicus*.

KEYWORDS

archaea, GDGT, cold stress, acid stress, *Saccharolobus islandicus*, thermoacidophile, stress response

1. Introduction

Many archaea produce membrane-spanning isoprenoid lipids called glycerol dibiphytanyl glycerol tetraethers (GDGTs) that are unique to the Archaeal domain of life (Pearson and Ingalls, 2013; Schouten et al., 2013; Villanueva et al., 2014). Individual GDGTs may contain 0–8 cyclopentyl ring moieties denoted as GDGT-0 through GDGT-8 (De Rosa et al., 1980a; De Rosa and Gambacorta, 1988). It has been proposed that archaea modify the rigidity and permeability of their membranes by modifying the numbers of cyclopentyl rings in their GDGTs, where GDGTs with a higher number of cyclopentyl rings can pack together more tightly to decrease membrane permeability and increase heat stability (Gliozzi et al., 1983; Gabriel and Chong, 2000; Shinoda et al., 2005). By adjusting membrane stability and permeability, archaea demonstrate homeoviscous adaptation in response to environmental stressors (Oger and Cario, 2013). Previous experiments showed that shifts in temperature, pH, or energy availability each caused thermoacidophilic archaea to alter the average cyclization of their membrane GDGTs (e.g., De Rosa et al., 1980a; Zillig et al., 1983; De Rosa and Gambacorta, 1988; Boyd et al., 2011; Jensen S. et al., 2015; Feyhl-Buska et al., 2016; Cobban et al., 2020; Quehenberger et al., 2020; Zhou et al., 2020). Despite these observations, the underlying expression systems leading to changes in GDGT cyclization as part of the archaeal stress response are poorly understood.

Few studies have examined how archaeal transcript and protein expression correlate with changes in membrane GDGT composition, due in part to the fact that enzymes for several key steps in the GDGT biosynthesis pathway were only recently discovered. These enzymes include a calditol synthase (Cds) that synthesizes calditol headgroups important for acidic pH conditions, a tetraether synthase (Tes) required for tetraether formation, and GDGT ring synthases (GrsAB) responsible for cyclopentyl ring formation (Zeng et al., 2018, 2019, 2022; Lloyd et al., 2022). GrsA and GrsB are homologous radical SAM enzymes that successively insert cyclopentyl rings at distinct positions on the GDGT core lipid (Zeng et al., 2019). GrsA introduces cyclopentyl rings at C-7 positions forming GDGTs 1 through 4, followed by GrsB activity at C-3 positions forming GDGTs 5 through 8 (Zeng et al., 2019). The protein abundance of GrsA and the transcript and protein abundance of GrsB showed pH and temperature dependence supporting a temporal expression model of the GDGT ring synthases that enables *Sulfolobus acidocaldarius* to modify its GDGT cyclization under environmental fluctuations (Yang et al., 2023). With the characterization of GrsA and GrsB, we can now examine how the expression of GDGT ring synthases corresponds with changes in GDGT cyclization of other archaea facing environmental stressors.

Here we performed parallel acid and cold stress experiments with *Saccharolobus islandicus* REY15A, hypothesizing that these stresses would induce changes in core GDGT cyclization, transcriptomes and proteomes. *S. islandicus* (formerly known as *Sulfolobus islandicus*) is a genetically tractable model thermoacidophilic archaeon in the order *Sulfolobales* (Zhang et al., 2018; Counts et al., 2021). We chose *S. islandicus* REY15A as our study organism because its genome has been described, and we expected its lipid profile to be responsive to environmental

stresses and growth phase based on a prior study with another strain of *S. islandicus* (Guo et al., 2011; Jensen S. et al., 2015). While *S. islandicus* produces both diether and tetraether lipids with various polar headgroups that can contribute to membrane adaptation, we focus our efforts on core GDGTs specifically (Oger and Cario, 2013; Jensen S. M. et al., 2015). No prior study to our knowledge has combined transcriptomics or proteomics with GDGT profiling to characterize the stress response of *S. islandicus* to changes in temperature or pH. The *S. islandicus* REY15A genome contains single copies of both *grsA* and *grsB*, which simplifies expression analysis of these genes and their encoded proteins in relation to changes in core GDGT cyclization. Overall, we characterize the physiological, proteomic, transcriptomic, and core lipid responses of *S. islandicus* to acid and cold stress, with targeted analysis of lipid biosynthesis genes. These observations inform a broader understanding of *Sulfolobales* cellular stress response to major environmental parameters in their hot and acidic natural habitats.

2. Materials and methods







2.1. Growth experiments


We performed experiments with *S. islandicus* strain REY15A under three conditions defined by different pH and temperature combinations (Figure 1). These included the optimal growth condition control (pH 3.4, 76°C), an acid stress condition (pH 2.4, 76°C), and a cold stress condition (pH 3.4, 66°C). All cultures were grown as batch cultures in 1-L Erlenmeyer flasks containing 500-ml DY medium (Zhang et al., 2013) with D-arabinose (Sigma-Aldrich, USA) as the carbon and energy source in place of dextrin, hereafter AY medium. Cultures were incubated in Innova-42 shaking incubators (Eppendorf, Hauppauge, NY, USA) at 150 RPM and 66 or 76°C. To minimize culture evaporation at these high incubation temperatures, open trays of water were maintained in the incubator to ensure a saturated atmosphere. Culture flasks had loose-fitting plastic caps to allow for continuous atmospheric oxygen exchange with the media. We monitored culture growth by optical density measurements at 600 nm (OD₆₀₀) on a Genesys 10S UV-Vis spectrophotometer (Thermo Scientific, Waltham, MA, USA). OD₆₀₀ data were used to calculate growth rates using a rate calculation script from Cobban et al. (2020) and available on GitLab.¹ Calculations of growth rates for each condition only incorporated OD₆₀₀ data from the triplicate cultures designated for early stationary phase harvest given that monitoring of mid-log-designated cultures ceased after harvest.

Experiments for each growth condition consisted of one uninoculated control and six inoculated replicates. Cultures were inoculated with actively growing pre-cultures to starting optical densities (OD₆₀₀) of 0.02. Triplicate cultures were harvested at timepoints corresponding to mid-log and early stationary phase of each growth condition (Figure 2). Harvest timepoints were based on preliminary growth experiments characterizing *S. islandicus* growth under each growth condition (data not shown). For

¹ https://git.dartmouth.edu/leavitt_lab/cobban-saci-lipids-batch-and-fed-batch-2020

Growth Experiment Overview

		Optimal	Acid stress	Cold stress
Growth conditions	Temperature	76°C	76°C	66°C
	Starting pH	3.4	2.4	3.4
Growth phase harvested	Mid-logarithmic			
	Early stationary			

 = 500-mL culture of *S. islandicus* REY15A

Analyses Overview

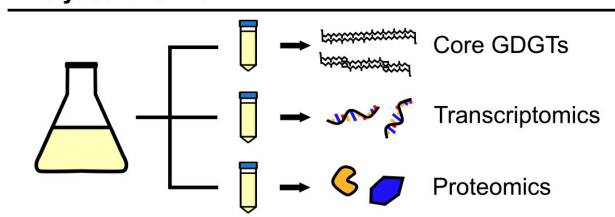


FIGURE 1

Overview of the growth experiments and downstream analyses, illustrating the conditions corresponding to the 18 samples of *S. islandicus* REY15A from which all lipid and expression data in this study were derived.

each harvest, biomass samples from the triplicate cultures were sacrificed, processed, and preserved for core lipid analysis, RNA sequencing, and proteomics (Figure 1). Given the anticipated biomass requirements for each downstream analysis and the variable OD₆₀₀ of each growth phase and condition, sampling volumes differed between experiments and are summarized in Supplementary Table 1. We also measured culture pH at each harvest to constrain average pH drifts over the course of growth under all conditions (Supplementary Table 1).

For each sampling time point, we prioritized biomass for RNA-seq as follows. The harvested culture was immediately mixed with an equal volume of chilled, sterile basal salts medium (AY medium without arabinose or tryptone) and placed on ice until centrifugation. Biomass pellets were collected by centrifuging the culture:basal salts solution for 18 min in an Eppendorf centrifuge with swinging bucket rotor set at $3,163 \times g$ (max speed) and 4°C, and decanting the resulting supernatant. Pellets were immediately placed on dry ice and then transferred to a -80°C freezer until RNA extraction. In parallel, to preserve biomass for proteomics and lipid extractions, harvested culture was centrifuged ($3,163 \times g$, 4°C, 18 min) and decanted, and resulting pellets were frozen in a -80°C freezer until further processing.

2.2. Lipid analysis

Frozen biomass pellets of *S. islandicus* harvested from growth experiments and designated for lipid extraction were freeze-dried prior to lipid extraction. We obtained core GDGT fractions from freeze-dried biomass using acid hydrolysis and solvent extraction as described previously, with modified 3N methanolic HCl incubation conditions of 65°C for 90 min (Zhou et al., 2020). Given that calditol-linked GDGTs found in *Sulfolobales* are resistant to hydrolysis by the methanolic HCl method used here, our

results only include non-calditol core GDGTs. GDGT abundances were measured by ultra-high-performance liquid chromatography (UHPLC) coupled to an Agilent 6410 triple-quadrupole mass spectrometer (MS) as originally described by Becker et al. (2013) and previously implemented for related organisms (Zhou et al., 2020). To enable analysis of small sample aliquots, data was collected using a Selected Ion Method (SIM) for core GDGTs only. The resulting peak areas of GDGTs in each sample were used to calculate ring indices (RI) according to Eq. 1. Like previous studies of *S. acidocaldarius* (Cobban et al., 2020; Zhou et al., 2020), multiple isomers of GDGTs 3–5 were detected in some *S. islandicus* samples. Peak areas of minor isomers were summed with their respective major components before RI calculation.

RI

$$\begin{aligned}
 &1^* [\text{GDGT-1}] + 2^* [\text{GDGT-2}] + 3^* [\text{GDGT-3}] \\
 &+ 4^* [\text{GDGT-4}] + 5^* [\text{GDGT-5}] + 6^* [\text{GDGT-6}] \\
 &+ 7^* [\text{GDGT-7}] + 8^* [\text{GDGT-8}] \\
 = &\frac{\quad}{[\text{GDGT-1}] + [\text{GDGT-2}] + [\text{GDGT-3}] \\
 &+ [\text{GDGT-4}] + [\text{GDGT-5}] + [\text{GDGT-6}] \\
 &+ [\text{GDGT-7}] + [\text{GDGT-8}]} \quad (1)
 \end{aligned}$$

Two sample *t*-tests were used to identify significant differences (*p*-value < 0.05) in RI and individual GDGT abundances between growth conditions.

2.3. RNA-seq

2.3.1. RNA isolation

Saccharolobus islandicus RNA was extracted using RNeasy Midi or Mini kits (Qiagen, Hilden, Germany) with modifications based on published methods (Buetti-Dinh et al., 2016). Briefly, frozen biomass pellets were defrosted on wet-ice, resuspended in 1 ml lysis

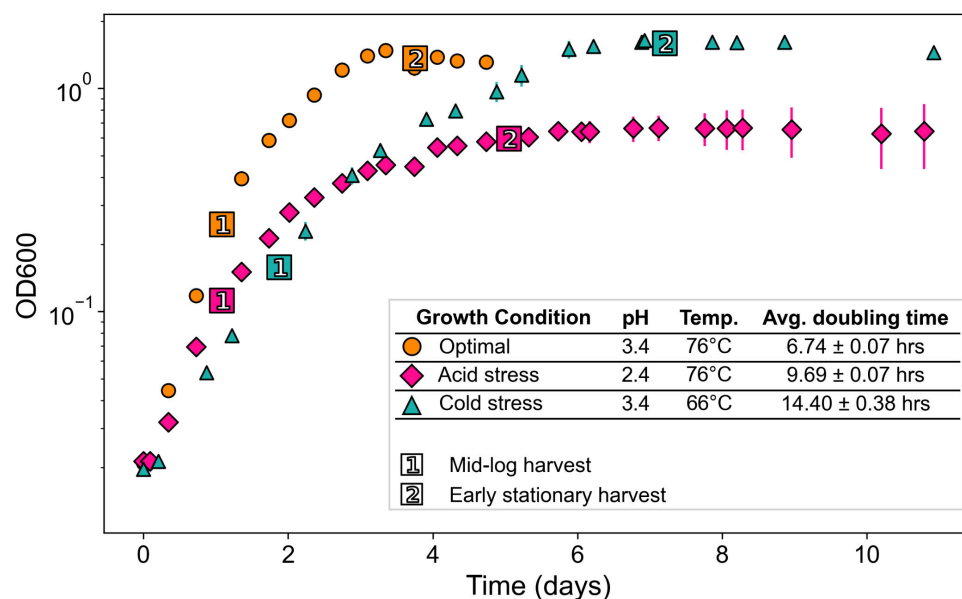


FIGURE 2

Growth curves of *S. islandicus* from all tested growth conditions. All series are the average of triplicates and error bars represent standard error. (Inset) Growth conditions with resulting growth rates and lipid ring index values. Temp., temperature; ML, mid-logarithmic phase; ES, early stationary phase.

buffer, transferred to tubes containing 0.5 mm glass beads, then subjected to three cycles of freeze-thaw with bead beating to aid in cell lysis. We followed kit instructions for the remainder of the RNA extraction. For samples extracted with RNeasy Midi kits, we used our maximum centrifugation speed ($3,163 \times g$) for all steps and increased centrifugation time of cell lysate through the filter column to 25 min and the Buffer RW1 wash step to 8 min.

Total RNA was treated with a TURBO DNA-free DNase treatment kit (Invitrogen, Waltham, MA, USA). RNA was confirmed pure by Nanodrop (Thermo Scientific, Waltham, MA, USA) measurements, quantified by Qubit RNA HS assay (Invitrogen), and checked for quality on a Lonza FlashGel RNA cassette gel (Lonza, Basel, Switzerland). Extracted RNA samples were stored at -80°C and mailed on dry-ice to the Joint Genome Institute for library preparation and sequencing.

2.3.2. Library preparation and sequencing

The Joint Genome Institute (CA, USA) performed RNA library preparation and sequencing. Briefly, rRNA was removed from 10 ng of total RNA using Qiagen FastSelect 5S/16S/23S for bacterial rRNA depletion (and additional FastSelect plant and/or yeast rRNA depletion) (Qiagen) with RNA blocking oligo technology. The fragmented and rRNA-depleted RNA was reverse transcribed to create first strand cDNA using Illumina TruSeq Stranded mRNA Library prep kit (Illumina) followed by the second strand cDNA synthesis which incorporated dUTP to quench the second strand during amplification. The double stranded cDNA fragments were polyA-tailed and ligated to JGI dual indexed Y-adapters, followed by an enrichment of the library by 10 cycles of PCR. The prepared libraries were quantified using KAPA Biosystems' next-generation sequencing library qPCR kit and run on a Roche LightCycler 480 real-time PCR instrument. Sequencing of the flowcell was performed on the Illumina NovaSeq sequencer using

NovaSeq XP V1 reagent kits, S4 flowcell, following a 2x151 indexed run protocol.

2.3.3. RNA-seq data processing and differential gene expression analysis

RNA-seq data were processed by JGI according to their RNA-Seq gene expression analysis pipeline as described here. The JGI QC pipeline was used for initial filtering and trimming of raw fastq file reads. Using BBDuk,² raw reads were evaluated for artifact sequence by kmer matching (kmer = 25), allowing 1 mismatch and detected artifact was trimmed from the 3' end of the reads. RNA spike-in reads, PhiX reads and reads containing any Ns were removed. Quality trimming was performed using the phred trimming method set at Q6. Following trimming, reads under the length threshold (25 bases or $\frac{1}{3}$ of the original read length, whichever was longer) were removed.

Filtered reads from each library were aligned to the *S. islandicus* REY15A reference genome (GenBank accession no. CP002425) using HISAT2 version 2.2.0 (Guo et al., 2011; Kim et al., 2015). Strand-specific coverage bigWig files (fwd and rev) were generated using deepTools v3.1 (Ramírez et al., 2014). featureCounts (Liao et al., 2014) was used to generate the raw gene counts file using gff3 annotations. Only primary hits assigned to the reverse strand were included in the raw gene counts (-s 2 -p -primary parameters). Raw gene counts were used to evaluate the level of correlation between biological triplicates from each growth condition at each growth phase using Pearson's correlation. All replicates had Pearson correlation coefficients >0.9 within replicate group. One replicate from the cold stress condition at early stationary phase had relatively poor correlation within its replicate group with 0.71

² <https://sourceforge.net/projects/bbmap/>

as the highest correlation coefficient. However, this replicate was still used for downstream differential gene expression analysis as it did not have a correlation coefficient $> 0.71 + 0.05$ with any libraries outside of its replicate group.

DESeq2 (version 1.30.0) (Love et al., 2014) was used to perform differential gene expression analysis between pairs of growth conditions. There were triplicate biological replicate datasets associated with every combination of growth condition and growth phase. To analyze the response of *S. islandicus* to acidic pH stress and cold temperature stress, we only analyzed the differential expression results from pairwise comparisons of each stress growth condition to optimal growth condition at the same growth phase (e.g., acidic pH stress at mid-log growth phase vs. optimal growth condition at mid-log growth phase). We considered genes to be differentially transcribed between pairs of conditions if they showed at least twofold change (\log_2 fold change ≤ -1 or ≥ 1) and had adjusted p -values < 0.05 (default p -value adjustment by DESeq2 analysis using the Benjamini–Hochberg method). Negative \log_2 fold change (\log_2 FC) values for transcripts indicate increased expression in the given stress condition relative to optimal conditions while positive \log_2 FC values indicate decreased expression in the given stress condition.

2.4. Proteomics

2.4.1. Protein extraction and LC/MS analysis

Cell pellets were extracted by vortexing and heating (95°C, 20 min) in a reducing and denaturing SDS (1%) / Tris (200 mM, pH 8.0) / DTT (10 mM) buffer, and cysteine thiols alkylated with 40 mM iodoacetamide. Proteins were purified by a modified eFASP (enhanced filter-aided sample preparation) protocol (Erde et al., 2014), using Sartorius Vivacon 500 concentrators (30 kDa nominal cut-off). Proteins were digested with MS-grade trypsin (37°C, overnight), and peptides were eluted from the concentrator dried by vacuum centrifugation. For quantitative analysis, peptides were isotopically labeled at both N- and C-termini using the diDO-IPTL methodology (Waldbauer et al., 2017). Briefly, C-termini were labeled with either oxygen-16 or -18 by enzymatic exchange in isotopic water of > 98 atom % enrichment. N-termini were labeled with either un- or dideuterated formaldehyde via reductive alkylation using sodium cyanoborohydride. Peptide extracts from each sample were split and aliquots labeled separately with $\text{CD}_2\text{O}/^{16}\text{O}$ and $\text{CH}_2\text{O}/^{18}\text{O}$; the latter were pooled to serve as a common internal standard for quantification. Aliquots of the ^{16}O -labeled peptides and ^{18}O -labeled internal standard were mixed 1:1 v/v and analyzed by LC-MS for protein expression quantification.

For LC-MS analysis, peptide samples were separated on a monolithic capillary C18 column (GL Sciences Monocap Ultra, 100 μm I.D. \times 200 cm length) using a water-acetonitrile + 0.1% formic acid gradient (2–50% AcN over 180 min) at 360 nl/min using a Dionex Ultimate 3000 LC system with nanoelectrospray ionization (Proxeon Nanospray Flex source). Mass spectra were collected on an Orbitrap Elite mass spectrometer (Thermo) operating in a data-dependent acquisition (DDA) mode, with one high-resolution (120,000 $m/\Delta m$) MS1 parent ion full scan triggering Rapid-mode 15 MS2 CID fragment ion scans of selected precursors. Proteomic mass spectral data were analyzed using

MorpheusFromAnotherPlace (MFAP; Waldbauer et al., 2017), using the predicted proteome of *S. islandicus* REY15A as a search database. Precursor and product ion mass tolerances for MFAP searches were set to 20 ppm and 0.6 Da, respectively. Static cysteine carbamidomethylation and variable methionine oxidation, N-terminal (d4)-dimethylation, and C-terminal $^{18}\text{O}_2$ were included as modifications. False discovery rate for peptide-spectrum matches was controlled by target-decoy searching to $< 0.5\%$. Protein-level relative abundances and standard errors were calculated in R using the Arm post processing scripts for diDO-IPTL data (Waldbauer et al., 2017).³

2.4.2. Differential protein abundance analysis

Significantly differential protein expression between stressed and optimal conditions at each growth phase was determined by calculating a Z-score for protein abundance differences by taking the difference in the mean (\log_2 -transformed) protein abundance between conditions and dividing it by the sum of the total uncertainty estimate for that protein in the two conditions. This total uncertainty estimate for a given condition was taken as the root-square sum of the standard deviation of a protein's abundance across the biological replicates of that condition plus the average standard error of the protein's abundance across quantified spectra within each replicate. Z-scores were converted to p -values assuming a standard normal distribution and then the familywise error rate for significantly differential expression between conditions was controlled by adjusting p -values to correct for multiple testing (Benjamini and Hochberg, 1995) with a threshold of adjusted $p < 0.1$. For consistency with our differential transcript expression analysis, we also imposed a minimum threshold of at least a twofold change ($|\log_2\text{FC}| \geq 1$). Negative $\log_2\text{FC}$ values for proteins indicate decreased abundance in the given stress condition relative to optimal conditions while positive $\log_2\text{FC}$ values indicate increased abundance in the given stress condition. Proteomic mass spectral data are available via proteomeXchange under accession PXD037641 and the MassIVE repository⁴ under accession MSV000090562 (username for reviewers: MSV000090562_reviewer / password: rey15a).

3. Results

3.1. Acid and cold stress conditions impair *S. islandicus* growth

Acid and cold stress impaired *S. islandicus* growth relative to optimal conditions (Figure 2). Optimal growth conditions yielded the fastest average doubling time observed: 6.8 ± 0.1 h. Doubling times under acidic pH and cold temperature stress conditions were 9.8 ± 0.1 and 14.0 ± 0.5 h, respectively. Maximum culture densities also differed among growth conditions (Figure 2). Cultures grown under optimal conditions reached an average maximum optical density of 1.48 ± 0.06 while acidic pH and cold temperature cultures reached average maximum optical densities of 0.75 ± 0.11

³ github.com/waldbauerlab

⁴ massive.ucsd.edu

and 1.69 ± 0.04 , respectively. An ANOVA of average maximum cell densities indicated that at least one mean significantly differed from the others ($p < 0.001$), and a *post-hoc* Tukey HSD test showed that the acidic pH cultures had a significantly lower maximum cell density than optimal or cold temperature cultures.

3.2. *Saccharolobus islandicus* GDGT distributions change in response to acid and cold stress

Saccharolobus islandicus produced GDGT-0 through -6 in all experimental conditions, while average GDGT cyclization differed among growth conditions and growth phases (Figure 3). Small amounts of GDGT-7 occurred in profiles from optimal and acid stress conditions, but its relative abundance never exceeded 1% (Supplementary Table 2). As previously observed in *S. acidocaldarius*, we also detected the late eluting isomers (denoted by ' symbol) of GDGTs 3–5 in *S. islandicus* lipid profiles (Cobban et al., 2020; Zhou et al., 2020). Isomers were particularly abundant (up to 10% relative abundance) in GDGT profiles from acid stress conditions (Supplementary Table 2). The average cyclization of GDGTs, as interpreted using Ring Index (RI) values, ranged from 1.59 to 3.93 for each combination of growth condition and growth phase (Figure 3). Average RI values significantly increased from mid-log phase to early stationary phase in optimal and cold stress conditions, indicating greater GDGT cyclization during later growth (Figure 3). Average RI increased at later growth under acid stress as well, but not significantly ($p = 0.09$).

Both acid stress and cold stress induced significantly decreased average RI values relative to optimal conditions at each growth phase (Figure 3). At mid-log growth phase, *S. islandicus* grown under optimal conditions had an average RI value of 3.65 ± 0.05 while acid and cold stress resulted in average RI values of 1.59 ± 0.10 and 2.83 ± 0.05 , respectively. Similarly at early stationary phase, optimal conditions yielded an average RI value of 3.93 ± 0.03 while acid and cold stress resulted in average RI values of 2.08 ± 0.29 and 3.67 ± 0.04 , respectively. Though both stressors resulted in lower RI, comparison of GDGT profiles resulting from each stress show that acid and cold conditions induced distinct changes of different individual GDGT abundances (significance determined by *t*-test). For example, cold stress GDGT profiles showed decreased relative abundances of all detected GDGTs with ≥ 5 rings while GDGT-5 was the only higher-numbered ring showing decreased abundance in acid stress profiles (Figure 3). Unlike cold stress, the decrease in RI under acid stress conditions was driven largely by increased abundances of GDGTs 0–2 and a dramatic reduction of GDGT-4 (Figure 3). Differences in the underlying GDGT abundance changes induced by each stress condition highlights the value of interpreting complete lipid profiles in addition to RI values.

3.3. RNA-seq

We obtained a total of 108.4 million raw reads from the 18 *S. islandicus* RNA samples derived from our growth experiments. Raw reads per sample averaged at 6.0 million (range: 3.7–10.1

million). After quality filtering and mapping fragments to the *S. islandicus* REY15A genome, an average of 91% (range: 65–95%) of fragments could be assigned to genes. Complete read and mapping data can be found in Supplementary Table 3.

To understand how *S. islandicus* transcriptomes responded to acid and cold stress, we compared the RNA-seq data from either stress condition versus optimal growth conditions as our control for baseline expression. Differential gene expression analysis showed that acid stress and cold stress each induced differentially expressed genes (DEGs) at both sampled growth phases though more DEGs occurred at early stationary phase in both stress conditions (Figure 4A).

We compared the DEGs from both stresses by growth phase and directionality (up or downregulated) to examine the number of DEG responses that were unique to either stress or common to both stresses in this study (Figure 4B). At both growth phases and for both expression directions, acid stress always yielded the highest proportion of observed DEGs (Figure 4B). At early stationary phase, a significant proportion of DEGs were commonly induced by both acid and cold stress, potentially representing more generalized stress responses by *S. islandicus* to environmental perturbations. Complete differential gene expression results are available in Supplementary Table 4. We present specific RNA-seq results with an initial focus on GDGT biosynthesis-related DEGs followed by an overview of broader DEG responses grouped by inducing stressor(s).

3.3.1. Differentially expressed GDGT biosynthesis genes

To better understand the mechanisms by which GDGT cyclization is affected by environmental stresses, we looked for differential expression among GDGT biosynthesis genes (Figure 5), particularly the GDGT ring synthase genes *grsA* and *grsB* that are essential for the formation of cyclic moieties on GDGTs (Zeng et al., 2019). *S. islandicus* did not differentially express *grsA* (SiRe_1404) under any condition ($|\log_2FC|$ always < 1) while *grsB* (SiRe_1524) expression responded to both acid and cold stress. *S. islandicus* upregulated *grsB* at mid-log phase under acid stress conditions ($\log_2FC = -2.1$) and downregulated *grsB* in both growth phases under cold stress ($\log_2FC = 2.1$ and 2.5 at mid-log and early stationary phase, respectively).

Unique to cold stress conditions during the early stationary phase, *S. islandicus* upregulated its homolog of tetraether synthase (*tes*; SiRe_1758). Tetraether synthase catalyzes the condensation of pairs of archaeol to form GDGTs (Lloyd et al., 2022; Zeng et al., 2022), and the upregulation of *tes* observed here may indicate increased tetraether production induced by cold stress.

Several other GDGT biosynthesis genes showed differential expression under both stress conditions. Two genes encoding enzymes that catalyze consecutive steps in the formation of GDGT synthesis intermediate DGGGP displayed transcriptional shifts under both acid and cold stress conditions during early stationary phase: GGPP synthase (SiRe_1931) was upregulated while GGGP synthase (SiRe_1736) was downregulated (Figure 5). While GGPP and GGGP are both GDGT precursors, we note that both also serve as precursors for dialkyl glycerol diether (DGD) lipid synthesis, and GGPP serves as a precursor for other polyterpene compounds including quinones and polyprenols (Koga and Morii, 2007; Salvador-Castell et al., 2019).

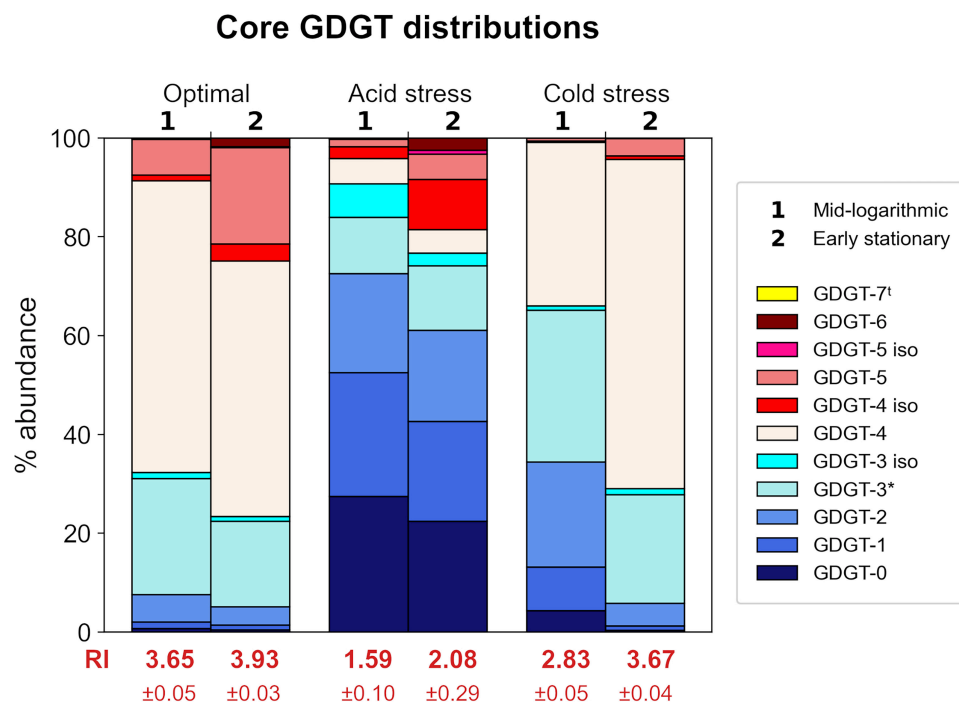


FIGURE 3

Relative abundances of core GDGTs from each growth condition at each time point harvested. Each stacked bar chart represents the average of triplicates. Ring Index (RI) values corresponding to each profile are indicated in red text. Isomers are indicated by “iso” in legend. †GDGT-7 was detected in some samples, but only in small amounts <1% relative abundance. *GDGT-3 abundances are based on peak areas inclusive of a shoulder observed in corresponding chromatogram peaks.

Saccharolobus islandicus upregulated one headgroup attachment related gene, the putative archaetidylinositol phosphate (AIP) phosphatase (SiRe_0209), under acid stress conditions during early stationary phase. AIP phosphatase may dephosphorylate archaetidylinositol AIP to form the archaetidylinositol (AI) headgroup (Morii et al., 2009). However, AIP synthase (SiRe_1447), which catalyzes the initial attachment of 1L-myo-inositol-1-phosphate (IP) headgroup to CDP-archaeol to form AIP, was not differentially expressed under either stress condition (Morii et al., 2009). Calditol synthase (*cds*), for calditol headgroup synthesis, was previously shown to be important for acid tolerance in *S. acidocaldarius* (Zeng et al., 2018). Despite this, the *S. islandicus* homolog of *cds* (SiRe_1626) was not a DEG in our dataset, suggesting that it may be constitutively expressed rather than transcriptionally regulated by acid stress conditions (Zeng et al., 2018). While *Sulfolobales* sp. lipids are known to have other possible headgroups besides IP and calditol, the genes for synthesizing these lipid headgroups are yet to be identified (Langworthy, 1977; De Rosa et al., 1980b; Daiyasu et al., 2005; Jensen S. et al., 2015; Jensen S. M. et al., 2015).

We also looked for DEGs related to the mevalonate (MVA) pathway, which forms isoprenoid building blocks as precursors for GDGT biosynthesis. We found five MVA pathway genes to be DEGs, all upregulated during early stationary phase. Both acid and temperature stress induced upregulation of the genes encoding for the first three steps of the MVA pathway: acetoacetyl-CoA thiolase (AACT; SiRe_1461), 3-hydroxy-3-methylglutaryl-CoA synthase (HMGS; SiRe_1462), and 3-hydroxy-3-methylglutaryl-CoA reductase (HMGR; SiRe_1459)

(Supplementary Figure 1). The final two MVA pathway-related DEGs were phosphomevalonate kinase (PMK; SiRe_2253), upregulated only under cold stress, and isopentenyl phosphate kinase (IPK; SiRe_1929), upregulated only under acid stress. PMK is a part of the classical MVA pathway and is absent in many archaea, but present in the *Sulfolobales* (Jain et al., 2014). IPK is part of the “modified” or “alternative” MVA pathway that is conserved among archaea (Jain et al., 2014). The observed PMK and IPK differential expression results suggest that their respective MVA pathways are favored under specific stress conditions.

3.3.2. Common transcriptional responses to both acid and cold stress

We analyzed the *S. islandicus* response to acid or cold stress across the entire transcriptome. Both environmental stressors, a pH unit shift in acidity (from pH 3.4 to 2.4) or a ten degree decrease in temperature (from 76 to 66°C), induced 467 common DEGs with the same directionality of expression change at early stationary phase (Figure 4B). We highlight several of these common DEGs below and suggest that they may provide candidate genes for studying generalized stress response in *S. islandicus* facing other types of environmental perturbations.

Differential gene expression results supported decreased energy production and motility as part of the acid and cold stress response in *S. islandicus*. At early stationary phase under both stresses, *S. islandicus* downregulated all three terminal oxidases (DoxBCE, SoxABCDL, and SoxEFGHIM) of its electron transport chain (ETC) (Figure 6). Succinate dehydrogenase, which reduces quinones upstream of cytochromes

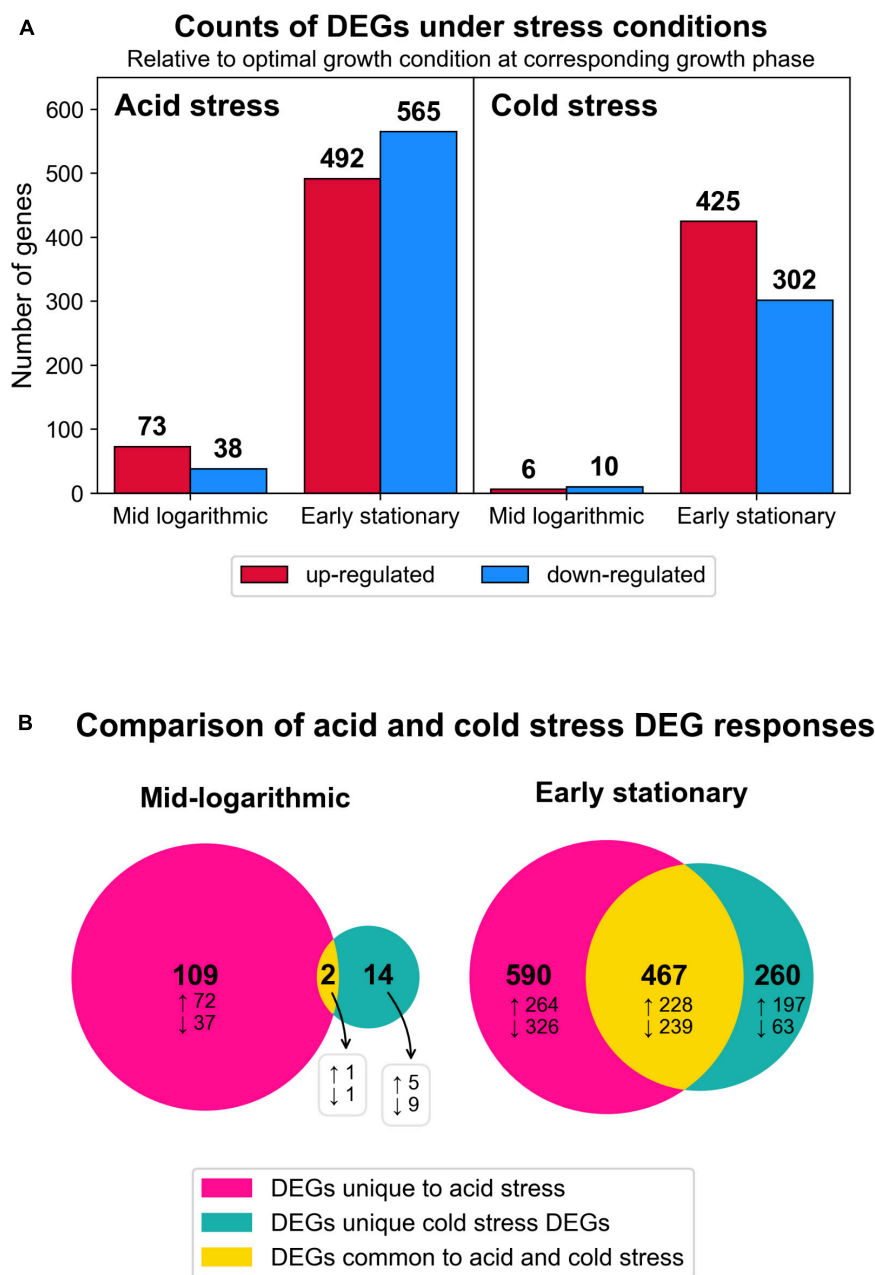


FIGURE 4

Summary of transcriptomic changes. **(A)** Counts of differentially expressed genes (DEGs) (\log_2 fold change ≥ 1 , adjusted $p < 0.05$) in response to both stress conditions at each time point harvested. All results are relative to optimal growth conditions at corresponding growth phase. **(B)** Venn diagrams depicting the numbers of DEG responses (transcript identity and direction of change) that were unique to a given stress condition or common to both stresses tested. Total numbers of DEG responses are in bold text and the subset of DEG responses that were up or downregulated are denoted by up and down arrows, respectively.

and terminal oxidases in the ETC, was downregulated in response to cold stress. Decreased transcription of ETC genes suggests decreased or slower energy production that is consistent with the slower growth rates observed in response to both acid and cold stress. Similarly, differential expression of archaeum genes and archaeum regulators suggested that *S. islandicus* decreased motility in response to both stressors. Most or all of the *S. islandicus* archaeum-encoding *fla* cluster showed downregulation in response to both stresses during early stationary phase (Figure 6). Additionally, differential expression

of archaeum regulators *abfR1*, *arnR*, and *arnC* in response to acid and/or cold stress all supported repression of the archaeum component *flaB* (Figure 6; Lassak et al., 2013; Orell et al., 2013; Hoffmann et al., 2017). The *slaA* and *slaB* genes encoding the *S. islandicus* S-layer, which anchors the archaeum, were also downregulated in response to acid and cold stress during later growth (Figure 6; Banerjee et al., 2015; Tsai et al., 2020). Interestingly, two hypothetical genes (SiRe_0125 and SiRe_0126) located between the *S. islandicus* *fla* cluster and *arnR* were also downregulated during early stationary phase under both stresses.

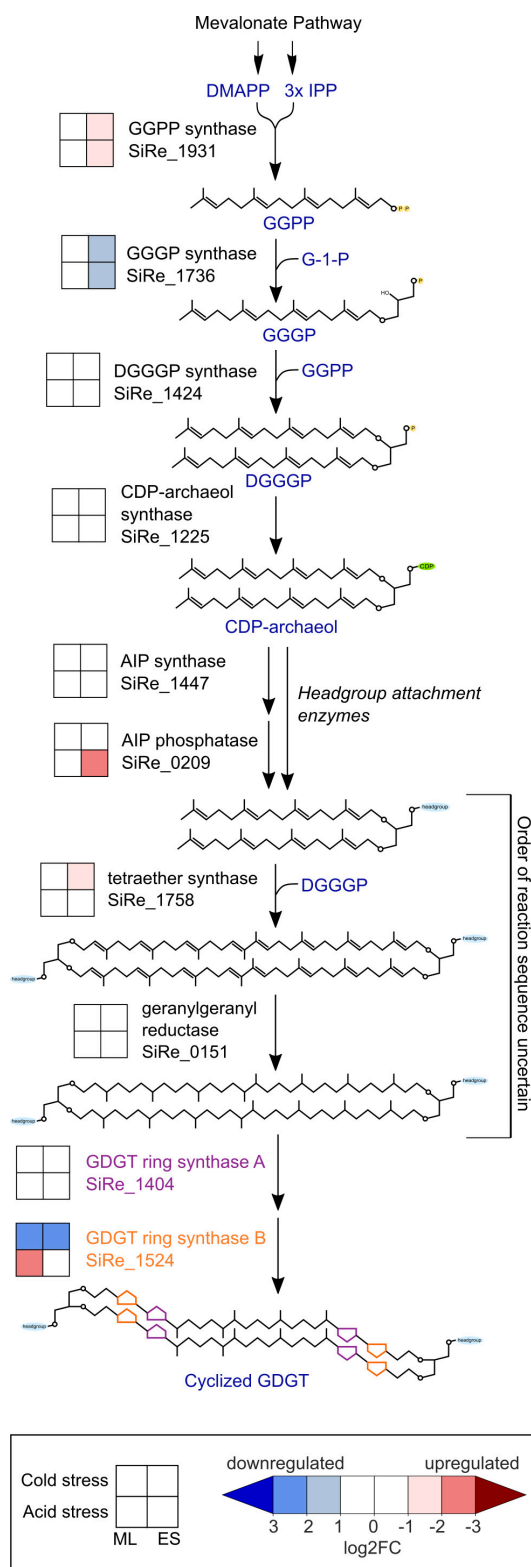


FIGURE 5

Transcriptomic changes within the GDGT biosynthesis pathway in response to cold stress or acid stress. GDGT, glycerol dibiphytanyl glycerol tetraether; DMAPP, dimethylallyl pyrophosphate; IPP, isopentenyl pyrophosphate; GGGP, geranylgeranyl diphosphate; GGPP, 3-O-geranylgeranyl-sn-glyceryl-1-phosphate; DGGGP, 2,3-bis-O-geranylgeranyl-sn-glyceryl-1-phosphate; CDP, cytidine diphosphate; AIP, archaetidylinositol phosphate; ML, mid-logarithmic phase; ES, early stationary phase.

These genes do not have homologs in *S. acidocaldarius*, the organism used for the majority of archaeal characterization work to date (Albers and Jarrell, 2015), but the genomic context and consistent expression pattern of these genes in *S. islandicus* suggest they contribute to *S. islandicus* archaeal-related motility.

While *S. islandicus* seemed to decrease energy production and motility, differential expression results indicated upregulation of other functions in response to acid and cold stressors. For example, upregulation of homologous recombination repair (HRR) genes encoding for the Holliday junction cleavage enzyme (SiRe_1431) and the Hel308A/Hjm DNA repair helicase (SiRe_0250) suggest increased DNA repair in acid and cold stress conditions during early stationary phase (Komori et al., 1999; Li et al., 2008). Similarly, *S. islandicus* upregulated numerous RNA polymerase genes, ribosomal protein genes, and translation initiation factors suggesting increased transcription and translation under both stressors. Differential expression, largely downregulation, of numerous genes encoding VapBC toxin-antitoxin system proteins (Supplementary Table 5), also supports changes in translation activity as part of the *S. islandicus* acid and cold stress responses. VapBC toxin-antitoxin systems affect the translation of specific RNAs (Walling and Butler, 2019), and have previously been associated with archaeal heat stress and uranium stress responses (Cooper et al., 2009, 2023; Maezato et al., 2011; Mukherjee et al., 2017).

In addition to DEGs supporting upregulated DNA repair, transcription, and translation, our data also suggested increased anabolic activity in response to acid and cold stress during early stationary phase. *S. islandicus* upregulated pathways for biosynthesis of the vitamin and cofactor cobalamin, the purine nucleotide precursor inosine 5'-monophosphate (IMP), the amino acid histidine, the versatile biochemical intermediate chorismate, and polyamines (Figure 6). *S. islandicus* concurrently increased its expression of anaplerotic enzyme-encoding genes (Figure 6), which replenish TCA cycle intermediates that are diverted for biosynthesis. Thus, the evidence for increased anaplerosis complements the evidence for increased biosynthesis in response to acid and cold stressors.

Interestingly, several nitrate reduction genes shifted in response to acid and cold stress. *S. islandicus* REY15A belongs to a subset of *S. islandicus* strains with nitrate reduction genes predicted in their genomes (Guo et al., 2011). During early stationary phase of both stress conditions, *S. islandicus* downregulated a qNOR nitric oxide reductase gene (*norZ*; SiRe_0538) and upregulated the catalytic and electron transfer subunits of the Nar nitrate reductase (*narG*; SiRe_0444 and *narH*; SiRe_0443, respectively) (Figure 6). qNORs use electrons derived from the quinone pool to reduce nitric oxide as part of denitrification or as a means of scavenging and detoxifying nitric oxide (Hendriks et al., 2000). Nar is a nitrate-reducing system typically associated with anaerobic nitrate respiration and generating proton motive force. An annotated sulfite reductase (SiRe_2311), which could be bifunctional for sulfite and nitrite, was also upregulated in cold stress conditions (Crane and Getzoff, 1996). Although anaerobic nitrate respiration by *S. islandicus* is not expected, particularly in our aerobic experiments, the differential expression of these dissimilatory nitrate reduction genes suggest they may play a role in the acid and cold stress responses of *S. islandicus*.

upregulated DNA damage sensing and repair genes including two HRR genes, *mre11* (SiRe_0063; log2FC = −1.0) and *rad50* (SiRe_0062; log2FC = −1.3), DEAD/DEAH box helicase domain genes (SiRe_1605 and SiRe_1977; log2FC = −1.9 and −1.8, respectively), and the thermophile-specific reverse gyrase, *topR1* (SiRe_1581; log2FC = −1.6) (Forterre, 2002; Kampmann and Stock, 2004; Fröls et al., 2007; Quaiser et al., 2008; Song et al., 2016). The thermosome (subunits α , SiRe_1214; β , SiRe_1716; γ , SiRe_2245), a chaperonin with roles in protein synthesis and repair, was also upregulated throughout growth (log2FC range: −2.1 to −1.3) under acid stress conditions (Klumpp and Baumeister, 1998). With regard to molecular degradation processes, we observed upregulation of the exosome complex (SiRe_1273-1275; log2FC range: −1.3 to −1.0), which mediates polyadenylation and RNA degradation (Portnoy et al., 2005), and upregulation of the PAC2 assembly factor (*pbaA*; SiRe_1724; log2FC = −1.1) for the protein-degrading 20S proteasome (Kusmierczyk et al., 2011).

Consistent with the observations of DEGs related to molecular turnover processes, *S. islandicus* may have also increased fatty acid degradation as an acid-specific response. Six genes belonging to a genomic cluster of fatty acid metabolism genes were upregulated under acid stress during mid-log and/or early stationary phase (Supplementary Figure 4). Accordingly, the fatty acid metabolism repressor, TetR (SiRe_0301), co-localized within this cluster did not show differential expression (Wang et al., 2019). However, an adjacent Fis family transcriptional regulator (SiRe_0302) was upregulated under acid stress suggesting that it may play a role in regulating fatty acid metabolism of *S. islandicus*. Previous work linked fatty acid metabolism genes to acid stress adaptation in *Saccharolobus solfataricus* as a possible means of accelerating membrane turnover (McCarthy et al., 2016).

Downregulated DEGs specific to acid stress may represent genes and functions that do not play important roles in *S. islandicus* under low pH conditions. Downregulated acid-specific DEGs included 15 *orfB* elements belonging to the IS200/IS605 family of transposable elements, all of which showed downregulation (log2FC range: 1.3–4.4) at early stationary phase. We also observed acid-specific downregulation of one small heat shock protein chaperone (SiRe_2601; log2FC = 2.8) and two HtpX proteases (SiRe_0639 and SiRe_2013; log2FC = 1.7 for both) suggesting that these genes may not be important to the protein repair and turnover processes in acidic conditions. Finally, of the four universal stress proteins (USPs) predicted in the *S. islandicus* REY15A genome, only one was an upregulated acid-specific DEG (SiRe_1457; log2FC = −1.4) while all others were downregulated (SiRe_0634, SiRe_2404, and SiRe_2055), two of which were acid-specific responses (SiRe_0634 and SiRe_2402; log2FC = 2.2 and 1.6, respectively).

3.3.4. Transcriptional responses unique to cold stress

Cold temperature stress induced unique transcriptional changes in *S. islandicus* that were not observed under acid stress. Although cold stress induced less than half the number of unique DEG responses compared to the number of acid-specific DEGs (Figure 4B), the majority of cold-specific DEGs occurred during early stationary phase rather than mid-log phase, consistent with observations of the acid-specific DEGs relative to growth phase

(Figure 4). The DEGs unique to cold stress may provide insight into cold-specific responses of thermoacidophilic archaea, and we highlight several observations related to carbon and energy metabolism, transposases, and DNA repair. At early stationary phase, *S. islandicus* downregulated all four subunits comprising succinate dehydrogenase (SiRe_0155-0158; log2FC range: 1.2–1.3), which plays a role in both the TCA cycle and the ETC (Figure 6). Cold stress also induced upregulation of numerous predicted transposase-encoding genes at early stationary phase including 11 *orfB* elements (log2FC range: −3.4 to −1.3), 6 ISH3 family transposases (log2FC range: −3.6 to −2.7), and 1 ISH6 family transposase (log2FC = −3). Upregulation of transposases in cold stress conditions contrasts the general downregulation of transposases observed in acid stress conditions. With regard to DNA repair, *S. islandicus* strongly upregulated the HRR-related Holliday junction endonuclease (*hje*, SiRe_0930; log2FC = −4.2) during early stationary phase. This particular Holliday junction resolvase has been proposed to have a viral origin and is only found in certain archaea (Kvaratskhelia and White, 2000a,b; Middleton et al., 2004). Strong upregulation of *hje* may indicate that it is particularly important for cold-induced DNA damage.

3.4. Proteomic responses to acid and cold stress

Quantitative proteomics revealed protein level changes associated with acid and cold stress responses of *S. islandicus*. We detected a total of 1,107 unique proteins, which account for 42% of the total predicted proteins (2,644) encoded by the *S. islandicus* REY15A genome. Differential abundance analysis revealed that 18 proteins significantly changed abundance levels under acid stress and 7 proteins significantly changed abundance under cold stress (Table 1). Of all the observed significant protein abundance changes, only six showed consistent transcriptomic changes at corresponding growth phase and stress condition including the heat shock protein (SiRe_2601) downregulated at early stationary phase under acid stress (Table 1). Another nine significant protein changes showed transcriptional changes at corresponding growth phase and stress condition, but with opposite directionality (Table 1 and Figure 7).

Some of the largest magnitude protein level changes in *S. islandicus* included the downregulation of formate dehydrogenase alpha subunit (SiRe_2464; log2FC = −3.31) and the upregulation of UDP-sulfoquinovose synthase Agl3 (SiRe_2623; log2FC = 2.98), both under acidic stress conditions during early stationary phase. SiRe_2464 is homologous to the non-canonical formate dehydrogenase (FDH) YjgC subunit from *Bacillus subtilis* (E-value 0.0, 53% identities), which complexes with a second subunit to form a FDH that couples formate oxidation with menaquinone reduction (Arias-Cartin et al., 2022). Agl3 forms UDP-sulfoquinovose, a component of S-layer and archaeum N-glycans that was important for *S. acidocaldarius* growth under high salt conditions (Meyer et al., 2013). Cold stress resulted in a strongly decreased abundance of the arginine biosynthesis protein argininosuccinate lyase ArgH (SiRe_1371; log2FC = −2.83) during mid-log phase. Despite these proteins showing some of the highest magnitude changes in our proteomic dataset, none had consistent

TABLE 1 Summary of *S. islandicus* proteomic responses to tested environmental stresses.

Condition and phase	Log2FC*	Annotation	Locus tag	Transcript change direction
Acid stress				
Mid-log	2.08	Ornithine carbamoyltransferase	SiRe_1208	N/A
	−3.01	Conserved hypothetical protein	SiRe_2064	↑
	−2.29	Ketol-acid reductoisomerase	SiRe_0753	N/A
Early stationary	2.98	UDP-sulfoquinovose synthase, Agl3	SiRe_2623	↓
	2.38	AMP-dependent synthetase and ligase	SiRe_2451	↓
	1.94	Mandelate racemase (MR) subfamily	SiRe_2246	N/A
	1.62	Aminotransferase class V	SiRe_0032	↓
	1.17	Conserved hypothetical protein	SiRe_2599	↓
	1.15	Conserved hypothetical protein	SiRe_2391	↑
	−3.31	Formate dehydrogenase, alpha subunit	SiRe_2464	↑
	−2.58	Thioredoxin/glutaredoxin-like protein	SiRe_1684	N/A
	−2.19	Heat shock protein (Hsp20)	SiRe_2601	↓
	−2.11	Thermosome	SiRe_1716	↑
	−2.01	Thermosome	SiRe_1214	↑
	−1.81	Deoxyhypusine synthase	SiRe_1119	↑
	−1.70	Glutamine synthetase, type I	SiRe_1633	N/A
	−1.61	CoA-binding domain protein	SiRe_0989	↓
	−1.59	Conserved hypothetical protein	SiRe_2064	↓
	−1.57	Extracellular solute-binding protein family 1	SiRe_1066	↓
Cold stress				
Mid-log	1.22	A-type ATPase subunit, AtpI	SiRe_1444	N/A
	−2.83	Arginosuccinate lyase	SiRe_1371	N/A
	−2.19	Ketol-acid reductoisomerase	SiRe_0753	N/A
	−1.44	Glycerol kinase	SiRe_0511	N/A
	−1.29	Beta-glucosidase	SiRe_2202	N/A
	−1.08	Amidohydrolase	SiRe_0643	N/A
Early stationary	1.85	Short-chain dehydrogenase/reductase SDR	SiRe_0099	N/A
	−1.17	Oxidoreductase molybdopterin binding protein	SiRe_2037	↓

*Negative log2FC values indicate decreased protein levels under stress conditions relative to optimal conditions. Positive log2FC values correspond with increased protein levels under the stress condition. Log2FC SE values can be found in [Supplementary Table 6](#). If applicable, “Transcript Change Direction” column indicates the direction of expression change of corresponding transcripts.

transcriptional changes at corresponding growth phase and stress conditions, suggesting *S. islandicus* might engage significant post-transcriptional regulation in its acid and cold stress responses.

4. Discussion

To better understand the cell biology of GDGT-forming archaea when they modify their membranes, we quantified the lipid, protein, and transcript responses of *S. islandicus* cultivated under acid or cold stress. Only one study to-date has used proteomics and transcriptomics in parallel with membrane lipid profiles to characterize a thermoacidophilic archaeon's stress response, specifically the *S. acidocaldarius* nutrient

limitation response (Bischof et al., 2019). We subjected the model thermoacidophile, *S. islandicus* REY15A, to acid or cold stress, inducing shifts in GDGT cyclization (Figure 3), protein expression (Table 1), and gene transcription (Figure 4). The RNA and protein data overlapped minimally, consistent with prior studies in another thermoacidophilic crenarchaeon, *S. acidocaldarius*, suggesting that stress responses in *Sulfolobales* sp. involve multiple levels of gene transcription and translation (Figure 7; Bischof et al., 2019; Benninghoff et al., 2021). Here, stress conditions induced hundreds of transcriptional changes and only tens of protein level changes in *S. islandicus*. While most of the differentially expressed proteins in *S. islandicus* under acid stress were also differentially transcribed, many of these loci showed opposite directionality between data type (Table 1). Overall, we focus this discussion first on GDGT profiles

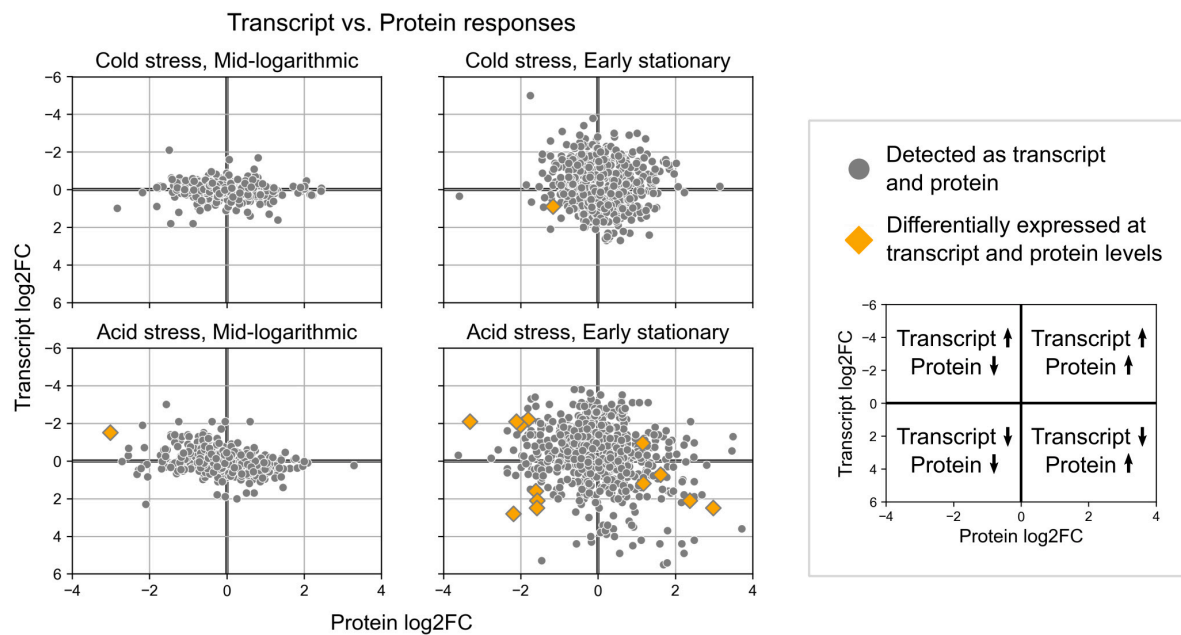


FIGURE 7

Expression change results from differential gene expression and differential protein abundance analyses for loci detected in both datasets. Orange diamonds highlight the few loci that showed significant changes in both transcript expression and protein abundances.

and expression responses of the GDGT ring synthases, then on the broader transcription responses of *S. islandicus* under acid or cold stress conditions.

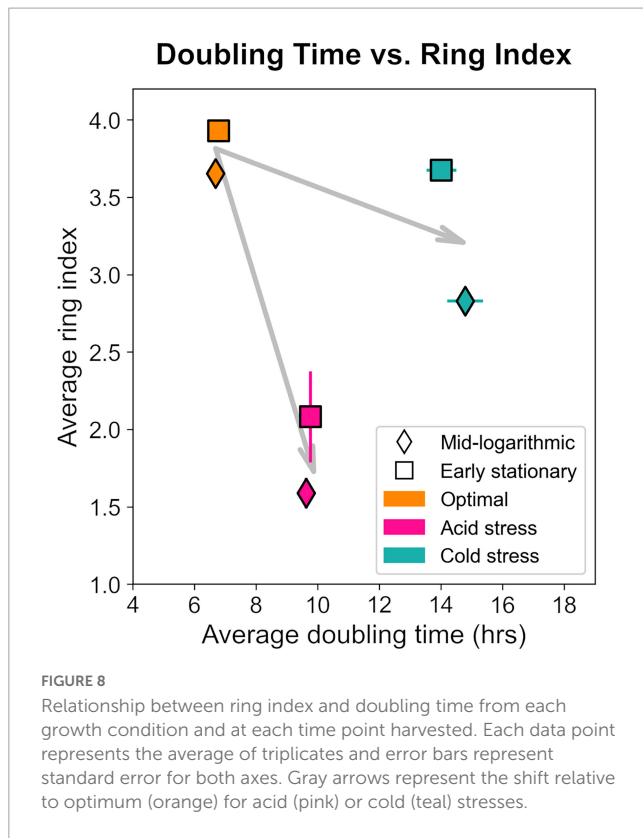
4.1. Acid or cold stress effects on GDGTs and GDGT biosynthesis genes

As expected, acid and cold stress both induced shifts in the average cyclization of *S. islandicus* GDGTs. While each stress caused changes to the relative abundances of different individual GDGTs, both sets of stress-induced changes resulted in lower RI values indicating lower average cyclization (Figure 3). The GDGT responses of various archaeal groups to a wide variety of stressors have been reviewed in the literature (e.g., Oger and Cario, 2013; Law et al., 2020; Rastädter et al., 2020), but we limit our discussion here to comparisons between our data and GDGT studies examining similar stress conditions. Lower RI at lower temperature is consistent with previous laboratory studies of thermoacidophiles as well as mesophilic neutrophiles (Figure 3; De Rosa et al., 1980a; Uda et al., 2001, 2004; Boyd et al., 2011; Elling et al., 2015; Jensen S. et al., 2015; Feyhl-Buska et al., 2016; Cobban et al., 2020).

Shifts in archaeal GDGT cyclization in response to pH changes are more complex. We observed lower RI at lower pH in *S. islandicus* (Figure 3), which is consistent with previous laboratory experiments with *S. acidocaldarius* (Cobban et al., 2020). However, a coherent empirical relationship between archaeal GDGT cyclization and pH has not yet been established (Siliakus et al., 2017). Moreover, comparison of our data to prior reports of GDGT shifts in response to pH is difficult given that several other studies only examined effects of higher pH's rather than lower pH

(e.g., Shimada et al., 2008; Boyd et al., 2011). Thermoacidophilic archaea such as *S. islandicus* maintain circumneutral intracellular pH while living in extremely acidic environments, resulting in steep proton gradients across their membranes. Shifting GDGT cyclization serves as a pH homeostasis mechanism for thermoacidophiles to maintain these proton gradients (Baker-Austin and Dopson, 2007; Slonczewski et al., 2009). Highly cyclized GDGTs form more tightly packed and less permeable membranes that may help limit passive proton diffusion and in turn protect against cytoplasmic acidification (Gabriel and Chong, 2000; Shinoda et al., 2005; Baker-Austin and Dopson, 2007). Thus, we initially predicted *S. islandicus* might increase RI as one means of counteracting the higher external concentration of protons at lower pH. In contrast, *S. islandicus* exhibited lower RI at more acidic pH, indicating that its membrane was more permeable. If increased membrane permeability allowed greater proton diffusion into the cell, the stunted maximum cell density of our cultures grown under acid stress (Figure 2) could represent cells succumbing to cytoplasmic acidification. Other factors can affect archaeal membrane permeability including GDGT headgroup composition and it follows that quantitative headgroup analysis and/or direct cytoplasmic pH measurements would help validate our proposed hypothesis (Oger and Cario, 2013).

Under both stress conditions, *S. islandicus* GDGT compositions differed by growth phase. RI increased at early stationary phase relative to mid-log phase within both the acid and cold stress conditions, as well as in optimal growth conditions (Figure 3). This is consistent with previous observations from other archaeal batch studies sampled at similar growth phases including *S. islandicus* and *S. acidocaldarius* as well as the with *Picrophilus torridus* and *Nitrosopumilus maritimus* (Elling et al., 2014; Jensen S. et al., 2015; Feyhl-Buska et al., 2016; Cobban et al., 2020). We suggest that



our observations from *S. islandicus* are also consistent with the findings from continuous culture experiments of *S. acidocaldarius* and *N. maritimus* showing that RI increases as electron availability decreases (Hurley et al., 2016; Quehenberger et al., 2020; Zhou et al., 2020). As our batch cultures grew, *S. islandicus* actively consumed medium nutrients and thereby depleted its available energy supply. Here we observed increased RI values from mid-log to early stationary phase, which we infer as a transition to more energy limited conditions.

Notably, *S. islandicus* RI and doubling time did not linearly covary when comparing data from all tested growth conditions suggesting that RI changes are not a result of growth rate alone (Figure 8). pH and temperature are distinct physical parameters that likely affect cell physiology and thereby GDGT composition differently. If we consider the general effects of acid stress and cold stress independently, it is clear that both stressors induced slower growth and decreased RI relative to optimal conditions. These observations are consistent with the broad trend of environmental stress causing longer doubling times and decreased RI as observed in *S. acidocaldarius* (Cobban et al., 2020).

To provide context for the shifts in *S. islandicus* GDGT compositions induced by acid and cold stress, we examined expression changes of known GDGT biosynthesis genes under each condition. Most notable were the expression changes in the *S. islandicus* GDGT ring synthase genes, *grsA* and *grsB*. The *grs* genes encode the GrsA and GrsB enzymes that are required for cyclopentane ring formation in *S. acidocaldarius* and *S. solfataricus* (Zeng et al., 2019). GrsA produces GDGTs 1 through 4 and GrsB subsequently acts on GDGT-4 to produce

GDGTs 5 through 8 (Zeng et al., 2019). A recent study using RT-qPCR and Western blot analysis examined GDGT ring synthase expression in *S. acidocaldarius* cultured at more acidic pH or colder temperature than its optimal conditions of pH 3.5 and 75°C (Yang et al., 2023). Acid stress (pH 2.5) induced increased *grsB* and GrsB expression in *S. acidocaldarius* while cold stress (65°C) caused decreased *grsB* expression and undetectable levels of GrsB (Yang et al., 2023). In contrast, *S. acidocaldarius* did not show differential expression of *grsA* nor GrsA in either stress condition at 48 or 72 h timepoints, which are the closest analogs to the mid-log and early stationary phase timepoints in our experiments (Yang et al., 2023). Results from our study indicated similar expression patterns of the *grs* genes in *S. islandicus* grown under acid stress or cold stress. *S. islandicus* did not differentially express *grsA* or GrsA in response to either stress, supporting constitutive transcription of *grsA* and insignificant GrsA protein level changes (Figure 5). However, *S. islandicus* showed differential expression of *grsB* as responses to both stress conditions in alignment with observations from *S. acidocaldarius* (Figure 5; Yang et al., 2023). Under cold stress at both mid-log and early stationary phases, *S. islandicus* downregulated *grsB*, and consistent with the gene's expected function, cold stress lipid profiles had lower abundances of GDGTs with ≥ 5 rings (Figure 3). Under acid stress, *S. islandicus* upregulated *grsB* at mid-log phase, but changes in acid stress lipid profiles did not correspond as predicted by GrsB function. That is, no GDGTs with ≥ 5 rings significantly increased under acid stress (Figure 3). The isomer of GDGT-5 (GDGT-5') increased under acid stress, but did not result in an overall increased abundance of GDGTs with 5 rings. The inconsistency between increased *grsB* transcription and the predicted changes in acid stress GDGT profiles may be related to the fact that like most of the DEGs in this study, the abundance of GrsB protein did not increase in tandem with higher transcription. Additionally, GrsB requires GDGT-4 as a substrate to produce more cyclized GDGTs, and GDGT-4 abundance was starkly lower in *S. islandicus* acid stress lipid profiles relative to optimal conditions (Figure 3; Zeng et al., 2019). Finally, GDGT profiles are likely not only affected by the GDGT ring synthase transcription and protein levels quantified here, but also by variable Grs enzyme activity under different temperature or pH conditions and by potential reworking of GDGT cyclization by as yet unidentified enzymes. Overall, acid and cold stress both affected the regulation of *grsB* in *S. islandicus*, but only cold stress conditions resulted in differential *grsB* expression and GDGT-level changes consistent with GrsB function. Our results suggest that transcriptional data may be helpful for predicting GDGT-level changes in cold stress conditions, but do not reflect Grs enzyme activity under all stress conditions.

Differential gene expression of other GDGT biosynthesis genes suggested that *S. islandicus* may have adapted its membrane to acid and cold stress by means other than GDGT cyclization. For example, the polar headgroups attached to membrane lipids can play a role in controlling membrane permeability (Shimada et al., 2008; Oger and Cario, 2013). Although we did not quantify headgroups in our experiments and despite only a few known headgroup synthesis genes, our RNA-seq data indicated that *S. islandicus* upregulated AIP phosphatase (SiRe_0209) under acid stress, suggesting that the proportion of archaetidylinositol headgroups may have increased in response to acidic conditions. We also observed differential expression of two sequential GDGT

biosynthesis genes in response to acid and cold stress, though the directionality of their expression changes differed. *S. islandicus* upregulated GGPP synthase (SiRe_1931) and downregulated GGGP synthase (SiRe_1736) (Figure 5). The downregulation of GGGP synthase and the lack of differential expression changes of numerous other GDGT synthesis genes (Figure 5) suggest that *S. islandicus* did not increase overall core lipid synthesis in response to acid or cold stress. However, the upregulation of GGPP synthase suggests increased production of GGPP, which is an intermediate for the syntheses of multiple polyterpenes beyond just core lipids. A potentially increased supply of GGPP could have been funneled toward the synthesis of polyterpenes like quinones and polyprenols, which have been proposed as archaeal membrane regulators for environmental stress response (Salvador-Castell et al., 2019). Broader characterizations of archaeal membrane components in response to environmental stressors could help shed light on these other adaptative mechanisms.

4.2. Broader transcriptional responses to acid or cold stress

RNA-seq enabled a comprehensive overview of *S. islandicus*' acid and cold stress responses beyond just lipid biosynthesis. Overall, *S. islandicus* differentially expressed many more genes during early stationary phase than mid-log phase under both stresses (Figure 4), possibly related to prolonged exposure to a given stressor and/or onset of growth limitation. These observations show the value in analyzing stress-induced expression changes at multiple growth phases and at later growth phases in particular. We also found that both acid and cold stress induced many of the same transcriptional responses in *S. islandicus*, which we propose to represent candidate general stress response genes that can be examined in future studies testing different environmental stressors. One common transcriptional response to both acid and cold stress included downregulation of ETC terminal oxidases, suggesting decreased oxidative phosphorylation and thereby a reduction in ATP generation and proton pumping. The drop in ATP supply would stymie energy-requiring processes, which is consistent with the slower growth rates we observed under both stress conditions (Figure 2). In addition to a slowed growth rate, acid stress also caused a significant decrease in maximum cell density not seen in cold stress conditions (Figure 2). Decreased growth yield suggests that *S. islandicus* acid stress response was energetically costly and/or that *S. islandicus* could not successfully counteract the negative effects of acid stress.

Cell motility also decreased in response to both stressors. Expression changes of the archaellum and related regulator genes indicated that *S. islandicus* decreased archaellum activity in response to both acid and cold stresses, contrasting the increased motility response by *S. acidocaldarius* to nutrient limitation (Szabó et al., 2007; Lassak et al., 2012; Haurat et al., 2017; Hoffmann et al., 2017; Bischof et al., 2019). This difference could indicate that unlike nutrient limitation, cells facing pH or temperature stress may conserve energy rather than swimming to find more favorable conditions.

Saccharolobus islandicus upregulated other cellular processes as common responses to acid and cold stress. Transcription

and translation activities increased, based on upregulation of genes encoding for RNA polymerase, ribosomal proteins, and translation initiation factors. Differential expression of VapBC toxin-antitoxin systems also indicated changes in translation across our experimental conditions. Upregulated HRR genes and purine synthesis genes suggest increased DNA repair activities, and increased expression of anaplerotic reaction genes suggests increased biosynthetic activity. Accordingly, DEGs supported increased biosynthesis of polyamines, histidine, cobalamin, and chorismate in response to both stressors.

The upregulation of polyamine synthesis under both acid and cold stress was notable given that polyamines are commonly referenced as cytoplasmic buffers and thereby an acid-specific adaptation (e.g., Slonczewski et al., 2009; Lewis et al., 2021). While upregulation of polyamine synthesis genes may have ameliorated cytoplasmic acidification in acid stress conditions, the advantage of increased cytoplasmic buffering is less clear under cold stress. As such, we suggest that polyamines may have contributed to the acid and cold stress responses of *S. islandicus* by some other mechanism(s). For example, polyamines are known to play an important role in translation as they are essential for ribosome activity (Londei et al., 1986). Given that observed DEGs indicated increased translation in response to both acid and cold stress, we propose that *S. islandicus* could have increased polyamine synthesis to support translation activity. Overall, upregulated polyamine synthesis as response to both stressors suggest that polyamines may contribute to the *S. islandicus* acid and cold stress responses by multiple mechanisms.

Low pH induced a strong physiological response in *S. islandicus* relative to cold temperature. The number of DEGs, magnitude of RI changes, and magnitude of cell density impairment were all greater in response to acidic stress. Consistent with the damage to protein and nucleic acids that can result from cytoplasmic acidification (Lund et al., 2014), acid-specific DEGs showed that *S. islandicus* upregulated genes related to degradation and repair processes for molecules including proteins, nucleic acids, and fatty acids. One of the most notable acid-specific transcriptional responses involved upregulation of an A-type ATPase that we suggest functions as a proton pump based on evidence from enzyme purification and gene sequence studies (Konishi et al., 1987; Lübbers and Schäfer, 1987; Hochstein and Stan-Lotter, 1992; Grüber et al., 2014). Proton pumping expels protons from the cytoplasm and is a common mechanism to combat pH stress (Baker-Austin and Dopson, 2007; Slonczewski et al., 2009). Terminal oxidases of the ETC also function as proton pumps, but *S. islandicus* downregulated all three of its terminal oxidases under acid stress suggesting a need for an alternate proton pump like the A-type ATPase. Altogether, the acid-specific DEGs such as the A-type ATPase provide insight to the ways that *S. islandicus* is affected by and responds to decreased pH.

The insertion sequence (IS) DEGs observed in our experiments notably showed opposite changes in expression between acid and cold stress conditions. Under both stressors, *S. islandicus* differentially expressed IS elements belonging to the IS200/605, ISH3, or ISH6 families of transposable elements, which are restricted to, or ancestral to, archaea (Filée et al., 2007). Acid stress induced consistent downregulation of these transposable elements whereas cold stress induced almost entirely upregulation shifts. These observations suggest that transposition frequency in

S. islandicus may increase under cold stress and decrease under acid stress.

5. Conclusion

Our study characterizes *S. islandicus* GDGT lipid and expression level responses to acid stress and cold stress. Both stress conditions led to decreased GDGT cyclization and differential expression of *grsB*. Our observations of *grsB* downregulation and decreased GDGT cyclization in response to cold stress show that *S. islandicus* modifies its GDGT profile and GDGT ring synthase transcription consistently with cold temperature responses of closely related thermoacidophilic Crenarchaea (e.g., De Rosa et al., 1980a; Jensen S. et al., 2015; Cobban et al., 2020; Yang et al., 2023). Acid stress induced upregulation of *grsB* in *S. islandicus* also supports previous observations of pH affecting *grsB* transcription in *S. acidocaldarius* (Yang et al., 2023). However, our observations of decreased cyclization in acid stress lipid profiles, despite *grsB* upregulation, was counter to our predictions that *S. islandicus* might increase GDGT cyclization to reduce membrane permeability and limit passive proton diffusion. These results suggest that higher GDGT cyclization is not or perhaps cannot always be employed as an acid stress response strategy by acidophilic archaea. Based on transcriptional evidence, we suggest that *S. islandicus* may selectively allocate ATP consumption toward other acid stress response mechanisms aside from GDGT cyclization. Acid-stressed *S. islandicus* only upregulated *grsB* during mid-log phase while other acid response mechanisms such as proton pumping and molecular repair were actively upregulated during mid-log and/or early stationary phase. Decreased GDGT cyclization and increased transcription of alternate pH homeostasis responses indicate that membrane modification may not be the primary mechanism enabling *S. islandicus* to survive acid stress. Together this work demonstrates how *S. islandicus* modifies its GDGTs and expression patterns in response to pH and temperature shifts. Future archaeal stress response studies using observations of both core and intact polar membrane lipid compositions, gene expression, and protein expression will help broaden our understanding of the intracellular factors that lead to archaeal membrane adaptations in response to environmental perturbations.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/[Supplementary material](#).

Author contributions

BC and WL conceptualized and designed the experiments. BC performed the experiments and RNA and lipid extractions. LZ and JW performed the protein extractions, proteomics analysis, and proteomic data generation. FE and AP performed the lipid analysis

and lipid data interpretation. BC and ÖM analyzed the expression data. BC and WL wrote the manuscript with contributions from JW and editorial expertise from EE. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by proposal: 10.46936/10.25585/60000984 (BC and WL) conducted by the U.S. Department of Energy Joint Genome Institute (<https://ror.org/04xm1d337>), a DOE Office of Science User Facility, is supported by the Office of Science of the U.S. Department of Energy operated under Contract No. DE-AC02-05CH11231. Other support was provided by the Simons Foundation, Award Number: 623881 (WL), the NSF-EAR Award #1928309 (WL), NSF Award #1843285 (AP), and the Junior Research Scholarship via Undergraduate Research and Advising at Dartmouth College (ÖM).

Acknowledgments

We thank Dr. Changyi Zhang and Dr. Rachel Whitaker for their gift of strain REY15A. We also thank Dr. Olga Zhaxybayeva and Dr. Zhengshuang Hua for RNA-seq analysis discussions, Dr. Yujiao Zhang and Dr. Yuki Weber for lipid extraction assistance, Dr. Claire Healy-Leavitt for initial data wrangling help, Fiona Harrigan for preliminary experimental assistance, and Dartmouth College WISP Freshman interns, Soyeon Cho, Madison Spivak, Crystal Igwe for participating in data exploration, and Middlebury College students Evan Fedorov and Amanda Girod for assistance with initial RNA processing and transcript verification.

Conflict of interest

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.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2023.1219779/full#supplementary-material>

References

- Albers, S.-V., and Jarrell, K. F. (2015). The archaeellum: How Archaea swim. *Front. Microbiol.* 6:23. doi: 10.3389/fmicb.2015.00023
- Arias-Cartin, R., Uzel, A., Seduk, F., Gerbaud, G., Pierrel, F., Broc, M., et al. (2022). Identification and characterization of a noncanonical menaquinone-linked formate dehydrogenase. *J. Biol. Chem.* 298:101384.
- Baker-Austin, C., and Dopson, M. (2007). Life in acid: PH homeostasis in acidophiles. *Trends Microbiol.* 15, 165–171. doi: 10.1016/j.tim.2007.02.005
- Banerjee, A., Tsai, C. L., Chaudhury, P., Tripp, P., Arvai, A. S., Ishida, J. P., et al. (2015). FlaF is a β -sandwich protein that anchors the archaeellum in the archaeal cell envelope by binding the S-layer protein. *Structure* 23, 863–872. doi: 10.1016/j.str.2015.03.001
- Becker, K. W., Lipp, J. S., Zhu, C., Liu, X. L., and Hinrichs, K. W. (2013). An improved method for the analysis of archaeal and bacterial ether core lipids. *Organ. Geochem.* 61, 33–44. doi: 10.1016/j.orggeochem.2013.05.007
- Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B* 57, 289–300. doi: 10.1111/j.2517-6161.1995.tb02031.x
- Benninghoff, J. C., Kuschmierz, L., Zhou, X., Albersmeier, A., Pham, T. K., Busche, T., et al. (2021). Exposure to 1-butanol exemplifies the response of the thermoacidophilic archaeon *Sulfolobus acidocaldarius* to solvent stress. *Appl. Environ. Microbiol.* 87:e02988–20. doi: 10.1128/AEM.02988-20
- Bischof, L. F., Haurat, M. F., Hoffmann, L., Albersmeier, A., Wolf, J., Neu, A., et al. (2019). Early response of *Sulfolobus acidocaldarius* to nutrient limitation. *Front. Microbiol.* 9:3201. doi: 10.3389/fmicb.2018.03201
- Boyd, E. S., Pearson, A., Pi, Y., Li, W.-J., Zhang, Y. G., He, L., et al. (2011). Temperature and pH controls on glycerol dibiphytanyl glycerol tetraether lipid composition in the hyperthermophilic crenarchaeon *Acidilobus sulfurireducens*. *Extremophiles* 15, 59–65. doi: 10.1007/s00792-010-0339-y
- Buetti-Dinh, A., Dethlefsen, O., Friedman, R., and Dopson, M. (2016). Transcriptomic analysis reveals how a lack of potassium ions increases *Sulfolobus acidocaldarius* sensitivity to pH changes. *Microbiology* 162, 1422–1434. doi: 10.1099/mic.0.000314
- Cobban, A., Zhang, Y., Zhou, A., Weber, Y., Elling, F. J., Pearson, A., et al. (2020). Multiple environmental parameters impact lipid cyclization in *Sulfolobus acidocaldarius*. *Environ. Microbiol.* 22, 4046–4056. doi: 10.1111/1462-2920.15194
- Cooper, C. R., Daugherty, A. J., Tachdjian, S., Blum, P. H., and Kelly, R. M. (2009). Role of *vapBC* toxin-antitoxin loci in the thermal stress response of *Sulfolobus solfataricus*. *Biochem. Soc. Trans.* 37, 123–126. doi: 10.1042/BST0370123
- Cooper, C. R., Lewis, A. M., Notey, J. S., Mukherjee, A., Willard, D. J., Blum, P. H., et al. (2023). Interplay between transcriptional regulators and VapBC toxin-antitoxin loci during thermal stress response in extremely thermoacidophilic Archaea. *Environ. Microbiol.* 25, 1200–1215. doi: 10.1111/1462-2920.16350
- Counts, J. A., Willard, D. J., and Kelly, R. M. (2021). Life in hot acid: A genome-based reassessment of the archaeal order Sulfolobales. *Environ. Microbiol.* 23, 3568–3584. doi: 10.1111/1462-2920.15189
- Crane, B. R., and Getzoff, E. D. (1996). The relationship between structure and function for the sulfite reductases. *Curr. Opin. Struct. Biol.* 6, 744–756. doi: 10.1016/S0959-440X(96)80003-0
- Daiyasu, H., Kuma, K., Yokoi, T., Morii, H., Koga, Y., and Toh, H. (2005). A study of archaeal enzymes involved in polar lipid synthesis linking amino acid sequence information, genomic contexts and lipid composition. *Archaea* 1, 399–410. doi: 10.1155/2005/452563
- De Rosa, M., and Gambacorta, A. (1988). The lipids of archaeobacteria. *Progr. Lipid Res.* 27, 153–175. doi: 10.1016/0163-7827(88)90011-2
- De Rosa, M., Esposito, E., Gambacorta, A., Nicolaus, B., and Bu'Lock, J. D. (1980a). Effects of temperature on ether lipid composition of *Caldariella acidophila*. *Phytochemistry* 19, 827–831. doi: 10.1016/0031-9422(80)85120-X
- De Rosa, M., Gambacorta, A., Nicolaus, B., and Bu'Lock, J. D. (1980b). Complex lipids of *Caldariella acidophila*, a thermoacidophile archaeobacterium. *Phytochemistry* 19, 821–825. doi: 10.1016/0031-9422(80)85119-3
- Elling, F. J., Könneke, M., Lipp, J. S., Becker, K. W., Gagen, E. J., and Hinrichs, K.-U. (2014). Effects of growth phase on the membrane lipid composition of the thaumarchaeon *Nitrosopumilus maritimus* and their implications for archaeal lipid distributions in the marine environment. *Geochim. Cosmochim. Acta* 141, 579–597. doi: 10.1016/j.gca.2014.07.005
- Elling, F. J., Könneke, M., Mußmann, M., Greve, A., and Hinrichs, K.-U. (2015). Influence of temperature, pH, and salinity on membrane lipid composition and TEX₈₆ of marine planktonic thaumarchaeal isolates. *Geochim. Cosmochim. Acta* 171, 238–255. doi: 10.1016/j.gca.2015.09.004
- Erde, J., Loo, R. R. O., and Loo, J. A. (2014). Enhanced FASP (eFASP) to increase proteome coverage and sample recovery for quantitative proteomic experiments. *J. Proteome Res.* 13, 1885–1895.
- Feyhl-Buska, J., Chen, Y., Jia, C., Wang, J.-X., Zhang, C. L., and Boyd, E. S. (2016). Influence of growth phase, pH, and temperature on the abundance and composition of tetraether lipids in the thermoacidophile *Picrophilus torridus*. *Front. Microbiol.* 7:1323. doi: 10.3389/fmicb.2016.01323
- Filée, J., Siguier, P., and Chandler, M. (2007). Insertion sequence diversity in Archaea. *Microbiol. Mol. Biol. Rev.* 71, 121–157. doi: 10.1128/MMBR.00031-06
- Forterre, P. (2002). A hot story from comparative genomics: Reverse gyrase is the only hyperthermophile-specific protein. *Trends Genet.* 18, 236–237. doi: 10.1016/S0168-9525(02)02650-1
- Fröls, S., Gordon, P. M. K., Panlilio, M. A., Duggin, I. G., Bell, S. D., Sensen, C. W., et al. (2007). Response of the hyperthermophilic archaeon *Sulfolobus solfataricus* to UV damage. *J. Bacteriol.* 189, 8708–8718. doi: 10.1128/JB.01016-07
- Gabriel, J. L., and Chong, P. L. (2000). Molecular modeling of archaeobacterial bipolar tetraether lipid membranes. *Chem. Phys. Lipids* 105, 193–200. doi: 10.1016/S0009-3084(00)00126-2
- Giozzio, A., Paoli, G., De Rosa, M., and Gambacorta, A. (1983). Effect of isoprenoid cyclization on the transition temperature of lipids in thermophilic archaeobacteria. *Biochim. Biophys. Acta* 735, 234–242. doi: 10.1016/0005-2736(83)90298-5
- Grüber, G., Manimekalai, M. S. S., Mayer, F., and Müller, V. (2014). ATP synthases from Archaea: The beauty of a molecular motor. *Biochim. Biophys. Acta* 1837, 940–952. doi: 10.1016/j.bbabo.2014.03.004
- Guo, L., Brügger, K., Liu, C., Shah, S. A., Zheng, H., Zhu, Y., et al. (2011). Genome analyses of icelandic strains of *Sulfolobus islandicus*, model organisms for genetic and virus-host interaction studies. *J. Bacteriol.* 193, 1672–1680. doi: 10.1128/JB.01487-10
- Haurat, M. F., Figueiredo, A. S., Hoffmann, L., Li, L., Herr, K., Wilson, A. J., et al. (2017). ArnS, a kinase involved in starvation-induced archaeellum expression. *Mol. Microbiol.* 103, 181–194. doi: 10.1111/mmi.13550
- Hendriks, J., Oubrie, A., Castresana, J., Urbani, A., Gemeinhardt, S., and Saraste, M. (2000). Nitric oxide reductases in bacteria. *Biochim. Biophys. Acta* 1459, 266–273. doi: 10.1016/S0005-2728(00)00161-4
- Hochstein, L. I., and Stan-Lotter, H. (1992). Purification and properties of an ATPase from *Sulfolobus solfataricus*. *Arch. Biochem. Biophys.* 295, 153–160. doi: 10.1016/0003-9861(92)90501-M
- Hoffmann, L., Schummer, A., Reimann, J., Haurat, M. F., Wilson, A. J., Beeby, M., et al. (2017). Expanding the archaeellum regulatory network – the eukaryotic protein kinases ArnC and ArnD influence motility of *Sulfolobus acidocaldarius*. *Microbiologyopen* 6:e00414. doi: 10.1002/mbo3.414
- Hurley, S. J., Elling, F. J., Könneke, M., Buchwald, C., Wankel, S. D., Santoro, A. E., et al. (2016). Influence of ammonia oxidation rate on thaumarchaeal lipid composition and the TEX₈₆ temperature proxy. *Proc. Natl. Acad. Sci. U.S.A.* 113, 7762–7767. doi: 10.1073/pnas.1518534113
- Jain, S., Caforio, A., and Driessen, A. J. (2014). Biosynthesis of archaeal membrane ether lipids. *Front. Microbiol.* 5:641. doi: 10.3389/fmicb.2014.00641
- Jensen, S. M., Brandl, M., Treusch, A. H., and Ejsing, C. S. (2015). Structural characterization of ether lipids from the archaeon *Sulfolobus islandicus* by high-resolution shotgun lipidomics: Mapping the lipidome of *Sulfolobus islandicus*. *J. Mass Spectrom.* 50, 476–487. doi: 10.1002/jms.3553
- Jensen, S., Neesgaard, V., Skjoldbjerg, S., Brandl, M., Ejsing, C., and Treusch, A. (2015). The effects of temperature and growth phase on the lipidomes of *Sulfolobus islandicus* and *Sulfolobus tokodaii*. *Life* 5, 1539–1566. doi: 10.3390/life5031539
- Kampmann, M., and Stock, D. (2004). Reverse gyrase has heat-protective DNA chaperone activity independent of supercoiling. *Nucleic Acids Res.* 32, 3537–3545. doi: 10.1093/nar/gkh683
- Kim, D., Langmead, B., and Salzberg, S. L. (2015). HISAT: A fast spliced aligner with low memory requirements. *Nat. Methods* 12, 357–360. doi: 10.1038/nmeth.3317
- Klumpp, M., and Baumeister, W. (1998). The thermosome: Archetype of group II chaperonins. *FEBS Lett.* 430, 73–77. doi: 10.1016/S0014-5793(98)00541-9
- Koga, Y., and Morii, H. (2007). Biosynthesis of ether-type polar lipids in Archaea and evolutionary considerations. *Microbiol. Mol. Biol. Rev.* 71, 97–120. doi: 10.1128/MMBR.00033-06
- Komori, K., Sakae, S., Shinagawa, H., Morikawa, K., and Ishino, Y. (1999). A Holliday junction resolvase from *Pyrococcus furiosus*: Functional similarity to *Escherichia coli* RuvC provides evidence for conserved mechanism of homologous recombination in Bacteria, Eukarya, and Archaea. *Proc. Natl. Acad. Sci. U.S.A.* 96, 8873–8878. doi: 10.1073/pnas.96.16.8873
- Konishi, J., Wakagi, T., Oshima, T., and Yoshida, M. (1987). Purification and properties of the ATPase solubilized from membranes of an acidothermophilic archaeobacterium, *Sulfolobus acidocaldarius*. *J. Biochem.* 102, 1379–1387. doi: 10.1093/oxfordjournals.jbchem.a122184
- Kusmierczyk, A. R., Kunjappu, M. J., Kim, R. Y., and Hochstrasser, M. (2011). A conserved 20S proteasome assembly factor requires a C-terminal HbYX motif for

- proteasomal precursor binding. *Nat. Struct. Mol. Biol.* 18, 622–629. doi: 10.1038/nsmb.2027
- Kvaratskhelia, M., and White, M. (2000a). An archaeal Holliday junction resolving enzyme from *Sulfolobus solfataricus* exhibits unique properties. *J. Mol. Biol.* 295, 193–202. doi: 10.1006/jmbi.1999.3363
- Kvaratskhelia, M., and White, M. (2000b). Two Holliday junction resolving enzymes in *Sulfolobus solfataricus*. *J. Mol. Biol.* 297, 923–932. doi: 10.1006/jmbi.2000.3624
- Langworthy, T. A. (1977). Comparative lipid composition of heterotrophically and autotrophically grown *Sulfolobus acidocaldarius*. *J. Bacteriol.* 130, 1326–1332. doi: 10.1128/jb.130.3.1326-1332.1977
- Lassak, K., Neiner, T., Ghosh, A., Klingl, A., Wirth, R., and Albers, S.-V. (2012). Molecular analysis of the crenarchaeal flagellum. *Mol. Microbiol.* 83, 110–124. doi: 10.1111/j.1365-2958.2011.07916.x
- Lassak, K., Peeters, E., Wróbel, S., and Albers, S.-V. (2013). The one-component system ArnR: A membrane-bound activator of the crenarchaeal archaeallum. *Mol. Microbiol.* 88, 125–139. doi: 10.1111/mmi.12173
- Law, K. P., Li, X., and Zhang, C. (2020). Lipidomics in archaeal membrane adaptation to environmental stresses and growth conditions: A review of culture-based physiological studies. *Sci. China Earth Sci.* 63, 790–807. doi: 10.1007/s11430-019-9571-2
- Lewis, A. M., Recalde, A., Bräsen, C., Counts, J. A., Nussbaum, P., Bost, J., et al. (2021). The biology of thermoacidophilic Archaea from the order *Sulfolobales*. *FEMS Microbiol. Rev.* 45:fuaa063. doi: 10.1093/femsre/fuaa063
- Li, Z., Lu, S., Hou, G., Ma, X., Sheng, D., Ni, J., et al. (2008). Hjm/Hel308A DNA helicase from *Sulfolobus tokodaii* promotes replication fork regression and interacts with Hjc endonuclease *In Vitro*. *J. Bacteriol.* 190, 3006–3017. doi: 10.1128/JB.01662-07
- Liao, Y., Smyth, G. K., and Shi, W. (2014). featureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30, 923–930. doi: 10.1093/bioinformatics/btt656
- Lloyd, C. T., Iwig, D. F., Wang, B., Cossu, M., Metcalf, W., Boal, A., et al. (2022). Discovery, structure and mechanism of a tetraether lipid synthase. *Nature* 609, 197–203. doi: 10.1038/s41586-022-05120-2
- Londei, P., Teixidó, J., Acca, M., Cammarano, P., and Amils, R. (1986). Total reconstitution of active large ribosomal subunits of the thermoacidophilic archaeobacterium *Sulfolobus solfataricus*. *Nucleic Acids Res.* 14, 2269–2285.
- Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15:550. doi: 10.1186/s13059-014-0550-8
- Lübbers, M., and Schäfer, G. (1987). A plasma-membrane associated ATPase from the thermoacidophilic archaeobacterium *Sulfolobus acidocaldarius*. *Eur. J. Biochem.* 164, 533–540. doi: 10.1111/j.1432-1033.1987.tb11159.x
- Lund, P., Tramonti, A., and De Biase, D. (2014). Coping with low pH: Molecular strategies in neutrophilic bacteria. *FEMS Microbiol. Rev.* 38, 1091–1125. doi: 10.1111/1574-6976.12076
- Maezato, Y., Daugherty, A., Dana, K., Soo, E., Cooper, C., Tachdjian, S., et al. (2011). VapC6, a ribonucleolytic toxin regulates thermophilicity in the crenarchaeote *Sulfolobus solfataricus*. *RNA* 17, 1381–1392. doi: 10.1261/rna.2679911
- McCarthy, S., Johnson, T., Pavlik, B. J., Payne, S., Schackwitz, W., Martin, J., et al. (2016). Expanding the limits of thermoacidophily in the archaeon *Sulfolobus solfataricus* by adaptive evolution. *Appl. Environ. Microbiol.* 82, 857–867. doi: 10.1128/AEM.03225-15
- Meyer, B. H., Peyfoon, E., Dietrich, C., Hitchen, P., Panico, M., Morris, H. R., et al. (2013). Agl16, a thermophilic glycosyltransferase mediating the last step of N-glycan biosynthesis in the thermoacidophilic crenarchaeon *Sulfolobus acidocaldarius*. *J. Bacteriol.* 195, 2177–2186.
- Middleton, C. L., Parker, J. L., Richard, D. J., White, M. F., and Bond, C. S. (2004). Substrate recognition and catalysis by the Holliday junction resolving enzyme Hje. *Nucleic Acids Res.* 32, 5442–5451. doi: 10.1093/nar/gkh869
- Morii, H., Kiyonari, S., Ishino, Y., and Koga, Y. (2009). A novel biosynthetic pathway of arachidyl-myo-inositol via arachidyl-myo-inositol Phosphate from CDP-archaeol and d-Glucose 6-Phosphate in methanococcal *Methanothermobacter thermautotrophicus* cells. *J. Biol. Chem.* 284, 30766–30774. doi: 10.1074/jbc.M109.034652
- Mukherjee, A., Wheaton, G. H., Counts, J. A., Ijeomah, B., Desai, J., and Kelly, R. M. (2017). VapC toxins drive cellular dormancy under uranium stress for the extreme thermoacidophile *Metallosphaera prunae*. *Environ. Microbiol.* 19, 2831–2842. doi: 10.1111/1462-2920.13808
- Oger, P. M., and Cario, A. (2013). Adaptation of the membrane in Archaea. *Biophys. Chem.* 183, 42–56. doi: 10.1016/j.bpc.2013.06.020
- Orell, A., Peeters, E., Vassen, V., Jachlewski, S., Schalles, S., Siebers, B., et al. (2013). Lrs14 transcriptional regulators influence biofilm formation and cell motility of Crenarchaea. *ISME J.* 7, 1886–1898. doi: 10.1038/ismej.2013.68
- Pearson, A., and Ingalls, A. E. (2013). Assessing the use of archaeal lipids as marine environmental proxies. *Annu. Rev. Earth Planet. Sci.* 41, 359–384. doi: 10.1146/annurev-earth-050212-123947
- Portnoy, V., Evgueniev-Hackenberg, E., Klein, F., Walter, P., Lorentzen, E., Klug, G., et al. (2005). RNA polyadenylation in Archaea: Not observed in *Haloflex* while the exosome polynucleotidylates RNA in *Sulfolobus*. *EMBO Rep.* 6, 1188–1193. doi: 10.1038/sj.embor.7400571
- Quaiser, A., Constantinesco, F., White, M. F., Forterre, P., and Elie, C. (2008). The Mre11 protein interacts with both Rad50 and the HerA bipolar helicase and is recruited to DNA following gamma irradiation in the archaeon *Sulfolobus acidocaldarius*. *BMC Mol. Biol.* 9:25. doi: 10.1186/1471-2199-9-25
- Quehenberger, J., Pittenauer, E., Allmaier, G., and Spadiut, O. (2020). The influence of the specific growth rate on the lipid composition of *Sulfolobus acidocaldarius*. *Extremophiles* 24, 413–420. doi: 10.1007/s00792-020-01165-1
- Ramírez, F., Dündar, F., Diehl, S., Grüning, B. A., and Manke, T. (2014). deepTools: A flexible platform for exploring deep-sequencing data. *Nucleic Acids Res.* 42, W187–W191. doi: 10.1093/nar/gku365
- Rastädter, K., Wurm, D. J., Spadiut, O., and Quehenberger, J. (2020). The cell membrane of *Sulfolobus* spp.—Homeoviscous adaption and biotechnological applications. *Int. J. Mol. Sci.* 21:3935. doi: 10.3390/ijms21113935
- Salvador-Castell, M., Tourte, M., and Oger, P. M. (2019). In search for the membrane regulators of Archaea. *Int. J. Mol. Sci.* 20:4434. doi: 10.3390/ijms20184434
- Schouten, S., Hopmans, E. C., and Sinninghe Damsté, J. S. (2013). The organic geochemistry of glycerol dialkyl glycerol tetraether lipids: A review. *Organ. Geochem.* 54, 19–61. doi: 10.1016/j.orggeochem.2012.09.006
- Shimada, H., Nemoto, N., Shida, Y., Oshima, T., and Yamagishi, A. (2008). Effects of pH and temperature on the composition of polar lipids in *Thermoplasma acidophilum* HO-62. *J. Bacteriol.* 190, 5405–5411. doi: 10.1128/JB.00415-08
- Shinoda, W., Shinoda, K., Baba, T., and Mikami, M. (2005). Molecular dynamics study of bipolar tetraether lipid membranes. *Biophys. J.* 89, 3195–3202. doi: 10.1529/biophysj.105.060962
- Siliakus, M. F., van der Oost, J., and Kengen, S. W. M. (2017). Adaptations of archaeal and bacterial membranes to variations in temperature, pH and pressure. *Extremophiles* 21, 651–670. doi: 10.1007/s00792-017-0939-x
- Slonczewski, J. L., Fujisawa, M., Dopson, M., and Krulwich, T. A. (2009). Cytoplasmic pH measurement and homeostasis in bacteria and Archaea. *Adv. Microb. Physiol.* 55, 1–79, 317. doi: 10.1016/S0065-2911(09)05501-5
- Song, X., Huang, Q., Ni, J., Yu, Y., and Shen, Y. (2016). Knockout and functional analysis of two DEXD/H-box family helicase genes in *Sulfolobus islandicus* REY15A. *Extremophiles* 20, 537–546. doi: 10.1007/s00792-016-0847-5
- Szabó, Z., Sani, M., Groeneveld, M., Zolghadr, B., Schelert, J., Albers, S.-V., et al. (2007). Flagellar motility and structure in the hyperthermoacidophilic archaeon *Sulfolobus solfataricus*. *J. Bacteriol.* 189, 4305–4309. doi: 10.1128/JB.00042-07
- Tsai, C., Tripp, P., Sivabalasarma, S., Zhang, C., Rodriguez-Franco, M., Wipfler, R., et al. (2020). The structure of the periplasmic FlaG–FlaF complex and its essential role for archaeal swimming motility. *Nat. Microbiol.* 5, 216–225. doi: 10.1038/s41564-019-0622-3
- Uda, I., Sugai, A., Itoh, Y. H., and Itoh, T. (2001). Variation in molecular species of polar lipids from *Thermoplasma acidophilum* depends on growth temperature. *Lipids* 36, 103–105.
- Uda, I., Sugai, A., Itoh, Y. H., and Itoh, T. (2004). Variation in molecular species of core lipids from the order *Thermoplasmatales* strains depends on the growth temperature. *J. Oleo Sci.* 53, 399–404.
- Villanueva, L., Damsté, J. S. S., and Schouten, S. (2014). A re-evaluation of the archaeal membrane lipid biosynthetic pathway. *Nat. Rev. Microbiol.* 12, 438–448. doi: 10.1038/nrmicro3260
- Waldbauer, J., Zhang, L., Rizzo, A., and Muratore, D. (2017). diDO-IPITL: A peptide-labeling strategy for precision quantitative proteomics. *Analyt. Chem.* 89, 11498–11504. doi: 10.1021/acs.analchem.7b02752
- Walling, L. R., and Butler, J. S. (2019). Toxins targeting transfer RNAs: Translation inhibition by bacterial toxin–antitoxin systems. *Wiley Interdiscip. Rev. RNA* 10:e1506. doi: 10.1002/wrna.1506
- Wang, K., Sybers, D., Maklad, H. R., Lemmens, L., Lewyllie, C., Zhou, X., et al. (2019). A TetR-family transcription factor regulates fatty acid metabolism in the archaeal model organism *Sulfolobus acidocaldarius*. *Nat. Commun.* 10:1542. doi: 10.1038/s41467-019-09479-1
- Yang, W., Chen, H., Chen, Y., Chen, A., Feng, X., Zhao, B., et al. (2023). Thermophilic archaeon orchestrates temporal expression of GDGT ring synthases in response to temperature and acidity stress. *Environ. Microbiol.* 25, 575–587. doi: 10.1111/1462-2920.16301
- Zeng, Z., Chen, H., Yang, H., Chen, Y., Yang, W., Feng, X., et al. (2022). Identification of a protein responsible for the synthesis of archaeal membrane-spanning GDGT lipids. *Nat. Commun.* 13:1545. doi: 10.1038/s41467-022-29264-x

- Zeng, Z., Liu, X.-L., Farley, K. R., Wei, J. H., Metcalf, W. W., Summons, R. E., et al. (2019). GDGT cyclization proteins identify the dominant archaeal sources of tetraether lipids in the ocean. *Proc. Natl. Acad. Sci. U.S.A.* 116, 22505–22511. doi: 10.1073/pnas.1909306116
- Zeng, Z., Liu, X.-L., Wei, J. H., Summons, R. E., and Welander, P. V. (2018). Calditol-linked membrane lipids are required for acid tolerance in *Sulfolobus acidocaldarius*. *Proc. Natl. Acad. Sci. U.S.A.* 115, 12932–12937. doi: 10.1073/pnas.1814048115
- Zhang, C., Cooper, T. E., Krause, D. J., and Whitaker, R. J. (2013). Augmenting the genetic toolbox for *Sulfolobus islandicus* with a stringent positive selectable marker for agmatine prototrophy. *Appl. Environ. Microbiol.* 79, 5539–5549. doi: 10.1128/AEM.01608-13
- Zhang, C., Phillips, A. P. R., Wipfler, R. L., Olsen, G. J., and Whitaker, R. J. (2018). The essential genome of the crenarchaeal model *Sulfolobus islandicus*. *Nat. Commun.* 9:4908. doi: 10.1038/s41467-018-07379-4
- Zhou, A., Weber, Y., Chiu, B. K., Elling, F. J., Cobban, A. B., Pearson, A., et al. (2020). Energy flux controls tetraether lipid cyclization in *Sulfolobus acidocaldarius*. *Environ. Microbiol.* 22, 343–353. doi: 10.1111/1462-2920.14851
- Zillig, W., Holz, I., Janekovic, D., Schäfer, W., and Reiter, W. D. (1983). The archaeobacterium *Thermococcus celer* represents, a novel genus within the thermophilic branch of the archaeobacteria. *Syst. Appl. Microbiol.* 4, 88–94. doi: 10.1016/S0723-2020(83)80036-8



OPEN ACCESS

EDITED BY

Arthur Charles-Orszag,
University of California, San Francisco,
United States

REVIEWED BY

Haruyuki Atomi,
Kyoto University, Japan
Parkson Lee-Gau Chong,
Temple University, United States

*CORRESPONDENCE

Buzz Baum
✉ bbaum@mrc-lmb.cam.ac.uk

RECEIVED 01 June 2023

ACCEPTED 17 August 2023

PUBLISHED 05 September 2023

CITATION

Cezanne A, Hoogenberg B and Baum B (2023)
Probing archaeal cell biology: exploring the use
of dyes in the imaging of *Sulfolobus* cells.
Front. Microbiol. 14:1233032.
doi: 10.3389/fmicb.2023.1233032

COPYRIGHT

© 2023 Cezanne, Hoogenberg and Baum. 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.

Probing archaeal cell biology: exploring the use of dyes in the imaging of *Sulfolobus* cells

Alice Cezanne¹, Baukje Hoogenberg^{1,2} and Buzz Baum^{1*}

¹Cell Biology Division, MRC Laboratory of Molecular Biology, Cambridge, United Kingdom, ²Faculty of Science, Utrecht University, Utrecht, Netherlands

Archaea are key players in many critical ecological processes. In comparison to eukaryotes and bacteria, however, our understanding of both the cell biology and diversity of archaea remains limited. While archaea inhabit a wide range of environmental conditions, many species are extremophiles, surviving in extreme temperature, salt or pH conditions, making their cell biology hard to study. Recently, our understanding of archaeal cell biology has been advanced significantly by the advent of live cell imaging *in extremis* as well as the development of genetic tools to exogenously express fluorescent proteins in some mesophilic archaeal model systems, e.g., *Haloferax volcanii*. However, for most archaeal species, especially thermophilic species or emerging model systems without well characterized genetic tools, live cell imaging remains dependent on fluorescent chemical probes to label and track the dynamics of living cells. While a wide range of fluorescent stains and markers that label different components of the cell are available commercially, their use has usually been optimized for use in a small number of eukaryotic cell systems. Here we report the successes and failures of the application of membrane, DNA, S-layer and cytoplasm markers in live cell imaging of archaea, as well as the optimization of fixation and immunolabelling approaches. We have applied these markers to the thermoacidophilic archaeon *Sulfolobus acidocaldarius*, but expect some to work in other archaeal species. Furthermore, those procedures that failed in *S. acidocaldarius* may still prove useful for imaging archaea that grow at a more neutral pH and/or at a less extreme temperature.

KEYWORDS

archaea, fluorescent imaging, live cell imaging, hyperthermophiles, molecular probes

1. Introduction

Archaea were first proposed to be a distinct domain of prokaryotic life by Carl Woese and colleagues (Woese and Fox, 1977). Since then, microbiology and phylogenetic studies have greatly furthered our understanding of the tree of life, as well as the extreme diversity of archaeal species and the wide variety environmental niches they occupy. However, the genetic tools and imaging procedures required to study the cell biology of archaea remain much less well developed than those available for studies in bacteria and eukaryotes. The difficulties are compounded by the fact that many archaea are extremophiles.

Much of the core information processing and cytoskeletal machinery present in eukaryotes appears to have an archaeal origin (Lake et al., 1984; Spang et al., 2015; Zaremba-Niedzwiedzka et al., 2017). Given their close evolutionary relationship with eukaryotes, many archaea possess simpler counterparts of the core machinery found in eukaryotes. This includes machinery involved in genome organization (Peeters et al., 2015; Mattioli et al., 2017), DNA replication

initiation and its elongation (Barry and Bell, 2006), transcription (Werner, 2007), rRNA processing (Omer et al., 2000), N-linked glycosylation (Jarrell et al., 2014), the Ubiquitin-ESCRT (Endosomal Sorting Complexes Required for Transport)-proteosomal system (Zwickl et al., 1992; Nunoura et al., 2011; Hennell James et al., 2017; Hatano et al., 2022), and the actin cytoskeleton (Akil and Robinson, 2018; Rodrigues-Oliveira et al., 2023). Understanding the cell biology of the relatively simple archaeal counterparts of eukaryotic proteins machineries can shed new light on their origins and can reveal underlying principles that are obscured by the complexity of the machinery present in eukaryotes. While recent attention has focused on the Asgard archaea from within which eukaryotes likely emerged (Eme et al., 2023), thus far only a few members of the Asgard archaea have been successfully cultivated. Moreover, in these cases the cells are present in mixed cultures, which include syntrophic partners, and must be grown under anaerobic conditions (Imachi et al., 2020; Rodrigues-Oliveira et al., 2023). As a result, most cell biology studies that aim to use archaea as simple models to study eukaryotic protein machineries have focused on the related TACK superphyla (Thaum-, Aig-, Cren-, and Korarchaeota), whose members share fewer molecular features with eukaryotes than Asgard archaea, but are far more experimentally tractable. These have proved useful model systems with which to probe the minimal components needed for cellular processes in eukaryotes, as has been done for ESCRT-III dependent cytokinesis (Lindås et al., 2008; Samson et al., 2008; Pulschen et al., 2020; Tarrason Risa et al., 2020; Hurtig et al., 2023).

While many cellular processes are shared between archaea and eukaryotes, archaea also possess unique chemical and biochemical features that are not found in other domains of life. One of the most prominent examples of this is the archaeal membrane, which is composed of unique lipid structures consisting of isoprenoid chains linked to glycerol-1-phosphate backbones by an ether linkage (Koga and Morii, 2007). In contrast, bacteria and eukaryotes share a phospholipid composition of fatty acid chains, linked by an ester linkage to glycerol-3-phosphate backbones (Peretó et al., 2004). This phenomenon is commonly termed the lipid divide, and has important implications for the emergence of bacteria and archaea as distinct domains of life as well as for eukaryogenesis (Peretó et al., 2004; Koga, 2011; Lombard et al., 2012; Villanueva et al., 2021). In addition, as many archaea are extremophiles, they possess unique strategies to survive the harsh environmental conditions in which they live: be it high salt, high temperature or low oxygen. While this presents experimental challenges, as we will discuss below, understanding the cell biology of extremophiles has a wide range of applications, including within industry (e.g., in drug and vaccine delivery) and in the search for life on other planets (Patel and Sprott, 1999; Jacquemet et al., 2009; Merino et al., 2019).

As a cell biological tool, fluorescence microscopy has enabled the characterization of many biological processes in living cells (Shimomura et al., 1962; Zimmer, 2002). However, its application in the field of archaeal cell biology is still under development (Bisson-Filho et al., 2018; Pulschen et al., 2020; Charles-Orszag et al., 2021). This in part reflects the challenges one faces when working with archaea. The solubility and stability of fluorescent probes must be tested for a range of media conditions including high salt, low pH and high temperatures. Further, a number of archaeal species are anaerobic which presents a challenge for light microscopy applications. Due to differences in membrane architecture, it must further

be considered whether fluorescent probes can cross the membrane to reach cytoplasmic targets or whether membrane targeting probes designed for eukaryotes and bacteria can interact with the archaeal lipid membrane at all. As most eukaryotes are studied within a narrow range of environmental conditions (neutral pH, 22–37°C), only few fluorescent probes developed for use in eukaryotes have been tested at extreme conditions. 2-photon imaging of Laurdan, for example, has been successfully used to characterize reconstituted archaeal membranes at low pH (pH 2.68) and high temperature (up to 64°C) (Bagatolli et al., 2000), however methods for imaging intact archaeal cells as they grow and divide are still in their infancy.

Here we explore the use of fluorescent markers for live cell imaging of *S. acidocaldarius*, currently the most experimentally tractable relative of eukaryotes. *S. acidocaldarius* is a member of the TACK superphylum, grows at pH 3 and 75°C, has a well-established molecular genetic toolbox (Lewis et al., 2021), and an ordered cell-cycle similar to that of eukaryotes (Bernander, 2007). For this analysis we explored the use of commercially available fluorescent probes to label different components of *S. acidocaldarius* including the membrane, DNA, S-layer, membrane proteins and cytoplasm (Figure 1A), as well as the optimization of different fixation methods for immunolabelling. The results of these tests will be useful to the wider archaeal community, and we hope they will help to fuel a growing interest in the cell biology of archaea (van Wolferen et al., 2022).

2. Materials and methods

2.1. Cell culturing

S. acidocaldarius DSM 639 (wild-type) or MW001 (uracil auxotrophic cloning strain) were grown in a shaking incubator at 75°C in Brock medium pH 2.9 supplemented with 0.1% N-Z-amine and 0.2% sucrose. MW001 cultures were supplemented with 4 µg/mL uracil. All cultures used for imaging were collected during exponential growth phase at an optical density at 600 nm (OD_{600nm}) of 0.1 to ~0.4.

2.2. Cell labelling

All dyes in listed in Tables 1–3 were dissolved in DMSO and added to a final concentration of no greater than 0.1% (v/v). Note that much higher concentrations of DMSO (>10%) can be added to cultures without it interfering with growth. Staining was performed for either 5 min at either room temperature (23°C) or for 1 min at 75°C. The signal from dyes that stained cells over background during this time-frame did not greatly improve upon longer incubation times. Similarly, dyes that failed to label cells within 5 min at room temperature did not successfully stain cells over incubation times of up to 30 min. To settle on conditions for live imaging we tested candidate probes at different concentrations over different exposure times at room temperature. Probes that successfully stained cells at room temperature were then re-tested at the same concentrations at 75°C using different exposure times. Probes that successfully stained cells at 75°C were then tested for cytotoxicity (without illumination) by performing growth curves over 24 h in presence or absence of the dye at concentrations determined above. Only DiO(C6) was found to

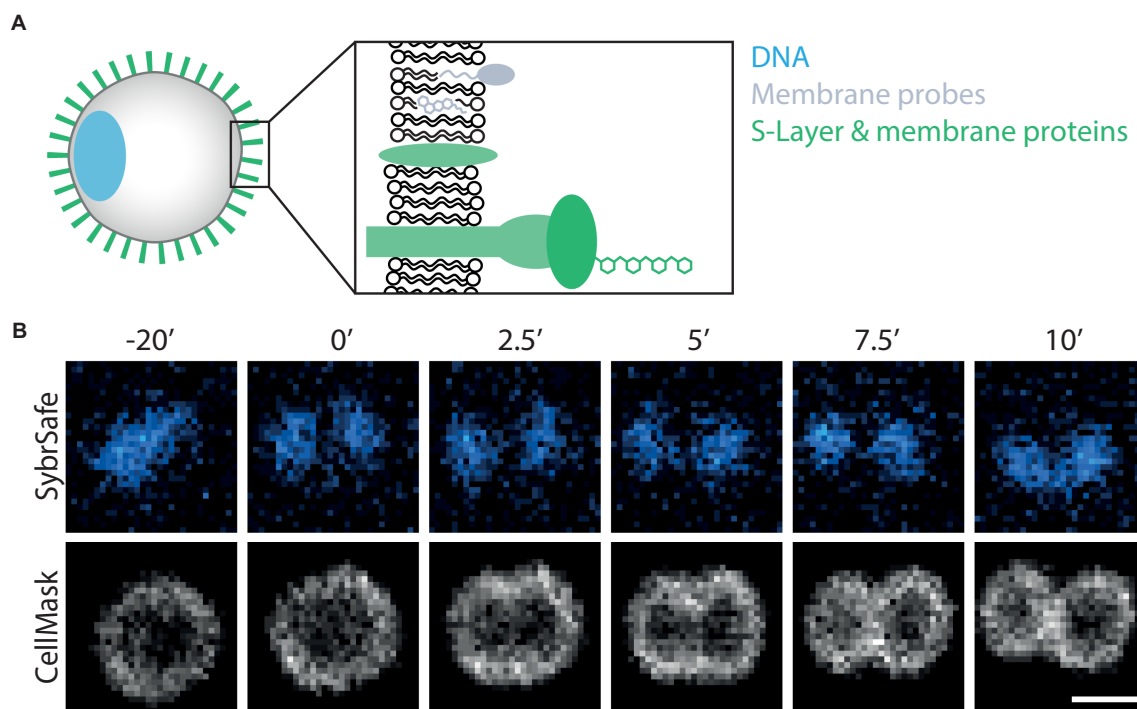


FIGURE 1

Visualizing *S. acidocaldarius*. (A) Schematic representation of labelling targets investigated in this study. (B) Time-lapse imaging of DNA and membrane dynamics in DSM 639 cells during cell division, time steps indicated in minutes above images. Cells were stained with SybrSafe and CellMask Deep Red Plasma Membrane stain and imaged at 75°C in Brock medium. Scale bar = 1 μ m.

be cytotoxic at concentrations used for live cell imaging at either room temperature or 75°C. Note that because cells were not washing following labelling, it likely that the pool of dye present in the medium can exchange with the cellular pool – reducing the effects of photobleaching.

2.3. Live cell-imaging at 75°C

Live-cell imaging was performed at 75°C using the “Sulfoscope” chamber described in Pulschen et al. (2020), with modifications to the hardware described by Hurtig et al. (2023). Briefly, 25 mm coverslips were washed with EtOH and H₂O, then assembled into commercial Attofluor chambers (Invitrogen A7816). Chambers were filled with 300 μ L of Brock medium and incubated at 75°C for at least 1 h or until the medium was dry. Afterwards, chambers were washed thoroughly with Brock medium, placed into the Sulfoscope chamber, and allowed to equilibrate to 75°C. Dyes detailed in Tables 1–3 were added to 5 mL of 75°C *S. acidocaldarius* cell culture (OD_{600nm} 0.15 to 0.3) immediately before imaging. For imaging, 400 μ L of cell suspension (OD_{600nm} 0.15 to 0.3) was added into the chamber and immobilized using heated, semi-solid gelrite pads (0.6% Gelrite, 0.5 \times Brock medium pH 5, and a final concentration of 20 mM CaCl₂). Pads were prepared in the following way: ~15 ml molten Gelrite Brock medium solution was added to 9cm plastic petri dishes, and allowed to set at room temperature (~5 minutes). Half-moon shapes were then cut from the plate with a 7 mm diameter circle punch, placed onto 13 mm circular coverslips, and incubated at 75°C for 5 to 10 min in a bead bath. During this period of incubation, pads equilibrated to the imaging

temperature and also dried slightly, causing the edges of the pad to curve downwards. Preheated pads were then placed in the chamber onto the cell suspension, such that the concave edge of the pad was in the center of the chamber. For the cell biological analysis, cells at the border of the immobilization pad were imaged since, in this area, cells are immobilized without being subjected to mechanical stress from the overlying Gelrite. Images were acquired on a Nikon Eclipse Ti2 inverted microscope equipped with a Yokogawa SoRa scanner unit and Prime 95B sCMOS camera (Photometrics). Imaging was performed with a 60 \times oil immersion objective (Plan Apo 60 \times /1.45, Nikon) using a custom formulated immersion oil for high temperature imaging (maximum refractive index matching at 70°C, $n = 1.515 \pm 0.0005$; Cargille Laboratories), using the $\times 2.8$ magnification of the SoRa unit (equivalent to a total magnification of $\times 168$). Images were acquired using a 15 ms exposure time and 10% laser power at intervals of 15 s for 2 to 3 h. After acquisition, XY drift was corrected using the ImageJ plugin StackReg (Thevenaz et al., 1998).

2.4. Imaging without fixation at room temperature

1 mL DSM 639 culture in Brock medium was allowed to cool to room temperature before staining with the dyes listed in Tables 1–3, as described above. Imaging was performed on cells confined using a 1% low melt agarose pad. In brief: a 1% low melting temperature agarose (Sigma Aldrich, A9414) was prepared in MilliQ water by microwaving until the agarose was completely dissolved. 100 μ L of molten agarose solution was pipetted onto a homemade spacer slide

TABLE 1 List of membrane labels used in this study.

Probe	$\lambda_{ex}/\lambda_{em}$ (nm)	Membrane Interaction	24°C (pH 3)		75°C (pH 3)		Toxic (75°C)
			Staining	Conc.	Staining	Conc.	
Nile Red	~552/636	Insertion into lipid core	Yes	5 µg/mL	Yes	2.5 µg/mL	No
CellMask™ Deep Red Plasma Membrane Stain	649/666	Lipophilic tail	Yes	0.5 µg/mL	Yes	1 µg/mL	No
CellMask™ Orange Plasma Membrane Stain	556/573	Lipophilic tail	Yes	0.5 µg/mL	Yes	1 µg/mL	No
Mitotracker® Green FM	490/516	Insertion into lipid core	Yes	0.336 µg/mL	Diffuse cytoplasmic signal		No
Mitotracker® Red CMXRos	579/599	Insertion into lipid core, Membrane potential dependent	Yes	0.265 µg/mL	Membrane and cytoplasmic signal	0.1325 µg/mL	No
BODIPY™ TR Ceramide	592/618	Insertion into lipid core	Weak signal	12.5 µg/mL	No		-
DiO C6	488/506	Either: intercalates between leaflets OR insertion or two lipid tails and fluorophore outside	Yes	2 µg/mL	No		Yes
DiO C18(3)	488/506		No		-		-
SP-DiO C18(3)	488/506		No		-		-
DiI C18(3)	550/570		No		-		-
DiI C18(3) DS	550/570		No		-		-
5,5'-Ph2-DiI C18(3)	550/570		No		-		-
SP-DiI C18(3)	550/570		No		-		-
DiA	450/585		No		-		-
DiR (DiI C18(7))	750/780		No		-		-
Mitotracker® Deep Red FM	644/665	Insertion into lipid core	No		No		-
CellMask™ Green Plasma Membrane Stain	522/535	Lipophilic tail	-		No		-
Nile Blue	626/668	Insertion into lipid core	No		No		-
FM™ 4-64X	~515/640	Anchored in outer leaflets, fluorescent in hydrophobic environments	No		No		-
FM™ 1-43FX	510/626	Anchored in outer leaflets, fluorescent in hydrophobic environments	No		No		-

Names colour coded according to success of staining. Green: optimal staining; Blue: conditionally applicable staining; Grey: no staining. Hyphens indicate that the condition was not tested.

consisting of a glass slide with 4 layers of lab tape wrapped around either end. A second glass slide was then placed on top of the agarose, which was allowed to harden for 2–5 min at room temperature. Once the agarose had hardened, the spacer slide was removed and 10 µL of labelled cell suspension was added onto the agarose pad and allowed to dry fully before a 13 mm borosilicate coverslip was placed on top for imaging. Imaging was performed using the inverted microscopy set-up described above. Images were acquired with a NIKON 100x oil immersion objective (Apo TIRF 100x/1.49) and type F2 immersion oil (Nikon) in addition to the 2.8x magnification lens in the SoRA unit giving a total magnification of 280x. Z-stack images were acquired with a 0.22 µm step size (10 slices, covering ~2 µm) using an exposure time of 50 ms with laser power set to 10% of maximum.

2.5. Fixation

For **Stepwise** ethanol fixation, 3 mL DSM 639 culture in Brock medium was added to 1.5 mL 4°C ethanol, incubated at 4°C for 10 min before adding a further 1.5 mL 4°C ethanol, incubated for 10 min, after which a final 4 mL of 4°C ethanol was added to a final concentration of 70%. For fixation in other buffer conditions, 3 mL of culture was spun for 3 min at 8000RPM in a table top centrifuge and resuspended in 3 mL of either Tris Buffer (25 mM Tris pH 7.4, 150 mM NaCl) or Citrate Buffer (25 mM Sodium Citrate, pH 3). For **Instant** ethanol fixation, 1 mL DSM 639 culture in Brock medium was added directly to 9 mL 77% 4°C ethanol. For **Formaldehyde** fixation 3 mL DSM 639 culture was spun for 3 min at 8000RPM in a table top centrifuge and

TABLE 2 List of cell contour and content markers used in this study.

Probe	Binding interaction	Readout	$\lambda_{ex}/\lambda_{em}$ (nm)	24°C (pH 3)		24°C (pH 5)		75°C (pH 7)		EtOH Fixation Compatible
				Staining	Staining	Staining	Conc.	Staining	Conc.	
CellTracker Green	-	Cytosol	492/517	Yes	4.65 µg/mL	-		Yes	0.93 µg/mL	-
CellBrite	Amines	Membrane Proteins	480/513	No		Yes	1x			No
ConA	Glycosylation	Contour (Glycosylated membrane proteins)	Multiple	No		No		No		Yes 50 µg/mL
Brilliant Blue	Amines	Contour (Membrane Proteins)	490/515	No		No		No		-

Names colour coded according to success of staining. Green, optimal staining; Blue, conditionally applicable staining; Grey, no staining. Hyphens indicate that the condition was not tested.

resuspended in 1 mL 4% freshly prepared paraformaldehyde (PFA) in H₂O and incubated at room temperature with shaking for 10 min. Cells were then washed with 1 mL phosphate-buffered saline supplemented with 0.1% Tween 20 (PBST), resuspended in phospho-buffered saline supplemented with 0.1% Triton X-100 and incubated at room temperature with shaking for 10 min in order to permeabilize. Cells were then washed with 1 mL PBST before proceeding to immunolabelling. All samples were stored at 4°C before labelling and imaging.

2.6. Imaging fixed cells using immunofluorescence

Immunolabelling was performed as described by [Hurtig et al. \(2023\)](#). Briefly, 1 mL fixed cells was spun in a tabletop centrifuge (3 min, 8000RPM) after which the supernatant was discarded and cells were washed twice in 1 mL PBST supplemented with 3% bovine serum albumin (PBSTA) to remove all remaining fixative. Cells were resuspended in a final volume of 100 µL PBSTA supplemented with 5% fetal bovine serum (FBS) and primary antibodies (in this case a lab generated anti-CdvB: [Tarrason Risa et al., 2020](#); [Hurtig et al., 2023](#)). Cells were incubated overnight at room temperature with 500 rpm agitation before washing with 1 mL PBSTA and resuspending in a final volume of 100 µL supplemented with secondary antibodies (either AlexaFluor-488 anti-rabbit (Thermo Fisher Scientific, A11034) or AlexaFluor 546 anti-rabbit (Thermo Fisher Scientific, A11035), 1:10,000) and 50 µg/mL Concanavalin A conjugated to Alexa Fluor 647 (Thermo Fisher Scientific, C21421). Cells were incubated for 3 h at room temperature and 500 rpm agitation after which cells were washed with 1 mL PBSTA and resuspended in a final volume of 1 mL supplemented with 3 µM DAPI (4',6-diamidino-2-phenylindole; Thermo Fisher Scientific, 62,248). For imaging, Lab-Tek chambered slides (Thermo Fisher Scientific, 177437PK) were coated with 2% polyethyleneimine (PEI) at 37°C for a minimum of 30 min. Coated chambers were washed with Milli-Q water before 200 µL cell suspension was added per well and spun down for 1 h at 750 relative centrifugal force (RCF). Imaging was performed, as for live-cell imaging, at room temperature using an exposure time of 200 ms for detection of secondary antibodies and an exposure time

of 500 ms for detection of DNA. Analysis and z-axis maximum projections were performed using ImageJ.

3. Results

In this paper we report our efforts to identify dyes and conditions that can be used to image thermoacidophilic archaea live. For this analysis, a variety of probes for proteins, lipids and nucleic acids were evaluated for their capacity to stain live *S. acidocaldarius* cells at 75°C or at room temperature. Unless otherwise indicated, cells were labelled with fluorescent probes in Brock medium at pH 2.9. To ensure that cells do not move during the imaging process, cells were immobilized under a soft gel pad. For imaging at 75°C, gelrite pads (Sigma Aldrich, G1910) were placed on top of a labelled cell suspension, and cells at the edge of the pad were imaged for up to 2 h as described by [Hurtig et al. \(2023\)](#). For room temperature imaging, labelled cells were placed between a low melting temp agarose pad and a glass coverslip as described above, and imaged for 5–10 min. In parallel, we optimized the visualization of DNA and protein structures in fixed cells using immunofluorescence. All membrane markers tested are summarized in [Table 1](#), S-layer and cytoplasmic markers in [Table 2](#), and DNA markers in [Table 3](#).

3.1. Imaging at 75°C

The microscopy set-up used for live-cell imaging includes a heated cap and stage that functions to maintain a temperature of 75°C for several hours without dehydration ([Pulschen et al., 2020](#)). In addition, we recently added a Yokogawa SoRa unit to our confocal microscope ([Azuma and Kei, 2015](#); [Hurtig et al., 2023](#)), enabling the resolution of discrete subcellular structures in ~1 µm diameter *S. acidocaldarius* cells. In previous live-imaging work, cells were stained with NileRed to mark the membrane and SybrSafe to label the DNA in order to visualize and characterize cell division in *S. acidocaldarius* ([Pulschen et al., 2020](#)). This is improved by using CellMask Deep Red Plasma Membrane Stain (CellMask) as a membrane stain ([Figure 1B](#)). CellMask provides a brighter and more specific membrane signal compared to NileRed, at both 75°C and room temperature

TABLE 3 List of DNA labels used in this study.

Probe	$\lambda_{ex}/\lambda_{em}$ (nm)	24°C (pH 3)		24°C (pH 7)		75°C (pH 3)		EtOH Fixation Compatible	
		Staining	Conc.	Staining	Conc.	Staining	Conc.		
SYBR TM Safe DNA Gel Stain	502/530	Yes	1:10000	Yes	1:10000	Yes	1:5000	No	
SYTO TM 11	508/527	Yes	5 μ M (~2 μ g/mL)	-	-	-	-	-	
SYTO TM 12	499/522	No		-	-	-	-	-	
SYTO TM 13	488/509	Weak signal		-	-	-	-	-	
SYTO TM 14	517/549	No		-	-	-	-	-	
SYTO TM 16	488/518	Yes	1 μ M (~0.45 μ g/mL)	-	-	-	-	-	
SYTO TM 17	621/634	No		-	-	-	-	-	
SYTO TM 21	494/517	No		-	-	-	-	-	
SYTO TM 24	490/515	No		-	-	-	-	-	
SYTO TM 59	622/645	Yes	5 μ M (~2.75 μ g/mL)	-	-	-	-	Yes	5 μ M
SYTO TM 60	652/678	No		-	-	-	-	-	
SYTO TM 61	628/645	Yes	5 μ M (~2.5 μ g/mL)	-	-	-	-	-	
SYTO TM 62	652/676	Yes	5 μ M (~2.75 μ g/mL)	-	-	-	-	-	
SYTO TM 63	657/673	Yes	5 μ M (~2.75 μ g/mL)	-	-	-	-	-	
SYTO TM 64	599/619	Yes	5 μ M (~2 μ g/mL)	-	-	-	-	-	
GelRed [®]	279/593	No		-	-	-	-	-	
DAPI	350/470	No		No	-	No	-	Yes	10 μ g/mL
Hoechst	350/461	No		No	-	No	-	Yes	1.23 μ g/mL

Names colour coded according to success of staining. Green, optimal staining; Blue, conditionally applicable staining; Grey, no staining. Hyphens indicate that the condition was not tested.

(Figure 2A). Unlike NileRed, which inserts into the lipid core, CellMask is composed of a hydrophilic fluorophore attached to a lipophilic tail which inserts into the membrane. The improved resolution of the membrane signal relative to the cytoplasmic signal may therefore reflect the inability of CellMask to cross the archaeal membrane.

Importantly, under these conditions, control cells (MW001) labelled with CellMask were observed undergoing constriction over a period of ~12 min at a rate of ~0.1 μ m/s (Hurtig et al., 2023), in line with speeds previously reported by Pulschen et al. (2020). This indicates that though CellMask is brighter and more specific than NileRed, it does not cause additional imaging stress. We did not observe significant phototoxicity over longer imaging periods of 2–3 h. Additional dyes that are compatible with imaging at 75°C, include the CellMask Orange Plasma Membrane Stain and Mitotracker Red CMXRos, although the latter also labels the cytoplasm (Table 1). Unfortunately, we were unable to identify a good label that could be used to image the proteinaceous surface layer at 75°C (see Table 2). A recent study however was able to visualize the cell contour of *S. acidocaldarius* by using a N-Hydroxysuccinimide (NHS)-ester functionalized Alexa Fluor dye to non-specifically label surface proteins in phospho-buffered saline at room temperature then returning cells into Brock medium for live-cell imaging at 75°C (Charles-Orszag et al., 2023).

While a range of dyes were found to strongly label the DNA of *S. acidocaldarius* cells at 75°C, the only DNA dye tested that proved compatible with long-term live cell imaging was SybrSafe (Table 3). Other dyes exhibited cytotoxicity and rapid bleaching. Nevertheless,

the hardware improvements we have put in place allow for precise visualization of DNA morphology and dynamics using SybrSafe. Under these conditions, the *S. acidocaldarius* genome appears to be crescent shaped, lying close to the membrane on one side of the cell during interphase. Prior to division, it then compacts to form two foci that align with the plane of cell cleavage (Figure 1B). The organization of the genome and the membrane dynamics visualized by live cell imaging at 75°C using SybrSafe and CellMask was used to validate the other stains and probes described below.

3.2. Imaging at room temperature

While only a limited number of markers were found to be compatible with live imaging at 75°C, many more were found to be compatible with imaging at room temperature. In addition to the dyes described above, Mitotracker dyes, which have been used to visualize the membranes of euryarchaeota and DPANN (Maslov et al., 2018; Hamm et al., 2023), were compatible with staining the membrane of *S. acidocaldarius* cells in Brock culture medium at room temperature (Figure 2A and Table 1). Interestingly, eukaryotic type lipid stains such as BODIPY-Ceramide were also found to weakly stain the membrane of *S. acidocaldarius* cells, providing hope that other tools developed to visualize lipids in eukaryotes can be adapted to the study of the archaeal membrane. Long-chain carbocyanine dyes (e.g., DiI C18 etc), commonly used to visualize eukaryotic and bacterial membranes, did not stain the archaeal bounding membrane in culture medium or in Tris Buffer (pH 7.4). The short-chain

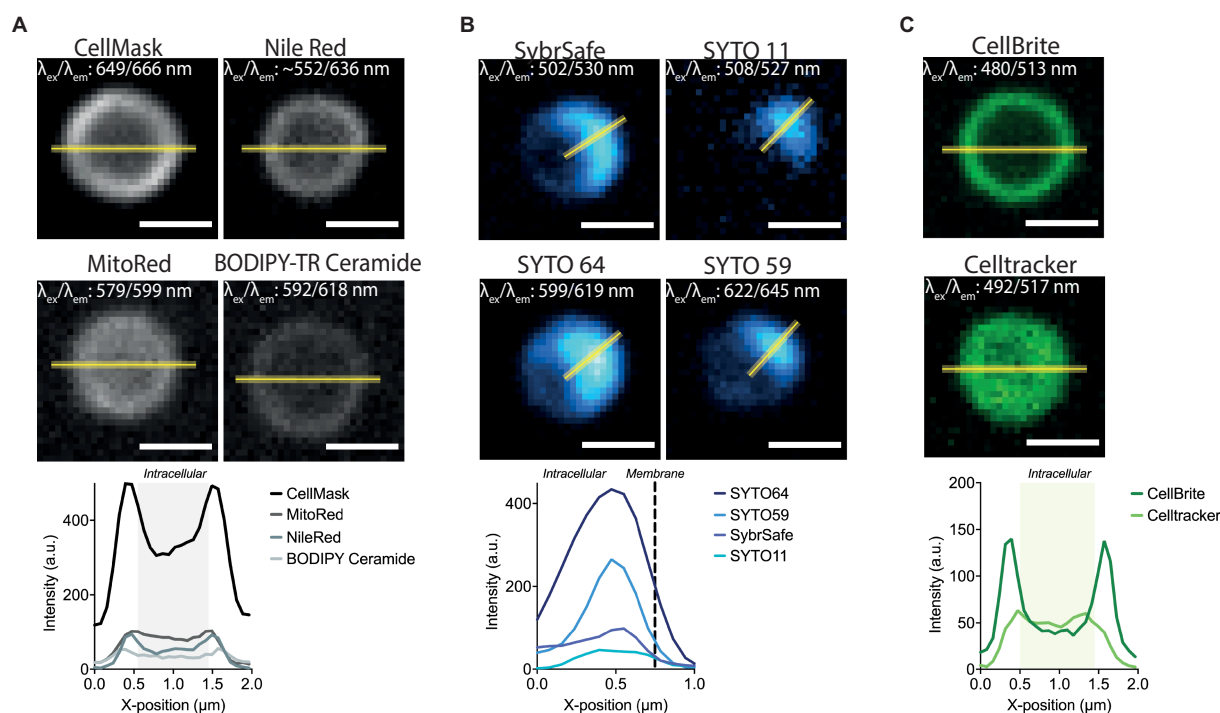


FIGURE 2

Labelling *S. acidocaldarius* at lower temperatures. **(A)** Representative membrane stains in Brock medium at 25°C (top) and intensity profiles through the equator of the cell (bottom). **(B)** Successful DNA stains in Brock medium at 25°C (top) and respective intensity profiles through the middle of the DNA signal (bottom). **(C)** Staining of membrane proteins (CellBrite, left top), general cytosolic staining (CellTracker, right top) in Brock medium at 25°C and intensity profiles through the equator of the cell (bottom). Line profiles are represented as a yellow line. Excitation and emission maxima for each dye are given in the top corner of each image. Scale bar = 1 μm .

carbocyanine DiO C6 was able to stain cells in Brock medium at room temperature, but proved to be cytotoxic at 75°C.

At room temperature, DNA could be stained with a number of SYTO nucleic acid stains with different spectral properties (Figure 2B). While the signal at the start of imaging was good, these dyes suffered significant photobleaching when compared with SybrSafe. Nevertheless, these dyes could be used for labelling DNA in fixed cells, where photobleaching is less of a problem.

In addition to testing DNA and membrane dyes, we also tested a host of markers that we hoped would label membrane proteins and the S-layer in live cells. Unfortunately, these did not stain cells in low pH Brock medium, Citrate Buffer, or Tris Buffer. An exception was CellBrite, which targets primary amines. Unfortunately, CellBrite only marked the cell periphery at a high pH (Figure 2C), conditions that compromise DNA and membrane organization and, likely, *S. acidocaldarius* viability. Interestingly, the cell content marker CellTracker CMFDA (5-chloromethylfluorescein diacetate) efficiently crossed the plasma membrane to provide a uniform labeling of the *S. acidocaldarius* cytoplasm in low pH Brock medium at both room temperature and 75°C.

3.3. Fixation and immunofluorescence

In the absence of genetically encoded fluorescent proteins for use in hyperthermophiles, visualizing proteins and protein structures in these organisms currently relies on immunofluorescence. Unfortunately, as detergent and solvent permeabilization both greatly

impact membrane integrity this precluded visualization of the membrane in fixed cells.

We have tested and optimized a range of fixation techniques for *S. acidocaldarius*. Our standard procedure for staining *S. acidocaldarius* cells employs a “Stepwise fixation” protocol in ethanol (Figure 3A, left). This involves adding 3 mL of culture to 1.5 mL of ethanol. After 10 min on ice a further 1.5 mL ethanol is added and after an additional 10 min on ice, the final volume is brought to 10 mL with ethanol, to a final concentration of 70% ethanol (Bernander and Poplawski, 1997; Han et al., 2017; Zhang et al., 2019). Importantly, this fixation protocol yields a DNA signal similar to that observed by live cell imaging. Cells fixed in this way, can also be effectively labelled with antibodies targeting ESCRT-III homologues to reveal division rings like those reported in previous studies (Samson et al., 2008, 2011; Tarrason Risa et al., 2020; Hurtig et al., 2023). In addition, we are able to use fluorescently conjugated Concanavalin A (ConA) to label glycosylated proteins and the cell contour in ethanol fixed cells. While this stain has been used as a proxy for the S-layer, which we know to be heavily glycosylated in *S. acidocaldarius* (Peyfoon et al., 2010), ConA likely also marks a range of other glycosylated membrane proteins. Note that ConA also causes cells to aggregate in a concentration dependent manner.

We attempted to streamline the two-step ethanol fixation protocol by adding 1 mL of culture directly to 9 mL of 77% ethanol, yielding a final concentration of 70% ethanol. Unfortunately, this “Instant fixation protocol” failed to preserve the structure of neither DNA nor division rings (Figure 3A, middle). While attempts to fix cells with formaldehyde (either paraformaldehyde, glutaraldehyde or a

combination of the two) were able to fix DNA so that its organization resembled that seen during live cell imaging, it did not preserve division rings (Figure 3A, right). Formaldehyde fixation was able to preserve ConA labelling of the membrane/S-layer, however, the staining appeared weaker than when using the Stepwise ethanol fixation. Combining formaldehyde fixation with ethanol fixation by replacing the detergent permeabilization portion of the formaldehyde fixation protocol with the stepwise ethanol fixation protocol did not appear to preserve CdvB division ring structures (data not shown).

As the Stepwise ethanol fixation was the only method tested that faithfully preserved DNA organization and division rings, we used this as a starting point to test the impact of different buffer conditions on immunolabelling (Figure 3B). For this analysis, cells were spun down and resuspended in either a minimal Brock medium (mBrock: pH 5, no supplementation with NZ-amine or FeSO₄), Tris or Citrate buffer, before being fixed using the Stepwise ethanol protocol described above. Interestingly, in both mBrock and Tris Buffer, CdvB protein structures appeared smoother and more continuous than when fixation was performed in Brock culture medium. However, DNA organization was entirely disrupted in Tris Buffer, in both fixed and live cell imaging at room temperature (Figure 3C). By contrast, fixation in Citrate buffer preserved both DNA and protein organization, but the signal was reduced in both cases compared to that observed following fixation in Brock medium. While the ESCRT-III signal appeared slightly more uniform in Citrate buffer, until the advent of thermostable GFP variants that work in *Sulfolobus*, it will not be possible to determine whether or not this represents the ring in its physiological state.

4. Discussion and conclusion

Here we present a set of tools and protocols that can be used to image live and fixed *S. acidocaldarius* cells. By cataloguing both the successes and failures of different visualization methods and dyes, we hope this paper will provide the archaeal biology field with a starting point from which to optimize the imaging of a range of archaea. Similar studies have been conducted for bacterial species which lack an established genetic toolbox (Atwal et al., 2016). Importantly, this analysis reveals that archaeal membranes can be labelled with a variety of lipid probe architectures – including probes that insert into the core of the lipid membrane, probes that intercalate between lipid tails, probes with a lipophilic anchor, and probes that are sensitive to membrane properties. This suggests that it may be possible to adapt other techniques used to visualize and characterize the membrane in eukaryotes for use in archaea, e.g., to localize specific lipid domains or species (Höglinger et al., 2017) or to measure membrane properties and organization (Klymchenko and Kreder, 2014; Colom et al., 2018). As the archaeal membrane is chemically, structurally, and functionally distinct from bacterial or eukaryotic membranes, this type of biophysical characterization will be an interesting direction for future work.

A number of the membrane dyes tested here are compatible with aldehyde fixation and can be used to complement immunofluorescence investigations in species that are amenable to aldehyde fixation. Unfortunately, however, *S. acidocaldarius* cells did not respond well to aldehyde fixation, even when used in conjunction with EtOH fixations. While fixation via the stepwise addition of ice cold EtOH preserves

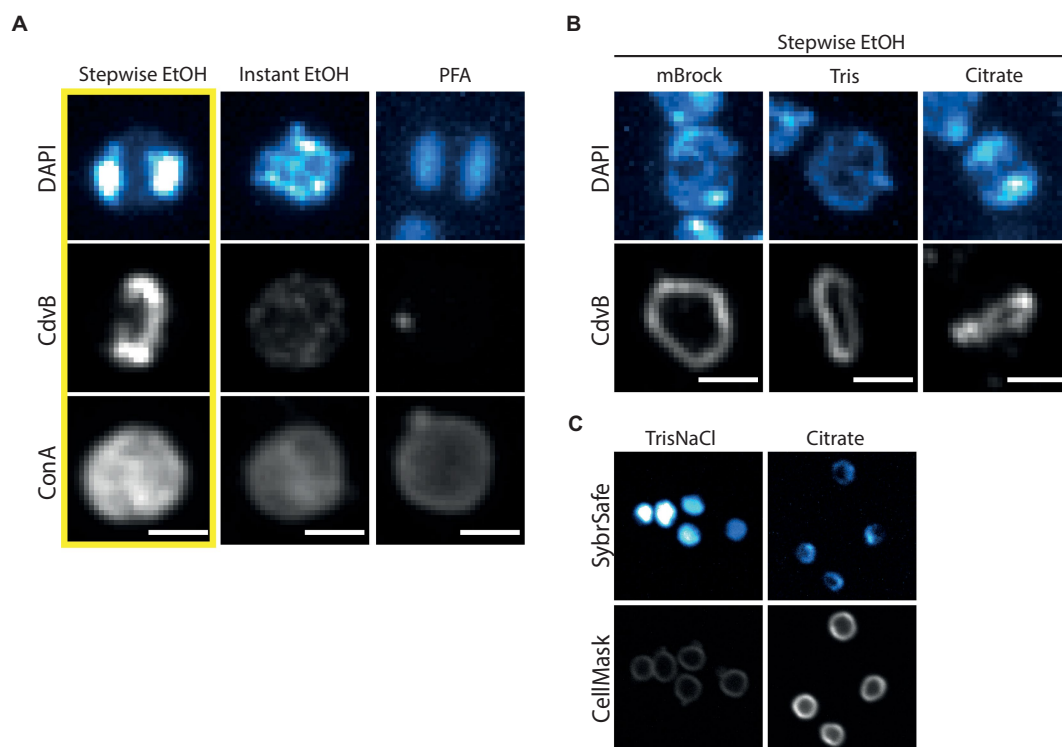


FIGURE 3

Fixation and immunofluorescence labelling of DNA, proteins and cell surface. (A) Comparison of different fixation methods in preserving DNA, protein and cell surface structures as represented by DAPI, immunofluorescence labelling of CdvB and ConA, respectively. Cells were fixed with Stepwise EtOH (left), Instant EtOH (middle) and Paraformaldehyde (right). (B) Stepwise Fixation performed in buffers of reduced complexity. (C) Live cell imaging at room temperature with SybrSafe and Cellmask in Tris (left) and Citrate Buffer (right). Scale bar = 1 μ m.

cell shape and DNA morphologies, this likely compromises the membrane. Thus, it will be important to try other methods, e.g., cryofixation, to visualize the membrane in fixed cells. We also noted that the medium in which the cells are fixed has a significant impact on the quality of fixation, especially in regards to DNA morphology.

While we have optimized our protocols for imaging *S. acidocaldarius* cells, it is hoped that these protocols can be adapted to label other archaeal species, as well as other thermoacidophilic organisms. By sharing this information, we hope to assist in the further development of a transparent and collaborative archaeal research community.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

AC and BB conceived the study with input from BH. Live cell imaging (RT and 75°C) was performed by AC, immunofluorescence and fixed cell imaging was performed by BH. AC prepared the figures and text with input from BH and BB. All authors contributed to the article and approved the submitted version.

Funding

AC was funded by an EMBO Postdoctoral fellowship (ALTF_1041-2021) and a Marie Skłodowska-Curie Individual Fellowship (101068523) provided by UKRI. BH was supported by

Wellcome Trust (203276/A/16/Z). BB received support from the MRC LMB, the Wellcome Trust (203276/Z/16/Z) and (222460/Z/21/Z), the VW Foundation (94933), the Life Sciences–Moore-Simons Foundation (735929LPI), and from the Gordon and Betty Moore Foundation's Symbiosis in Aquatic Systems Initiative (9346).

Acknowledgments

The authors would like to acknowledge all members of the Baum lab for their input throughout the project. Specifically, we would like to thank Andre Pulschen, Gabriel Tarrason Risa and Fredrik Hurtig for initially establishing the methods used in this paper; Jovan Traparic, Matthew Kenneth and Yin-wei Kuo for their input in method optimization and feedback on the manuscript. We would also like to thank the MRC LMB Light Microscopy facility for technical support.

Conflict of interest

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.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Akil, C., and Robinson, R. C. (2018). Genomes of Asgard archaea encode profilins that regulate actin. *Nature* 562, 439–443. doi: 10.1038/s41586-018-0548-6
- Atwal, S., Giengkam, S., VanNieuwenhze, M., and Salje, J. (2016). Live imaging of the genetically intractable obligate intracellular bacteria *Orientia tsutsugamushi* using a panel of fluorescent dyes. *J. Microbiol. Methods* 130, 169–176. doi: 10.1016/j.mimet.2016.08.022
- Azuma, T., and Kei, T. (2015). Super-resolution spinning-disk confocal microscopy using optical photon reassignment. *Opt. Express* 23, 15003–15011. doi: 10.1364/OE.23.015003
- Bagatolli, L., Gratton, E., Khan, T. K., and Chong, P. L.-G. (2000). Two-photon fluorescence microscopy studies of bipolar tetraether giant liposomes from thermoacidophilic archaeobacteria *Sulfolobus acidocaldarius*. *Biophys. J.* 79, 416–425. doi: 10.1016/S0006-3495(00)76303-X
- Barry, E. R., and Bell, S. D. (2006). DNA replication in the archaea. *Microbiol. Mol. Biol. Rev.* 70, 876–887. doi: 10.1128/MMBR.00029-06
- Bernander, R. (2007). The cell cycle of *Sulfolobus*. *Mol. Microbiol.* 66, 557–562. doi: 10.1111/j.1365-2958.2007.05917.x
- Bernander, R., and Poplawski, A. (1997). Cell cycle characteristics of thermophilic archaea. *J. Bacteriol.* 179, 4963–4969. doi: 10.1128/jb.179.16.4963-4969.1997
- Bisson-Filho, A. W., Zheng, J., and Garner, E. (2018). Archaeal imaging: leading the hunt for new discoveries. *Mol. Biol. Cell* 29, 1675–1681. doi: 10.1091/mbc.E17-10-0603
- Charles-Orszag, A., Lord, S. J., and Mullins, R. D. (2021). High-temperature live-cell imaging of cytokinesis, cell motility, and cell-cell interactions in the thermoacidophilic crenarchaeon *Sulfolobus acidocaldarius*. *Front. Microbiol.* 12:707124. doi: 10.3389/fmicb.2021.707124
- Charles-Orszag, A., van Wolferen, M., Lord, S. J., Albers, S. V., and Mullins, R. D. (2023). *Sulfolobus acidocaldarius* adhesion pili power twitching motility in the absence of a dedicated retraction ATPase. *bioRxiv*. doi: 10.1101/2023.08.04.552066
- Colom, A., Derivery, E., Soleimanpour, S., Tomba, C., Molin, M. D., Sakai, N., et al. (2018). A fluorescent membrane tension probe. *Nat. Chem.* 10, 1118–1125. doi: 10.1038/s41557-018-0127-3
- Eme, L., Tamarit, D., Caceres, E. F., Stairs, C. W., De Anda, V., Schön, M. E., et al. (2023). Inference and reconstruction of the heimdallarchaeal ancestry of eukaryotes. *Nature* 618, 992–999. doi: 10.1038/s41586-023-06186-2
- Hamm, J. N., Liao, Y., Kügelgen, A. von, Dombrowski, N., Landers, E., Brownlee, C., et al. (2023). The intracellular lifestyle of an archaeal symbiont. *bioRxiv* 02.24.529834. doi: 10.1101/2023.02.24.529834
- Han, W., Xu, Y., Feng, X., Liang, Y. X., Huang, L., Shen, Y., et al. (2017). NQO-induced DNA-less cell formation is associated with chromatin protein degradation and dependent on AOA1-ATPase in *Sulfolobus*. *Front. Microbiol.* 8:1480. doi: 10.3389/fmicb.2017.01480
- Hatano, T., Palani, S., Papatziomou, D., Salzer, R., Souza, D. P., Tamarit, D., et al. (2022). Asgard archaea shed light on the evolutionary origins of the eukaryotic ubiquitin-ESCRT machinery. *Nat. Commun.* 13:3398. doi: 10.1038/s41467-022-30656-2
- Hennell James, R., Caceres, E. F., Escasinas, A., Alhasan, H., Howard, J. A., Deery, M. J., et al. (2017). Functional reconstruction of a eukaryotic-like E1/E2/(RING) E3 ubiquitylation cascade from an uncultured archaeon. *Nat. Commun.* 8:1120. doi: 10.1038/s41467-017-01162-7
- Höglinger, D., Nadler, A., Haberkant, P., Kirkpatrick, J., Schifferer, M., Stein, F., et al. (2017). Trifunctional lipid probes for comprehensive studies of single lipid species in living cells. *Proc. Natl. Acad. Sci.* 114, 1566–1571. doi: 10.1073/pnas.1611096114

- Hurtig, F., Burgers, T. C. Q., Cezanne, A., Jiang, X., Mol, F. N., Traparić, J., et al. (2023). The patterned assembly and stepwise Vps4-mediated disassembly of composite ESCRT-III polymers drives archaeal cell division. *Sci. Adv.* 9:eade5224. doi: 10.1126/sciadv.ade5224
- Imachi, H., Nobu, M. K., Nakahara, N., Morono, Y., Ogawara, M., Takaki, Y., et al. (2020). Isolation of an archaeon at the prokaryote-eukaryote interface. *Nature* 577, 519–525. doi: 10.1038/s41586-019-1916-6
- Jacquemet, A., Barbeau, J., Lemiègre, L., and Benvegna, T. (2009). Archaeal tetraether bipolar lipids: structures, functions and applications. *Biochimie* 91, 711–717. doi: 10.1016/j.biochi.2009.01.006
- Jarrell, K. F., Ding, Y., Meyer, B. H., Albers, S.-V., Kaminski, L., and Eichler, J. (2014). N-linked glycosylation in archaea: a structural, functional, and genetic analysis. *Microbiol. Mol. Biol. Rev.* 78, 304–341. doi: 10.1128/MMBR.00052-13
- Klymchenko, A. S., and Kreder, R. (2014). Fluorescent probes for lipid rafts: from model membranes to living cells. *Chem. Biol.* 21, 97–113. doi: 10.1016/j.chembiol.2013.11.009
- Koga, Y. (2011). Early evolution of membrane lipids: how did the lipid divide occur? *J. Mol. Evol.* 72, 274–282. doi: 10.1007/s00239-011-9428-5
- Koga, Y., and Morii, H. (2007). Biosynthesis of ether-type polar lipids in archaea and evolutionary considerations. *Microbiol. Mol. Biol. Rev.* 71, 97–120. doi: 10.1128/MMBR.00033-06
- Lake, J. A., Henderson, E., Oakes, M., and Clark, M. W. (1984). Eocytes: a new ribosome structure indicates a kingdom with a close relationship to eukaryotes. *Proc. Natl. Acad. Sci.* 81, 3786–3790. doi: 10.1073/pnas.81.12.3786
- Lewis, A. M., Recalde, A., Bräsen, C., Counts, J. A., Nussbaum, P., Bost, J., et al. (2021). The biology of thermoacidophilic archaea from the order Sulfolobales. *FEMS Microbiol. Rev.* 45:fuaa063. doi: 10.1093/femsre/fuaa063
- Lindås, A.-C., Karlsson, E. A., Lindgren, M. T., Ettema, T. J. G., and Bernander, R. (2008). A unique cell division machinery in the archaea. *Proc. Natl. Acad. Sci. U. S. A.* 105, 18942–18946. doi: 10.1073/pnas.0809467105
- Lombard, J., López-García, P., and Moreira, D. (2012). The early evolution of lipid membranes and the three domains of life. *Nat. Rev. Microbiol.* 10, 507–515. doi: 10.1038/nrmicro2815
- Maslov, I., Bogorodskiy, A., Mishin, A., Okhrimenko, I., Gushchin, I., Kalenkov, S., et al. (2018). Efficient non-cytotoxic fluorescent staining of halophiles. *Sci. Rep.* 8:2549. doi: 10.1038/s41598-018-20839-7
- Mattioli, F., Bhattacharyya, S., Dyer, P. N., White, A. E., Sandman, K., Burkhart, B. W., et al. (2017). Structure of histone-based chromatin in archaea. *Science* 357, 609–612. doi: 10.1126/science.aaj1849
- Merino, N., Aronson, H. S., Bojanova, D. P., Feyhl-Buska, J., Wong, M. L., Zhang, S., et al. (2019). Living at the extremes: extremophiles and the limits of life in a planetary context. *Front. Microbiol.* 10:780. doi: 10.3389/fmicb.2019.00780
- Nunoura, T., Takaki, Y., Kakuta, J., Nishi, S., Sugahara, J., Kazama, H., et al. (2011). Insights into the evolution of archaea and eukaryotic protein modifier systems revealed by the genome of a novel archaeal group. *Nucleic Acids Res.* 39, 3204–3223. doi: 10.1093/nar/gkq1228
- Omer, A. D., Lowe, T. M., Russell, A. G., Ebhardt, H., Eddy, S. R., and Dennis, P. P. (2000). Homologs of small nucleolar RNAs in archaea. *Science* 288, 517–522. doi: 10.1126/science.288.5465.517
- Patel, G. B., and Sprott, G. D. (1999). Archaeobacterial ether lipid liposomes (archaeosomes) as novel vaccine and drug delivery systems. *Crit. Rev. Biotechnol.* 19, 317–357. doi: 10.1080/0738-859991229170
- Peeters, E., Driessen, R. P. C., Werner, F., and Dame, R. T. (2015). The interplay between nucleoid organization and transcription in archaeal genomes. *Nat. Rev. Microbiol.* 13, 333–341. doi: 10.1038/nrmicro3467
- Peretó, J., López-García, P., and Moreira, D. (2004). Ancestral lipid biosynthesis and early membrane evolution. *Trends Biochem. Sci.* 29, 469–477. doi: 10.1016/j.tibs.2004.07.002
- Peyfoon, E., Meyer, B., Hitchen, P. G., Panico, M., Morris, H. R., Haslam, S. M., et al. (2010). The S-layer glycoprotein of the crenarchaeote *Sulfolobus acidocaldarius* is glycosylated at multiple sites with chitobiose-linked N-glycans. *Archaea* 2010:754101. doi: 10.1155/2010/754101
- Pulschen, A. A., Mutavchiev, D. R., Culley, S., Sebastian, K. N., Roubinet, J., Roubinet, M., et al. (2020). Live imaging of a hyperthermophilic archaeon reveals distinct roles for two ESCRT-III homologs in ensuring a robust and symmetric division. *Curr. Biol.* 30, 2852–2859.e4. doi: 10.1016/j.cub.2020.05.021
- Rodrigues-Oliveira, T., Wollweber, F., Ponce-Toledo, R. I., Xu, J., Rittmann, S. K.-M. R., Klingl, A., et al. (2023). Actin cytoskeleton and complex cell architecture in an Asgard archaeon. *Nature* 613, 332–339. doi: 10.1038/s41586-022-05550-y
- Samson, R. Y., Obita, T., Freund, S. M., Williams, R. L., and Bell, S. D. (2008). A role for the ESCRT system in cell division in archaea. *Science* 322, 1710–1713. doi: 10.1126/science.1165322
- Samson, R. Y., Obita, T., Hodgson, B., Shaw, M. K., Chong, P. L.-G., Williams, R. L., et al. (2011). Molecular and structural basis of ESCRT-III recruitment to membranes during archaeal cell division. *Mol. Cell* 41, 186–196. doi: 10.1016/j.molcel.2010.12.018
- Shimomura, O., Johnson, F. H., and Saiga, Y. (1962). Extraction, purification and properties of aequorin, a bioluminescent protein from the luminous hydromedusa, *Aequorea*. *J. Cell. Physiol.* 59, 223–239. doi: 10.1002/jcp.1030590302
- Spang, A., Saw, J. H., Jørgensen, S. L., Zaremba-Niedzwiedzka, K., Martijn, J., Lind, A. E., et al. (2015). Complex archaea that bridge the gap between prokaryotes and eukaryotes. *Nature* 521, 173–179. doi: 10.1038/nature14447
- Tarrason Risa, G., Hurtig, F., Bray, S., Hafner, A. E., Harker-Kirschneck, L., Faull, P., et al. (2020). The proteasome controls ESCRT-III-mediated cell division in an archaeon. *Science* 369:eaa2532. doi: 10.1126/science.aaz2532
- Thevenaz, P., Rüttimann, U. E., and Unser, M. (1998). A pyramid approach to subpixel registration based on intensity. *IEEE Trans. Image Process.* 7, 27–41. doi: 10.1109/83.650848
- van Wolferen, M., Pulschen, A. A., Baum, B., Gribaldo, S., and Albers, S.-V. (2022). The cell biology of archaea. *Nat. Microbiol.* 7, 1744–1755. doi: 10.1038/s41564-022-01215-8
- Villanueva, L., von Meijenfildt, F. A. B., Westbye, A. B., Yadav, S., Hopmans, E. C., Dutilh, B. E., et al. (2021). Bridging the membrane lipid divide: bacteria of the FCB group superphylum have the potential to synthesize archaeal ether lipids. *ISME J.* 15, 168–182. doi: 10.1038/s41396-020-00772-2
- Werner, F. (2007). Structure and function of archaeal RNA polymerases. *Mol. Microbiol.* 65, 1395–1404. doi: 10.1111/j.1365-2958.2007.05876.x
- Woese, C. R., and Fox, G. E. (1977). Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc. Natl. Acad. Sci.* 74, 5088–5090. doi: 10.1073/pnas.74.11.5088
- Zaremba-Niedzwiedzka, K., Caceres, E. F., Saw, J. H., Bäckström, D., Juzokaite, L., Vancaester, E., et al. (2017). Asgard archaea illuminate the origin of eukaryotic cellular complexity. *Nature* 541, 353–358. doi: 10.1038/nature21031
- Zhang, C., Wipfler, R. L., Li, Y., Wang, Z., Hallett, E. N., and Whitaker, R. J. (2019). Cell structure changes in the hyperthermophilic crenarchaeon *Sulfolobus islandicus* lacking the S-layer. *MBio* 10, e01589–e01519. doi: 10.1128/mBio.01589-19
- Zimmer, M. (2002). Green fluorescent protein (GFP): applications, structure, and related photophysical behavior. *Chem. Rev.* 102, 759–782. doi: 10.1021/cr010142r
- Zwickl, P., Lottspeich, F., and Baumeister, W. (1992). Expression of functional *Thermoplasma acidophilum* proteasomes in *Escherichia coli*. *FEBS Lett.* 312, 157–160. doi: 10.1016/0014-5793(92)80925-7



OPEN ACCESS

EDITED BY

Jacques Oberto,
UMR9198 Institut de Biologie Intégrative de la
Cellule (I2BC), France

REVIEWED BY

Hannu Myllykallio,
INSERM U1182 Laboratoire d'Optique et
Biosciences (LOB), France
Haruyuki Atomi,
Kyoto University, Japan

*CORRESPONDENCE

Solenne Ithurbide
✉ solenne.ithurbide@umontreal.ca

†PRESENT ADDRESS

Solenne Ithurbide,
Département de Microbiologie, Infectiologie et
Immunologie, Université de Montréal,
Montréal, QC, Canada

RECEIVED 01 August 2023

ACCEPTED 11 September 2023

PUBLISHED 29 September 2023

CITATION

Patro M, Duggin IG, Albers S-V and
Ithurbide S (2023) "Influence of plasmids,
selection markers and auxotrophic mutations
on *Haloferax volcanii* cell shape plasticity".
Front. Microbiol. 14:1270665.
doi: 10.3389/fmicb.2023.1270665

COPYRIGHT

© 2023 Patro, Duggin, Albers and Ithurbide.
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.

"Influence of plasmids, selection markers and auxotrophic mutations on *Haloferax volcanii* cell shape plasticity"

Megha Patro ^{1,2}, Iain G. Duggin ³, Sonja-Verena Albers ¹ and
Solenne Ithurbide ^{1*†}

¹Molecular Biology of Archaea, Institute of Biology II, Faculty of Biology, University of Freiburg, Freiburg, Germany, ²Spemann Graduate School of Biology and Medicine, University of Freiburg, Freiburg, Germany, ³The Australian Institute for Microbiology and Infection, University of Technology Sydney, Sydney, NSW, Australia

Haloferax volcanii and other Haloarchaea can be pleomorphic, adopting different shapes, which vary with growth stages. Several studies have shown that *H. volcanii* cell shape is sensitive to various external factors including growth media and physical environment. In addition, several studies have noticed that the presence of a recombinant plasmid in the cells is also a factor impacting *H. volcanii* cell shape, notably by favoring the development of rods in early stages of growth. Here we investigated the reasons for this phenomenon by first studying the impact of auxotrophic mutations on cell shape in strains that are commonly used as genetic backgrounds for selection during strain engineering (namely: H26, H53, H77, H98, and H729) and secondly, by studying the effect of the presence of different plasmids containing selection markers on the cell shape of these strains. Our study showed that most of these auxotrophic strains have variation in cell shape parameters including length, aspect ratio, area and circularity and that the plasmid presence is impacting these parameters too. Our results indicated that $\Delta hdrB$ strains and *hdrB* selection markers have the most influence on *H. volcanii* cell shape, in addition to the sole presence of a plasmid. Finally, we discuss limitations in studying cell shape in *H. volcanii* and make recommendations based on our results for improving reproducibility of such studies.

KEYWORDS

Haloferax volcanii, haloarchaea, cell shape, HDRB, plasmid, auxotrophy

1. Introduction

In the last couple of years, the field of cell biology of Archaea has developed immensely (van Wolferen et al., 2022). Amongst the current model organisms, *Haloferax volcanii* for which many tools have been developed has facilitated most of the cell biology discoveries in the Euryarchaeota phylum and helped build up the foundations of new paradigms in archaeal cell division, cell envelope biology and cell shape (Duggin et al., 2015; Pohlschroder and Schulze, 2019; Ithurbide et al., 2022). In general, how is cell shape established is an intriguing question in Haloarchaea as these species are commonly characterized as pleomorphic organisms that can harbor shapes ranging from rods and plates (also called disks or discoid) to triangles, squares, and more exotic forms (Mullakhanbhai and Larsen, 1975; Bisson-Filho et al., 2018; Walsh et al., 2019; Liao et al., 2021; Schwarzer et al., 2021; Tittes et al., 2021). In addition, several studies have shown that one

species can adopt one shape or another depending on the growth phase. Indeed, in early log phase, rods can be observed in *H. volcanii* liquid cultures. Then, it transitions into a mixed population where rod cells coexist with plate cells as they grow in exponential phase to finally tend towards a homogenous population of plate cells of a smaller size by the time cells enter stationary phase (Li et al., 2019; de Silva et al., 2021). This growth phase dependent cell shape transitions have also been shown for *Haloferax gibbonsii*, *Haloarcula hispanica* and *Haloarchula californiae* (Schwarzer et al., 2021; Tittes et al., 2021). However, differences occur between the different species regarding the window of time when shape transitions occur, as well as the type of cell shape that are observed. Interestingly, another haloarchaeal species, *Halobacterium salinarum*, was shown to be mostly rod at all growth stages, with more than 90% of rods even in stationary phase (Eun et al., 2018). Furthermore, it has been shown that other external factors can influence cell shape including the media the cells are grown in and mechanical forces, further supporting the natural plasticity of haloarchaea cell shape (Bisson-Filho et al., 2018; Eun et al., 2018; de Silva et al., 2021).

Despite the interesting pleomorphism of Haloarchaea, little is known about the mechanisms involved in cell shape development and transition regulation. The archaeal specific tubulin CetZ1 was found to be essential for rod formation and knock out mutants remained in plate morphology (Duggin et al., 2015; de Silva et al., 2021). To the contrary, a mutant containing an insertion of a transposon upstream of HVO_2176 was found to maintain the rod shape and failed to transition to the plate morphology (Schiller et al., 2023). A comparative proteomics approach of the rod-only and plate-only mutants, and WT, in early log and in late log, identified several proteins implicated in cell shape development and transition (Schiller et al., 2023). These included HVO_2175, a SMC family homologue (Structural Maintenance of Chromosomes) called Sph3 whose disruption led the cells to only form rods, and its two neighboring genes *rdFA* (HVO_2174), important for rod formation, and *ddfA* (HVO_2176), important for plate formation. This study also identified “volactin”, a so-far uncharacterized homologue of actin, which appears to represent another group of cytoskeletal proteins alongside the tubulin family proteins, CetZ and FtsZ, important for the rod-to-plate cell shape transition (Schiller et al., 2023). Further studies are needed to determine the role of these genes in cell shape determination. Cell shape is a spatial cue for many core cell biology functions, including placement of the *H. volcanii* division plane (Walsh et al., 2019), the motility and chemotaxis machineries (Li et al., 2019; Nußbaum et al., 2020) and possibly many others.

Despite these advances, a recurring observation made by several laboratories might provide a limitation to consistency in future studies. Several studies have reported that morphological changes are induced in *H. volcanii* strains due to presence or absence of a plasmid based on the endogenous plasmid pHV2 (Supplementary Table S2). The first report of such a phenomenon was by Abdul Halim et al. (2016). While they were studying the phenotype of a deletion of the archaeosortase *artA* gene on cell morphology, they observed that during the first liquid culture inoculated from colonies on agar plates, about 20% of the cells were rods both in WT (H53 Δ trp Δ pyrE2) and in Δ artA strains, whereas the same strains displayed 90% and 100% of rods, respectively, when harboring a plasmid (Abdul Halim et al., 2016). After a subsequent dilution of the liquid cultures, the Δ artA mutant only

displayed plate cells whereas the same background strain bearing the plasmid (pTA963) were all rod shaped. It was also observed that the presence of pTA963 negatively impacted the growth rate of the Δ artA mutant compared to the strain without the plasmid. Corroborating these observations, de Silva et al., 2021 have shown that the early log rod development is much more visible for H26 (Δ pyrE2) or H98 (Δ hdrB, Δ pyrE2) containing the plasmid pTA962 compared to the plasmid-free strains in a rich liquid medium (de Silva et al., 2021).

Further potential complexities arise from this plasmid related cell shape variations in strains harboring mutations that affect cell morphology. For example, *ftsZ2* knockout strains show a very severe cell division defect and grow as very large plate cells whereas the same strain transformed with a plasmid is filamentous (Liao et al., 2021). The same is true for knock-down of *sepF*, another cell division protein (Nußbaum et al., 2021), whereas Δ ftsZ1 cells do not become elongated in presence of the plasmid (Liao et al., 2021). These effects on cell shape, and their dependence on strain background and plasmid presence, potentially complicates the interpretation of the results of such genetic analyses.

To better understand how the plasmid used in these studies influences cell shape, here we characterized the cell shape of several *H. volcanii* strains that are commonly used as genetic backgrounds for selection during strain engineering namely: H26, H53, H77, H98, and H729 (Allers et al., 2004), and investigated the influence of the plasmid and various auxotrophic markers. Our results indicate that Δ hdrB marker strains and *hdrB* selection markers have the most influence on *H. volcanii* cell shape in addition to the sole presence of a plasmid. Finally, we discuss the current limitations in studying cell shape in *H. volcanii* and make recommendations based on our results for improving reproducibility of such studies.

2. Materials and methods

2.1. Strains and growth conditions

The list of strains and plasmids are available in Supplementary Table S1. *Haloferax volcanii* cells were grown in Hv-YPC or Hv-CA media.

Hv-YPC medium consisted of 0.5% (w/v) yeast extract (Oxoid 1447488-02), 0.1% (w/v) peptone (Oxoid 2145582), and 0.1% (w/v) casamino acids (Difco 0314259) dissolved in 18% buffered salt water (SW) 144 g/L NaCl (Roth 0962.3), 18 g/L MgCl₂ * 6 H₂O (Roth A537.1), 21 g/L MgSO₄*7H₂O (Roth P027.2), 4.2 g/L KCl (Roth P017.3), 12 mM Tris/HCl (Roth 9090.2), pH 7.5, supplemented with 3 mM CaCl₂ (Roth 5239.1), adjusted to a pH of 7.2 with KOH.

Hv-CA medium consisted of: 0.5% (w/v) casamino acids (Difco 0314259) dissolved in 18% SW, supplemented with 3 mM CaCl₂ (Roth 5239.1), and 0.8 µg/mL of thiamine, and 0.1 µg/mL of biotin, adjusted to a pH of 7.2 with KOH.

For strains with auxotrophic mutations, the details of the culture conditions and additives to support growth in Hv-YPC or Hv-CA are listed in Supplementary Table S1. The water used for preparing media was deionized water from a diH₂O system. Solid agar media were prepared by addition of (1.5%) Bacto Agar (Difco 214010). Plates with strains were incubated at 45°C in a plastic container or plastic bag to limit evaporation.

Two methods were used for liquid cultures. For culture volume <5 mL, cells were grown in 15 mL glass tubes in a rotating platform and for cultures >5 mL, cells were grown in an Erlenmeyer flask covered with aluminum foil on a shaking platform. All cultures were grown at 45°C with a rotation of 110 rpm.

Transformation of *H. volcanii* via the Polyethylene glycol 600 (PEG600) method was performed as described before (Allers et al., 2004) and appropriate selection was used depending on the strain's genetic background and the selection marker on the transforming plasmid. Non-methylated plasmids were extracted from *Escherichia coli dam⁻/dcm⁻* (C2925I, NEB) prior to transformation of *H. volcanii*. *E. coli* strains were cultured in LB medium, with the necessary antibiotics (100 µg/mL ampicillin, 30 µg/mL chloramphenicol, or 25 µg/mL kanamycin) and grown at 37°C under constant shaking.

2.2. Growth curves of *Haloferax volcanii* strains

Strains from glycerol stocks were streaked on solid agar medium plus auxotrophic requirements as needed, or, for strains with plasmids, on HV-CA agar plus required additives (see [Supplementary Table S1](#)), and then incubated at 45°C for 4–5 days to obtain isolated colonies. A single colony from plates was used to inoculate 5 mL of media on day 1. On day 2, the obtained culture was used to inoculate 50 mL of Hv-CA supplemented with appropriate additives (see [Supplementary Table S1](#)) so that the culture would reach an OD₆₀₀ of 0.05 on the morning of day 3 and incubated at 45°C with shaking at 120 rpm. Measurement of optical density at 600 nm (OD₆₀₀) was started on day 3 at an OD₆₀₀ of 0.05 and manual sampling was performed up to 70 h.

2.3. Studies of *Haloferax volcanii* cell shape

2.3.1. Growth conditions and sampling

Strains were streaked on Hv-YPG media for strains without a plasmid (see [Supplementary Table S1](#)) and appropriate Hv-CA media for strains with plasmids (see [Supplementary Table S1](#)) and grown for 4–5 days to obtain isolated colonies. A single colony from plate was inoculated in 5 mL of appropriate Hv-CA media for all the strains and grown overnight. The following day, the 5 mL culture was used to inoculate 15 mL of Hv-CA media in a 100 mL Erlenmeyer flasks with foil lid so that the culture would reach an OD₆₀₀ of 0.01 the following morning. Samples were collected the following day at OD₆₀₀ of 0.01, 0.03, 0.06, 0.1 and 0.2 and prepared for phase-contrast light microscopy.

2.3.2. Phase-contrast light microscopy

A 5 µL culture sample was dropped onto an agarose pad [1% (w/v) agarose (Roth 3810.3), dissolved in 18% SW] and covered with a cover slip upon drying of the surface liquid. Micrographs were acquired on an inverted phase-contrast light microscope (Zeiss Axio Observer Z.1) at a magnification of 100X with an oil-immersion objective (Zeiss Plan-Apochromat 100×/1.40 NA Oil M27, with Immersol™ 518 F) and an exposure of 50 ms.

2.3.3. Image and data analysis

Phase-contrast images were analyzed using Fiji combined with the MicrobeJ plugin (Schindelin et al., 2012; Ducret et al., 2016). Segmentation of archaeal cells was performed in MicrobeJ using the following settings: default mode for thresholding, scale 0.065 µm and area cutoff: 0.1—max and cell shape descriptors (circularity, length, area, and aspect ratio) were collected. For the analysis, cells which are present as aggregates or fragmented were manually discarded from the segmentation results.

For each strain and OD₆₀₀, cell circularity was plotted as a frequency distribution obtained from the total number of cells analyzed (calculated and plotted in GraphPad Prism 6). The distribution was grouped into bins between 0 to 1 with increments of 0.1. Data were represented as frequency polygons and presented in a ridgeline plot to allow comparisons. Violin plots of the different cell shape parameters were created in GraphPad Prism 6, and plotted with the mean and standard deviation overlaid as a bar graph.

Categorization of cells into shape types rod (R), intermediate (I) and plates (P) was done in Excel with IF/AND functions based on the cell circularity features extracted from the segmentation/analysis in MicrobeJ and using the following parameters: a circularity between 1 and 0.8 included (closest to a circle) are corresponding to plate cells (P), values strictly in between 0.8 and 0.6 correspond to a mixed population of intermediate shapes (I) and values equal or below 0.6 (closest to an elongated shape) are rod cells (R). Percentages of each shape types from the total number of cells analyzed at a given OD₆₀₀ were calculated in Excel and plotted as bar graphs.

Statistical analyses were performed in GraphPad Prism 6. If comparison was between two groups, an unpaired non-parametric Mann–Whitney test was carried out. For comparison between three or more groups, a non-parametric Kruskal–Wallis and Dunn's multiple comparison tests were carried out. For the generation of the *p*-value heatmaps, *p*-values were calculated in python using SciPy (scipy.stats.kruskal) function and plotted as heatmaps with the seaborn package.

3. Results

3.1. Rod shape is transient during growth of wild-type *Haloferax volcanii* DS2

Cell shape development during the growth phases of *H. volcanii* has only been reported for two of the commonly used auxotrophic strains, including H26 and H53 (Abdul Halim et al., 2016; Li et al., 2019; de Silva et al., 2021, [Supplementary Table S2](#)). To better characterize *H. volcanii* cell shape during growth in liquid medium and study the potential effects of the auxotrophic mutations on *H. volcanii* cell shape, we first studied the *H. volcanii* DS2 parental wild-type strain. Cells were grown in liquid Hv-CA medium, sampled at different stages of growth (OD₆₀₀ 0.01, 0.03, 0.06, 0.1, and 0.2) and imaged by phase-contrast microscopy to obtain cell shape parameters, including cell length, aspect ratio, area, and circularity ([Figure 1](#)). Cell circularity was used as a proxy to characterize the degree of cell elongation and cells were categorized as plate cells (P, circularity 0.8–1), intermediate cells (I, circularity 0.6–0.8) or rod cells (R, circularity <0.6).

H. volcanii DS2 was pleomorphic throughout the different growth stages, showing a mixture of rods, intermediate and plate cells, with

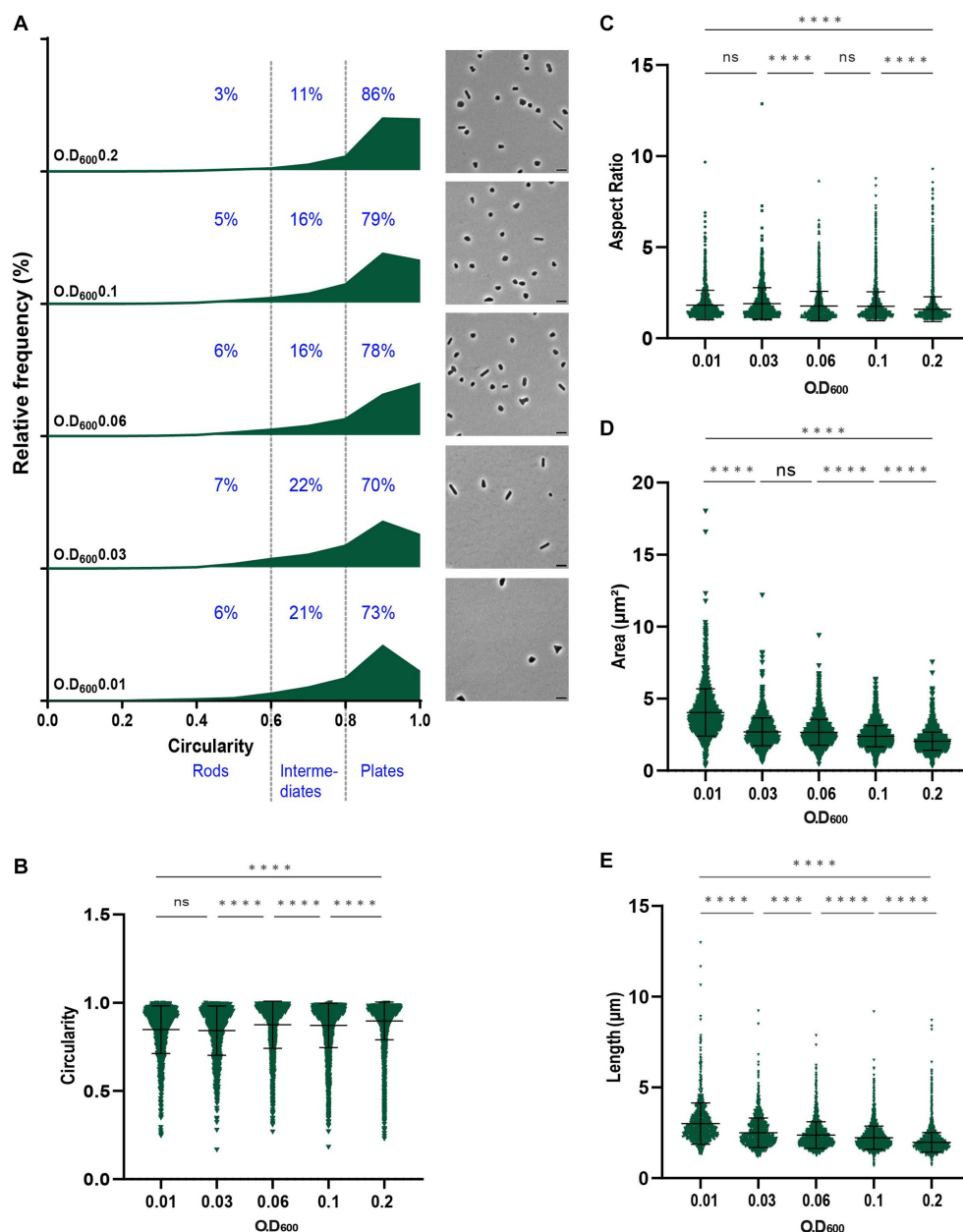


FIGURE 1

Analysis of the morphology of *H. volcanii* DS2 at different stages of growth. **(A)** Relative frequency distribution of cell circularity (left) measured from phase contrast image analysis (right) of DS2 samples collected at OD₆₀₀ 0.01 to 0.2 (bottom to top). Cell types R (rods), I (intermediates) and P (plates) are determined depending on the cell circularity as described in the text and each cell shape percentage is indicated on the graphs. The Y-axis indicates the percentage of cells at each OD. Sum of the graph height per OD₆₀₀ equals 100%. Scale bars on micrographs represent 4 μm. **(B)** Violin plot distribution of cells circularity at different OD₆₀₀. Data set is the same as in **(A)**. **(C)** Violin plot distribution of cells aspect ratio at different OD₆₀₀. **(D)** Violin plot distribution of cells area (μm²) at different OD₆₀₀. **(E)** Violin plot distribution of cells length (μm) at different OD₆₀₀. The statistical analysis in **(B–E)** were performed using Kruskal–Wallis-test in GraphPad Prism and data represent more than 1,100 cells from three independent experiments. Black line indicates mean; bottom and top lines indicate the standard deviation. Additional results of Kruskal–Wallis-tests are represented in [Supplementary Figure S5](#).

the majority of cells (~75%) existing as plates ([Figure 1A](#)). The proportion of plate cells increased as the culture progressed, with up to 86% of the population classified as plates at OD₆₀₀ = 0.2 ([Figure 1A](#)). These data can also be seen as a significant increase of the mean of the circularity ([Figure 1B](#)) and a decrease of the mean of the aspect ratio ([Figure 1C](#)). This suggested that the formation of rods in early log phase, as described earlier in H26 and H53 ([Li et al., 2019; de Silva et al., 2021](#)), is also characteristic for the wild-type *H. volcanii* DS2.

Cell area ([Figure 1D](#)) and length ([Figure 1E](#)) measurements revealed a decrease in cell size during culture progression, that was also seen in each of the three shape classes ([Supplementary Figure S1](#)). Notably, rod cells and intermediate cells were significantly larger than plate cells at the same stages ([Supplementary Figure S1](#)). Intermediate cells generally had a size in between rod and plate cells, suggesting they may be in transition between plate and rod shapes.

3.2. Differing cell shape transitions in wild-type and auxotrophic *Haloferax volcanii* strains

Many of the common genetic tools of *H. volcanii* are based on selection via auxotrophic marker genes (Allers et al., 2004) and recombinant plasmids based on the small endogenous plasmid pHV2, which has been cured from wild-type DS2 (Wendoloski et al., 2001). Previous studies have also noticed some *H. volcanii* strains would form more rods with plasmid compared to those without plasmid. These observations, summarized in Supplementary Table S2, were made in the background strains H26 ($\Delta pyrE2$), H53 ($\Delta pyrE2\Delta trpA$), and H98 ($\Delta pyrE2\Delta hdrB$) bearing plasmids based on pTA962/pTA963 derivatives containing the selection marker cassette $p_{fdx-pyrE2::hdrB}$ (Allers et al., 2004). Intrigued by the facts that (1) all the strains in which this phenomenon was observed share the $\Delta pyrE2$ mutation and that the strains H53 and H98 are derived from H26 and (2) that all the plasmids used were pTA962/963 derivatives with the same $p_{fdx-pyrE2::hdrB}$ selection cassette, we investigated whether the genetic background or the markers present on the plasmid could be responsible for the plasmid-dependent cell shape changes.

To test whether the auxotrophic mutations affect cell shape and may thereby be linked to the plasmid-mediated effects previously observed, we compared the cell shape development of five commonly used auxotrophic strains: H26 ($\Delta pyrE2$), H53 ($\Delta pyrE2\Delta trpA$), H98 ($\Delta pyrE2\Delta hdrB$) as well as two other independent strains H77 ($\Delta trpA$) and H729 ($\Delta hdrB$) to wild-type DS2 (Supplementary Table S1).

For the strains H26, H53 and H77, most cells were plates, representing more than 65% of the population at all stages of culture sampled similarly to DS2 (Figures 1, 2). The proportion of rods and intermediate cells decreased as the culture progressed into the exponential phase and cells transitioned to plate shape (~93% plates at $OD_{600}=0.2$, Figures 2B,C). H26 showed the lowest proportion of rods and intermediate cells and this proportion decreased over time whereas most of the transition to plate cells occurred somewhat later, between $OD_{600}=0.06$ and $OD_{600}=0.1$, for H53 and H77.

Strains H98 and H729, however, showed a greater proportion of rods and intermediate cells compared to DS2, H26, H53, and H77 (Figure 2) and kept a ratio of around 50% rods and intermediate cells from $OD_{600}=0.01$ to $OD_{600}=0.06$ (Figures 2B,C). For H98, an abrupt transition to plate shape happened between $OD_{600}=0.1$ to $OD_{600}=0.2$ changing from 64% to 95% plates. However, H729 still displayed around 38% of rods and intermediate cells for 62% of plates at $OD_{600}=0.2$. This might suggest that H729 is delayed in the transition to plate shape compared to H98 and to the other strains.

In addition to differences in the ratio between rods, intermediates and plates, our study showed differences of cell size between the strains and over time. Similar to DS2, the strains H53, H77, H729, and H98 showed the tendency to become smaller (see area Supplementary Figure S2C and length Supplementary Figure S2D) as the OD_{600} increased. However, H26 showed relatively small cells at $OD_{600}=0.01$ and were larger by $OD_{600}=0.2$ (see area Supplementary Figure S2C and longer Supplementary Figure S2D). The cells of H98, H53, H77, and H729 were overall larger than DS2 throughout, with H98 being the smallest of the four and H729 the largest (Supplementary Figures S2C,D). No substantive differences in the growth curves of these strains were observed (Supplementary Figure S3A). The H729 strain displayed the highest number of rods at all ODs (Figure 2C)

and the highest aspect ratio and length (Supplementary Figures S2B,D) showing that this strain is also more elongated than the others; H729 rods cells sometimes appeared irregular and filamentous (Figure 2A). This phenotype will be discussed further below.

Overall, we observed that for all the strains tested, the general tendencies of *H. volcanii* cells are consistent with the previous studies of single strains: (1) the cells appear generally pleomorphic throughout the monitored growth period (early-mid log phase), (2) rods cells are more prevalent during the earlier stages of exponential growth in the conditions used, and (3) the cells of most strains become smaller as the OD increases. However, strain dependent features appear, including (1) the proportion of rods, intermediate and plate cells, (2) how long rod cells persist in the culture, (3) the size of the cells and (4) how elongated rods are.

3.3. Rod development is both plasmid and background strain dependent

3.3.1. Cell shape in H26 backgrounds

We then analyzed the effect of various plasmids carrying different selection markers on the cell shape of the auxotrophic strains H26, H53, H98, H77, and H729.

H26 ($\Delta pyrE2$) was transformed with pTA1392, a derivative from pTA962 containing the selection cassette $p_{fdx-pyrE2::hdrB}$ (Haque et al., 2020), or pTA230, containing the selection cassette $p_{fdx-pyrE2}$ (Allers et al., 2004).

As previously observed in H26 pTA962 (de Silva et al., 2021), the presence of pTA1392 or pTA230 resulted in a greater proportion of rod and intermediate cells during the early logarithmic phase compared to H26 without the plasmid (Figures 3A,B). No obvious rod-to-plate transition timing could be observed for H26 without plasmid, and rod shape was clear in H26 strains bearing a plasmid, which persisted to $OD_{600}=0.2$ (Figures 3A,B).

The presence of the plasmid in the H26 background affected the proportions of the different cell shapes and the cell dimensions. While the H26 area increased overtime, the opposite happened for H26 pTA1392 and H26 pTA230 (Figure 3C) which is more similar to what happened in DS2 (Figure 1D) and the other backgrounds strains studied in Supplementary Figure S2C. The two plasmid-bearing strains were overall more elongated than H26 without plasmid (Figures 3D,E) and showed lower circularity (Supplementary Figure S4A). H26 pTA1392 was marginally the most elongated of the three strains (Figures 3D,E). Note that H26 pTA230 and H26 pTA1392 were both grown in Hv-CA without additional uracil as the $p_{fdx-pyrE2}$ marker was present on the plasmids. The main difference between these two conditions is the additional presence of the *hdrB* marker expressed from p_{fdx} on pTA1392. This increases the gene copy number of the *hdrB* gene that is present both in the chromosome and the plasmid, expected to cause overexpression of the gene. These results would suggest that the additional expression of *hdrB* differentially impacts *H. volcanii* cell shape.

3.3.2. Cell shape in H53 and H77 backgrounds

We complemented H53 ($\Delta pyrE2\Delta trpA$) with pTA230 ($p_{fdx-pyrE2}$), pTA231 ($p_{fdx-trpA}$) and pTA1392 ($p_{fdx-pyrE2::hdrB}$), similar to

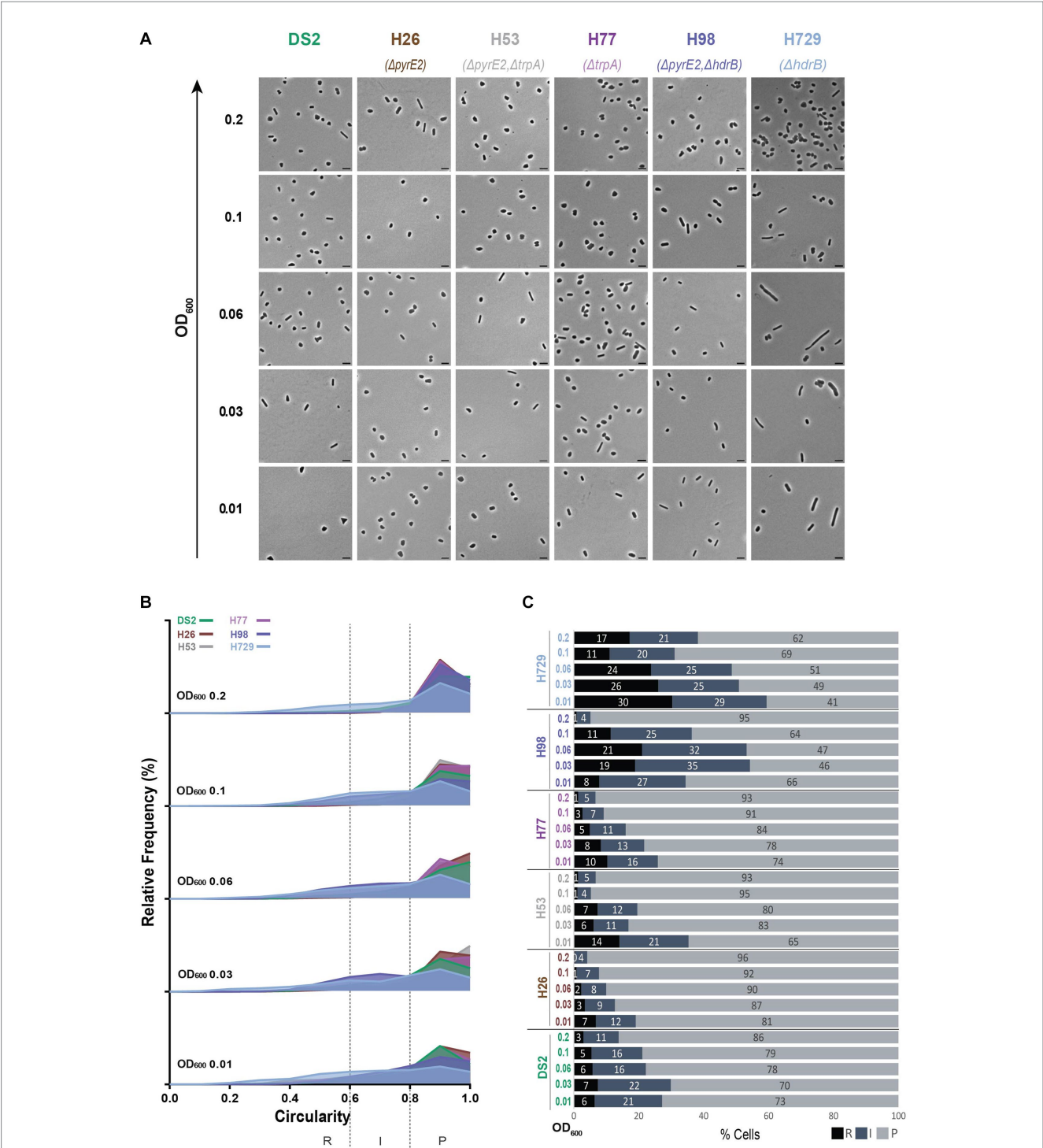


FIGURE 2 Morphological analysis of different auxotrophic *H. volcanii* strains at different stages of growth. **(A)** Phase contrast micrographs showing DS2, H26, H53, H77, H98 and H729 sampled at different growth stages from OD₆₀₀ 0.01 (bottom) to 0.2 (top). Scale bars represent 4 μm. The genotype of each strain (coloured) is indicated at the bottom of the micrographs. **(B)** Relative frequency distribution of cell circularity measured from micrographs in **(A)** at OD₆₀₀ 0.01 to 0.2 (bottom to top). Dashed vertical lines delimit the different cell types R (rods), I (intermediates) and P (plates) determined depending on the cell circularity as described in the text. The Y-axis indicates the percentage of cells at each OD. Sum of the graph height per OD₆₀₀ equals 100%. **(C)** Bar graph indicating the percentage of each cell shape types at each sampled OD₆₀₀ for each strain background. Colour representation—R (rods): black, I (intermediate): blue and P (plates): grey. **(B,C)** (N > 500 from three independent experiments). Data set for strain DS2 is the same as in Figure 1.

the plasmids used in previous works (Abdul Halim et al., 2016; de Silva et al., 2021, Supplementary Table S2), and H77 (ΔtrpA) was complemented with pTA231 (p_{tdx}-trpA). In the H53 background, the plasmids resulted in a higher proportion of rods and intermediate cells, noticeable particularly with pTA1392 (Figures 4A,B; Supplementary Figure S4B). For H53 pTA1392 and H53 pTA230, the

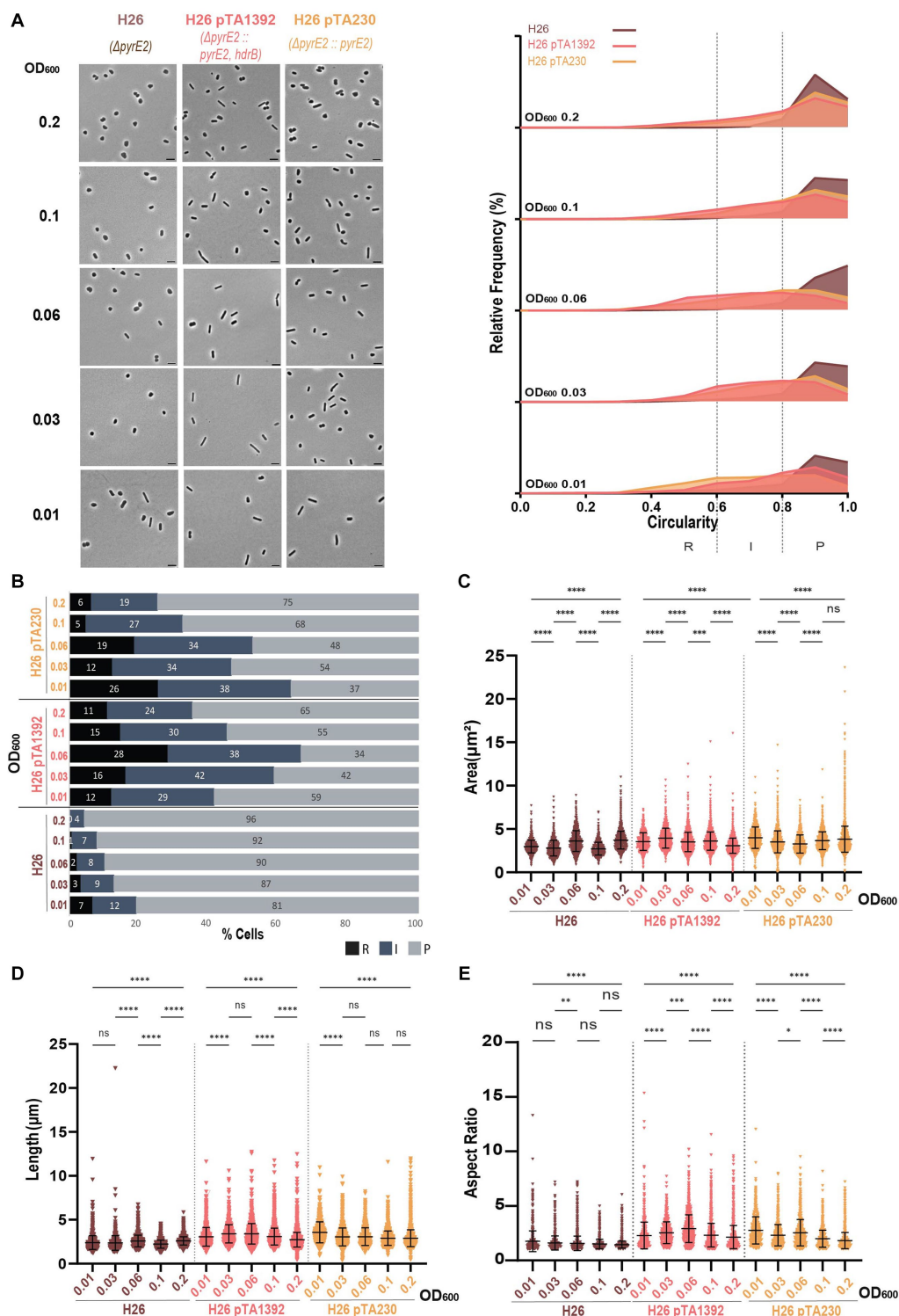


FIGURE 3

Morphological analysis of H26 with plasmids compared to H26 without plasmid at different stages of growth. (A) Left—phase contrast micrographs showing H26, H26 pTA1392, H26 pTA230 sampled at different growth stages from OD₆₀₀ 0.01 (bottom) to 0.2 (top). Scale bars represent 4 μ m. The genotype of each strain (coloured) is indicated at the bottom of the micrographs. Right—Relative frequency distribution of cell circularity measured from micrographs in (A) at OD₆₀₀ 0.01 to 0.2 (bottom to top). Dashed vertical lines delimit the different cell types R (rods), I (intermediates) and P (plates) determined depending on the cell circularity as described in the text. The Y-axis indicates the percentage of cells at each OD. Sum of the graph height per OD₆₀₀ equals 100%. (B) Bar graph indicating the percentage of each cell shape types at each sampled OD₆₀₀ for each strain background. Colour representation—R (rods): black, I (intermediate): blue and P (plates): grey. (C) Violin plot distribution of cells area (μ m²) at different OD₆₀₀. (D) Violin plot distribution of cell length (μ m) at different OD₆₀₀. (E) Violin plot distribution of cells aspect ratio at different OD₆₀₀. The statistical analysis in (C–E) were performed using Kruskal–Wallis-test in GraphPad Prism and data represent more than 1,300 cells from three independent experiments. Black line indicates mean; bottom and top lines indicate the standard deviation. Additional results of Kruskal–Wallis-tests are represented in [Supplementary Figure S6](#). Data set for strain H26 is the same as in [Figure 1](#).

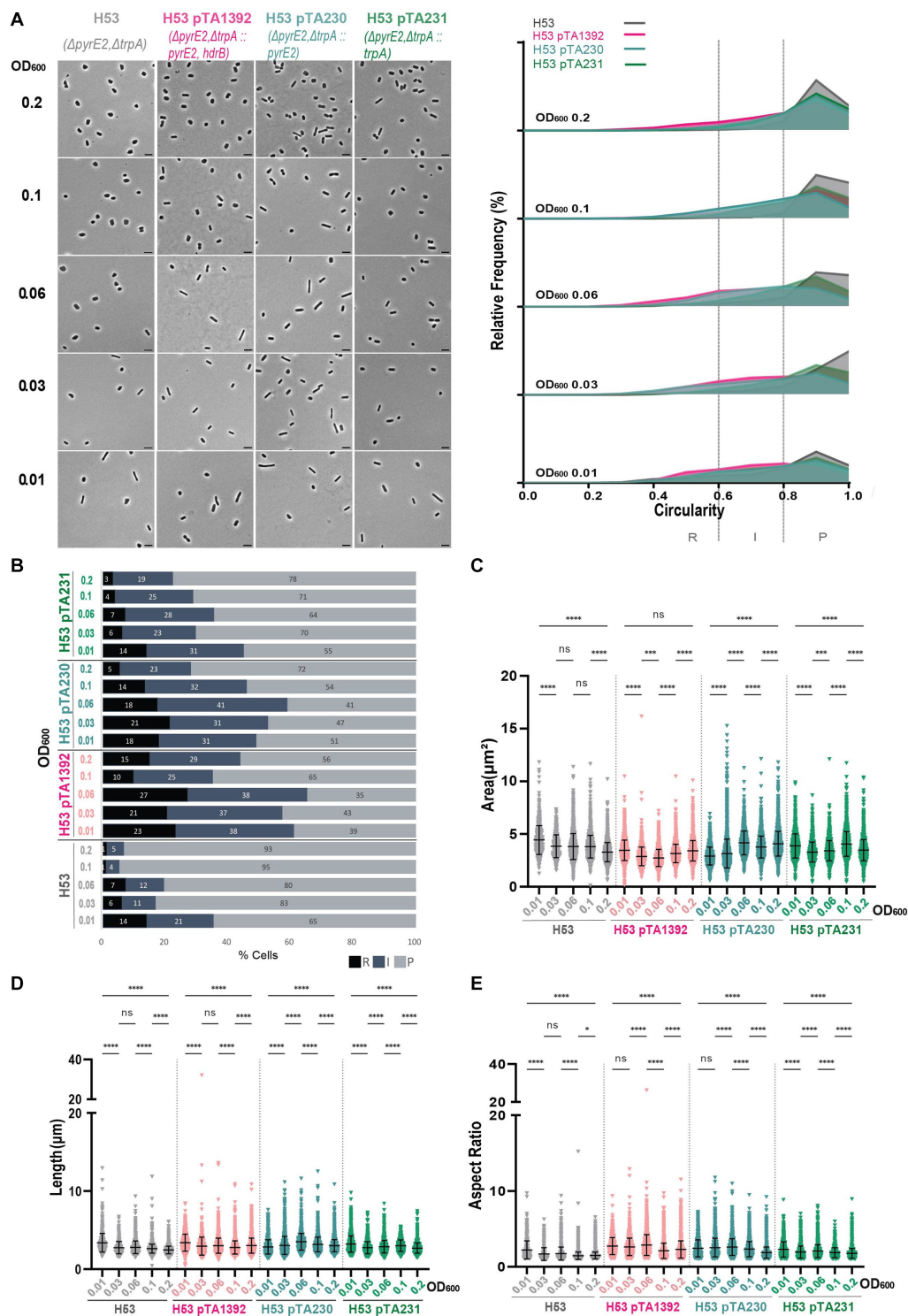


FIGURE 4

Morphological analysis of H53 with plasmids compared to H53 without plasmid at different stages of growth. (A) Left—phase contrast micrographs showing H53, H53 pTA1392, H53 pTA230, H53 pTA231 sampled at different growth stages from OD₆₀₀ 0.01 (bottom) to 0.2 (top). Scale bars represent 4 μ m. The genotype of each strain (coloured) is indicated at the bottom of the micrographs. Right—relative frequency distribution of cell circularity measured from micrographs in (A) at OD₆₀₀ 0.01 to 0.2 (bottom to top). Dashed vertical lines delimit the different cell types R (rods), I (intermediates) and P (plates) determined depending on the cell circularity as described in the text. The Y-axis indicates the percentage of cells at each OD. Sum of the graph height per OD₆₀₀ equals 100%. (B) Bar graph indicating the percentage of each cell shape types at each sampled OD₆₀₀ for each strain background. Colour representation—R (rods): black, I (intermediate): blue and P (plates): grey. (C) Violin plot distribution of cells area (μ m²) at different OD₆₀₀. (D) Violin plot distribution of cell length (μ m) at different OD₆₀₀. (E) Violin plot distribution of cells aspect ratio at different OD₆₀₀. The statistical analysis in (C–E) were performed using Kruskal–Wallis-test in GraphPad Prism and data represent more than 500 cells from three independent experiments. Black line indicates mean; bottom and top lines indicate the standard deviation. Additional results of Kruskal–Wallis-tests are represented in [Supplementary Figure S7](#). Data set for strain H53 is the same as in [Figure 1](#).

proportion of rods and intermediate cells increased until $OD_{600}=0.06$ to reach a peak of 65% and 59%. In between $OD_{600}=0.06$ and $OD_{600}=0.1$ the population cell shape transition occurred, and the proportion of cell shapes reversed in favor of plate cells. However, for H53 pTA231 the peak of rods and intermediates happened early on at $OD_{600}=0.01$ and a rather smooth decrease of the proportion of rod cells was visible up to $OD_{600}=0.2$.

Interestingly, H53 pTA1392 cells were generally smaller than the cells of the other strains (see area and length Figures 4C,D) but their aspect ratio was relatively high (Figure 4E) suggesting that in H53, pTA1392 results in greater elongation and thinning of the cells. The similarity to the results for H26 pTA1392 would suggest a similar effect of the additional *hdrB* marker on the cell morphology in strains deleted for *pyrE2*.

The presence of pTA231 in H77 induced a relatively high proportion of rods and intermediate cells starting from 63% at $OD_{600}=0.01$ and the transition towards a majority of plate cells occurred in between $OD_{600}=0.06$ and $OD_{600}=0.1$ (Figures 5A,B; Supplementary Figure S4C). Interestingly in the case of the H77 background, the presence of the plasmid did not increase the cell area (Figure 5C). However, cells bearing pTA231 are generally longer (Figure 5D) and have a higher aspect ratio than H77 without plasmid (Figure 5E) suggesting that in this strain, the presence of the plasmid pTA231 induces a remodeling of the cell towards elongated rods.

3.3.3. Cell shape in H729 and H98 backgrounds

The same experiments were conducted for H98 ($\Delta pyrE2\Delta hdrB$) transformed with pTA1392 ($p_{fdx-pyrE2::hdrB}$), pTA230 ($p_{fdx-pyrE2}$) and pTA233 ($p_{fdx-hdrB}$), and H729 ($\Delta hdrB$) transformed with pTA1392 and pTA233.

For H729, which had the largest cells compared to the other strains without plasmid and showed the highest proportion of rods and intermediate cells (Figure 2; Supplementary Figure S2), the presence of the plasmids pTA1392 or pTA233 did not increase the percentage of rods and intermediate cells (Figures 6A,B). To the contrary, at $OD_{600}=0.2$, H729 pTA1392 and H729 pTA233 showed less rods than H729 without plasmid. Additionally, the cells harboring pTA1392 and pTA233 were globally smaller, shorter, less elongated and more circular than H729, pTA233 conferring the smaller size and highest circularity (Figures 6C–E; Supplementary Figure S4E).

The difference between H729 without plasmid and with plasmid, might suggest a fundamental difference between the addition of thymidine and hypoxanthine in the media and the expression of *hdrB* from the plasmid for supporting H729 cell development in Hv-CA. However, it is interesting to note that differences could be observed between H729 pTA1392 and H729 pTA233 in terms of cell size and growth despite their same growth medium (Figure 6; Supplementary Figure S3F). These differences might be due to the extra copy of *pyrE2* expressed from pTA1392 or that *hdrB* is transcribed from the $p_{fdx-pyrE2::hdrB}$ cassette in a multicistronic mRNA and that the expression of *hdrB* in that context is most likely different than for $p_{fdx-hdrB}$ from pTA233.

Interestingly, H98 pTA233 grown in Hv-CA + uracil also showed cells that were smaller, shorter, and more circular compared to H98 or H98 pTA1392 (Figures 7C,D; Supplementary Figure S4D). This further supports the idea that an external addition of 40 $\mu\text{g/mL}$ of thymidine and hypoxanthine, or the expression of *hdrB* from $p_{fdx-pyrE2::hdrB}$ or $p_{fdx-hdrB}$ does not restore the cell's prototrophy to the

same extent and differently affects the growth and the cell shape development in $\Delta hdrB$ strains as compared to the wildtype (Figure 7; Supplementary Figures S3, S4D).

The presence of pTA233 in H98 did not drastically increase the percentage of rods and intermediate cells compared to H98 without plasmid but seemed to help maintain a rather stable proportion of the different cell shapes overtime. Indeed, H98 transitioned to plate shape between $OD_{600}=0.1$ and $OD_{600}=0.2$ to attain 95% of plates whereas H98 pTA233 still showed 33% of rods and intermediates at $OD_{600}=0.2$ (Figures 7A,B).

Both pTA1392 and pTA230 increased the percentage of rods and intermediate cells in H98 but in different manners (Figures 7A,B). While the percentage of rods and intermediate cells was largely increased compared to H98 at $OD_{600}=0.01$ for both strains, this percentage gradually decreased overtime for H98 pTA1392 down to 27% at $OD_{600}=0.2$ (transition between $OD_{600}=0.1$ and $OD_{600}=0.2$) whereas it stayed at high levels for H98 pTA230 (52%). Indeed, while pTA1392 reduced the overall size of the cells compared to H98, the presence of pTA230 greatly increased the size (area and length, Figures 7C,D) and cells were much more elongated (see aspect ratio and circularity, Figure 7E; Supplementary Figure S4D) than all the other H98 background strains. The genotypes and phenotypes of H98 pTA230 ($\Delta pyrE2\Delta hdrB/p_{fdx-pyrE2}$) resembled the one of H729 ($\Delta hdrB$), in the sense that the *pyrE2* deletion is compensated by the plasmidic expression of $p_{fdx-pyrE2}$ and addition of thymidine and hypoxanthine is required for growth in Hv-CA. In both cases, the cells were larger and formed irregular and elongated rods (Figures 7A,C–E). This phenotype might reflect the cumulative effect of the incomplete support of cell morphology development when thymidine and hypoxanthine are provided in the media and the presence of a plasmid.

4. Discussion

This study reports an analysis of the cell shape of the *H. volcanii* DS2 strain during the different stages of the exponential growth and could show that the rod development and cell shape transition is a natural behavior of *H. volcanii*. As previously seen in other genetic backgrounds, the development of rods is restricted to the early stages of exponential growth (Abdul Halim et al., 2016; Li et al., 2019; de Silva et al., 2021). However, the percentage of rods is relatively small for DS2 in the tested conditions. The analysis of the various cell shape parameters clearly showed that the cells become smaller as the culture goes through the exponential phase as seen earlier for other *Haloarchaea* (Schwarzer et al., 2021). A phenomenon that has been observed for batch cultures of bacteria too. Indeed, the cell mass as well as the growth rate of *E. coli* K-12 strain MG1655 cultivated in batch cultures in Luria-Bertani broth drop considerably from $OD_{600}=0.3$ and keep decreasing gradually until stationary phase (Sezonov et al., 2007) and cell length has been shown to decrease over time too (Shi et al., 2021). It was shown that this reduction of cell mass was concomitant with the impoverishment of utilizable carbon source in the media (Sezonov et al., 2007). While it has been shown that *E. coli* surface area/volume varies through time, the mechanisms that govern the control of cell width, length, volume and surface area are still debated (Shi et al., 2017; Colavin et al., 2018; Oldewurtel et al., 2021). Apart from nutrient availability, DNA replication could affect cell size and shape as observed in bacteria (Westfall and Levin, 2017).

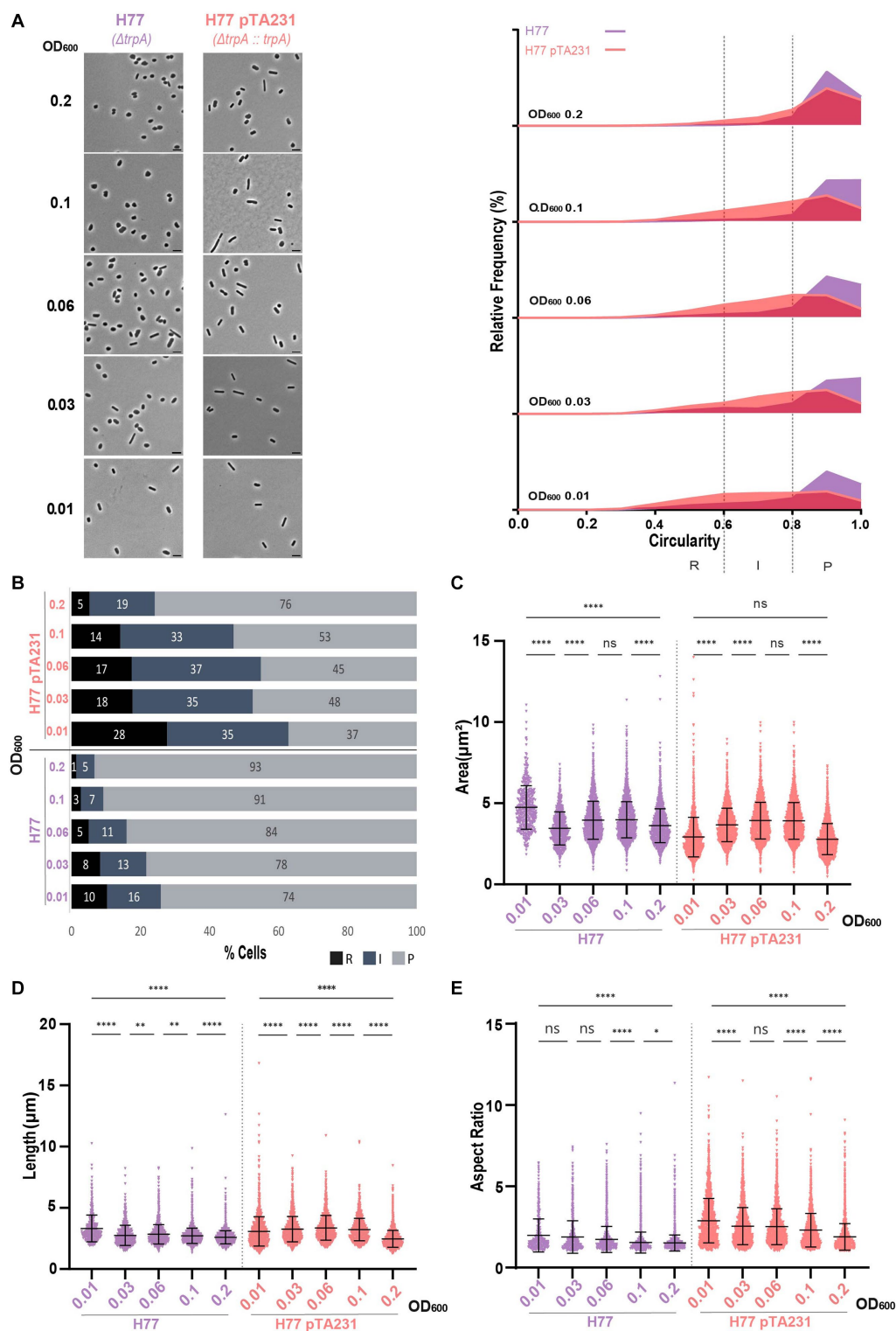


FIGURE 5

Morphological analysis of H77 with plasmids compared to H77 without plasmid at different stages of growth. **(A)** Left—phase contrast micrographs showing H77, H77 pTA231 sampled at different growth stages from OD₆₀₀ 0.01 (bottom) to 0.2 (top). Scale bars represent 4 μm . The genotype of each strain (coloured) is indicated at the bottom of the micrographs. Right—relative frequency distribution of cell circularity measured from micrographs in **(A)** at OD₆₀₀ 0.01 to 0.2 (bottom to top). Dashed vertical lines delimit the different cell types R (rods), I (intermediates) and P (plates) determined depending on the cell circularity as described in the text. The Y-axis indicates the percentage of cells at each OD. Sum of the graph height per OD₆₀₀ equals 100%. **(B)** Bar graph indicating the percentage of each cell shape types at each sampled OD₆₀₀ for each strain background. Colour representation—R (rods): black, I (intermediate): blue and P (plates): grey. **(C)** Violin plot distribution of cells area (μm^2) at different OD₆₀₀. **(D)** Violin plot distribution of cells length (μm) at different OD₆₀₀. **(E)** Violin plot distribution of cells aspect ratio at different OD₆₀₀. The statistical analysis in **(C–E)** were performed using Kruskal–Wallis-test in GraphPad Prism and data represent more than 500 cells from three independent experiments. Black line indicates mean; bottom and top lines indicate the standard deviation. Additional results of Kruskal–Wallis-tests are represented in [Supplementary Figure S8](#). Data set for strain H77 is the same as in [Figure 1](#).

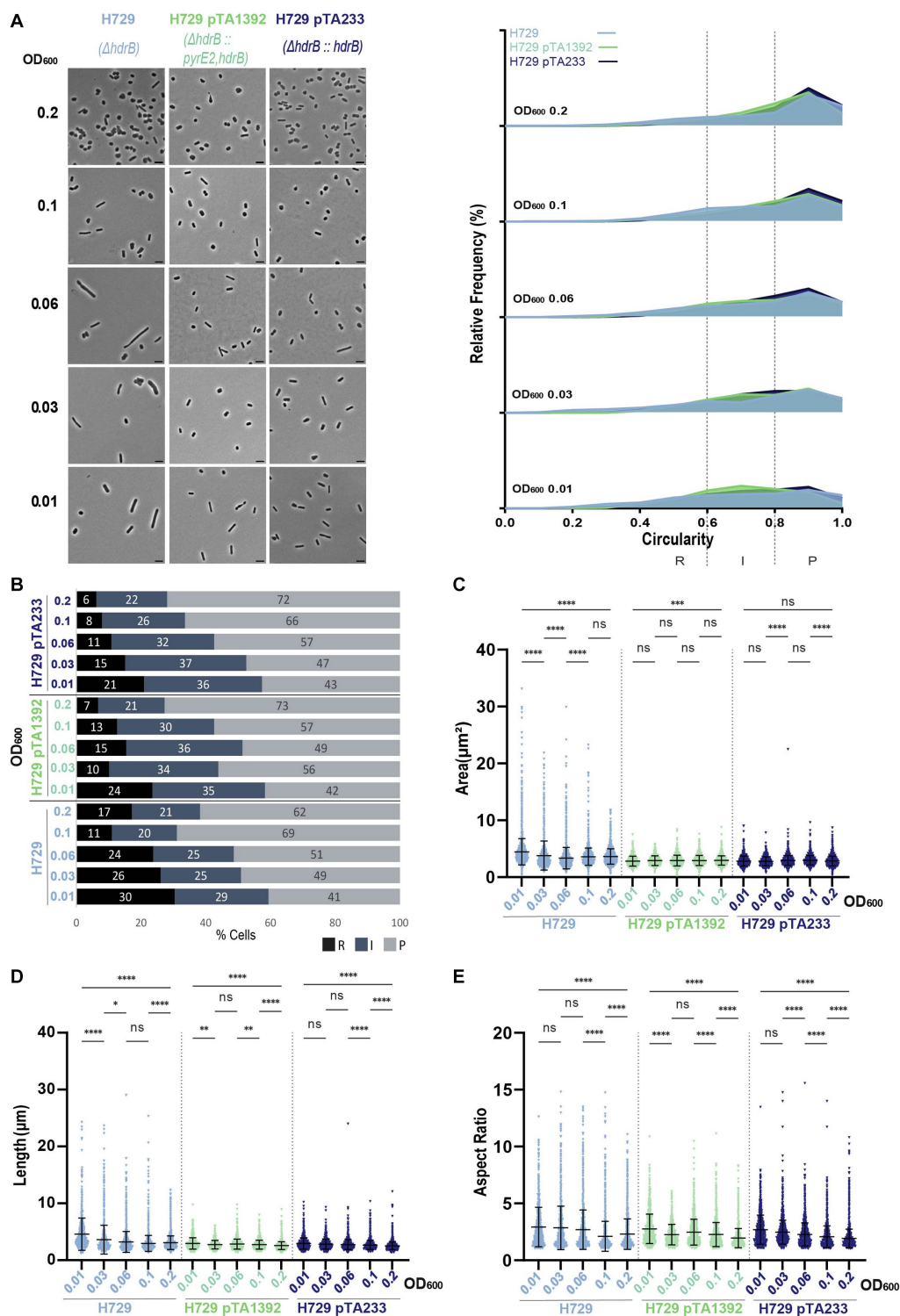


FIGURE 6

Morphological analysis of H729 with plasmids compared to H729 without plasmid at different stages of growth. **(A)** Left—phase contrast micrographs showing H729, H729 pTA1392, H729 pTA233 sampled at different growth stages from OD₆₀₀ 0.01 (bottom) to 0.2 (top). Scale bars represent 4 μm . The genotype of each strain (coloured) is indicated at the bottom of the micrographs. Right—relative frequency distribution of cell circularity measured from micrographs in **(A)** at OD₆₀₀ 0.01 to 0.2 (bottom to top). Dashed vertical lines delimit the different cell types R (rods), I (intermediates) and P (plates) determined depending on the cell circularity as described in the text. The Y-axis indicates the percentage of cells at each OD. Sum of the graph height per OD₆₀₀ equals 100%. **(B)** Bar graph indicating the percentage of each cell shape types at each sampled OD₆₀₀ for each strain background. Colour representation—R (rods): black, I (intermediate): blue and P (plates): grey. **(C)** Violin plot distribution of cells area (μm^2) at different OD₆₀₀. **(D)** Violin plot distribution of cell length (μm) at different OD₆₀₀. **(E)** Violin plot distribution of cells aspect ratio at different OD₆₀₀. The statistical analysis in **(C–E)** were performed using Kruskal–Wallis-test in GraphPad Prism and data represent more than 900 cells from three independent experiments. Black line indicates mean; bottom and top lines indicate the standard deviation. Additional results of Kruskal–Wallis-tests are represented in [Supplementary Figure S9](#). Data set for strain H729 is the same as in [Figure 1](#).

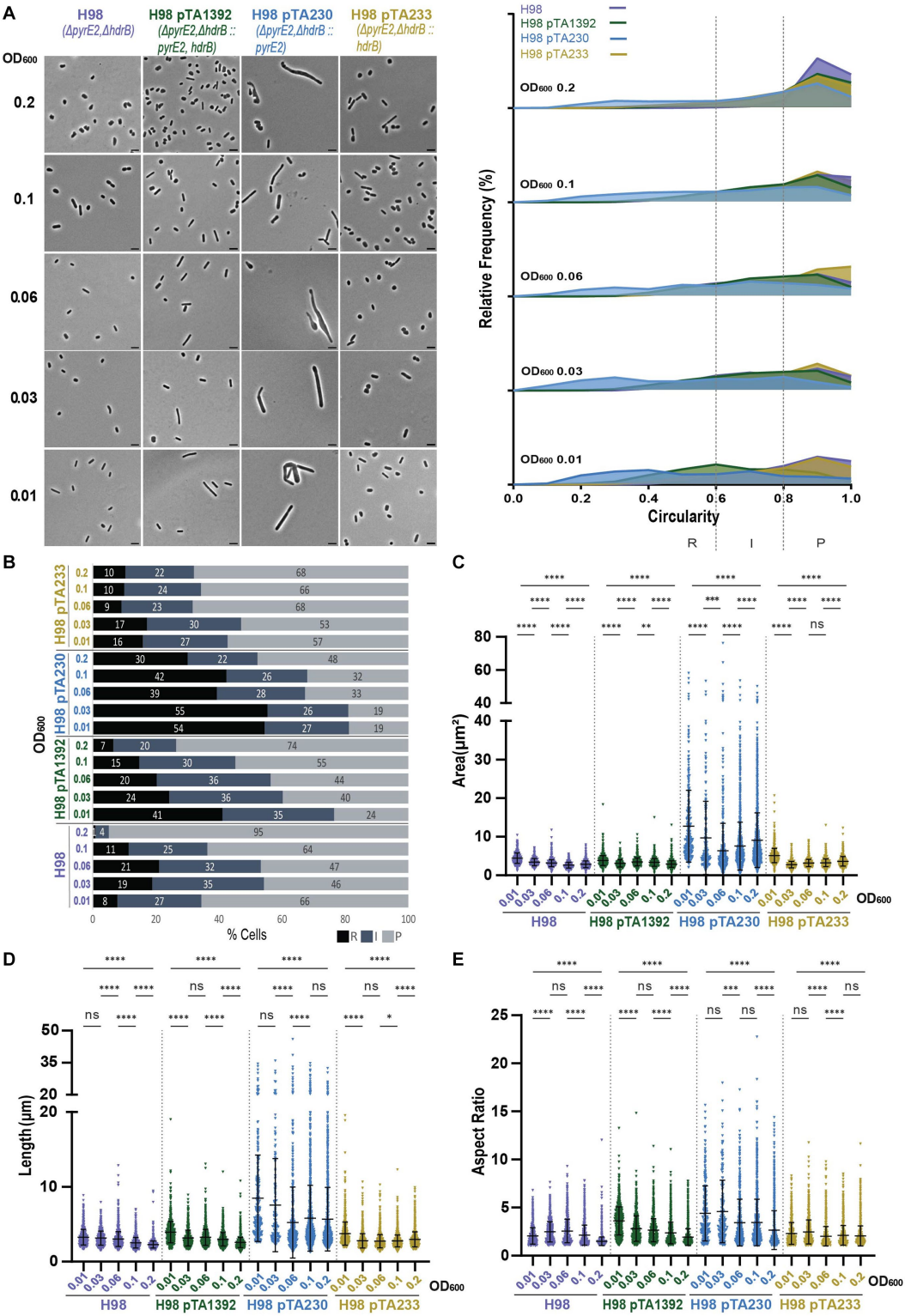


FIGURE 7 Morphological analysis of H98 with plasmids compared to H98 without plasmid at different stages of growth. **(A)** Left—phase contrast micrographs showing H98, H98 pTA1392, H98 pTA230, H98 pTA233 sampled at different growth stages from OD₆₀₀ 0.01 (bottom) to 0.2 (top). Scale bars represent 4 μm. The genotype of each strain (coloured) is indicated at the bottom of the micrographs. Right—relative frequency distribution of cell circularity measured from micrographs in **(A)** at OD₆₀₀ 0.01 to 0.2 (bottom to top). Dashed vertical lines delimit the different cell types R (rods), I (intermediates) and P (plates) determined depending on the cell circularity as described in the text. The Y-axis indicates the percentage of cells at each OD. Sum of the graph height per OD₆₀₀ equals 100%. **(B)** Bar graph indicating the percentage of each cell shape types at each sampled OD₆₀₀ for each strain background. Colour representation—R (rods): black, I (intermediate): blue and P (plates): grey. **(C)** Violin plot distribution of cells area (μm²) at different OD₆₀₀. **(D)** Violin plot distribution of cell length (μm) at different OD₆₀₀. **(E)** Violin plot distribution of cells aspect ratio at different OD₆₀₀. The statistical analysis in **(C–E)** were performed using Kruskal–Wallis-test in GraphPad Prism and data represent more than 1,700 cells from three independent experiments. Black line indicates mean; bottom and top lines indicate the standard deviation. Additional results of Kruskal–Wallis-tests are represented in [Supplementary Figure S10](#). Data set for strain H98 is the same as in [Figure 1](#).

H. volcanii has a multipartite and multicopy genome with multiple active origins of replication (for review Pérez-Arnaiz et al., 2020). So far the existence of a clear cell cycle in *H. volcanii* and the coupling of DNA replication, cell growth, cell division and chromosome segregation mechanisms are unknown. However, it is known that the genome copy number decreases between exponential phase and stationary phase from around 18 copies to around 10 copies (Breuert et al., 2006) so a possible link between the reduction of genome copy number and the reduction of cell size would be worth studying.

Additionally to a diminution of the population cell size over time, our study showed that the rod cells and intermediate cells in low OD cultures are also larger (larger area) than plate cells. It would be interesting to study how biomass is accumulated in the different cell shapes. *H. volcanii* cell envelope is constituted of an S-layer Glycoprotein and has been proposed to be newly incorporated at the division site (Abdul-Halim et al., 2020). However, nothing is known about how growth and S-layer incorporation are regulated, nor how this is coordinated with cell shape determination. A recent study in another haloarchaeon *Halobacterium salinarum* demonstrated that this species grows following the adder model where cells grow by adding a constant volume between two cell cycles (Eun et al., 2018). However, this study was conducted in conditions where all the cells were maintained in rod shape so there is currently no idea whether plate cells are growing the same way.

Another yet challenging side of *H. volcanii* cell shape plasticity is that the presence of a plasmid affects the cell shape of *H. volcanii* whether it is by supporting the development of rods (de Silva et al., 2021) or changing the morphological phenotypes of mutant strains (Abdul Halim et al., 2016; Liao et al., 2021; Nußbaum et al., 2021). Such effects may complicate studies of cell division, cell growth, cell morphology and cell motility where cell shape is an important component of the mechanisms studied. Based on this we sought to conduct a comparative study of the cell shape of the most commonly used background strains without and with a transformed plasmid in order to either find conditions where this would not happen or provide guidance for improving the current system. While we are aware that some limitations apply to our study, like for example the fact that each strain was cultivated in different media, i.e., Hv-CA supplemented adequately depending on the strains' auxotrophic state, and that direct comparison of them is somewhat questionable or the fact that we did not compare with the strain DS70 depleted from the plasmid pHV2 (Wendoloski et al., 2001) from which H26, H53, H77, H729, and H98 are derived from (Allers et al., 2004), we were able to highlight several patterns reflecting the presence or absence of a plasmid on *H. volcanii* cell shape. Most commonly, the plasmid increases the percentages of rods and intermediate cells compared to strains without plasmid. Sometimes it rather affects the dimensions of the cells (area, aspect ratio, length) (example H729) rather than changing the percentage of the different cell shapes. Other times it would change the timing during rods are developed and most of the times it is a combination of several of these phenomena. These phenotypes are most likely due to a combinatorial effect of the plasmid presence and of the different auxotrophic deletions and markers used for selection. Indeed, we have shown that the deletion of the various auxotrophic markers does have an impact on the cell shape, whether it is by increasing the percentage of rods and intermediate cells or/ and changing the cell dimensions independently of the presence of a plasmid. For example, the $\Delta trpA$ deletion (present in H53 and H77)

seems to increase the cell size but does not change the relative proportion of the cell types nor the timing of development of the rods compared to DS2.

The most striking impact on cell shape observed in this study was for the $\Delta hdrB$ deletion. Indeed, the deletion of *hdrB* itself induces a greater proportion of elongated cells, as seen both in H729 and H98 backgrounds. The cells are not only more elongated but also larger than DS2 especially for H729. A phenotype that could be partially reversed when the *hdrB* gene was provided on a plasmid. It shows that the external addition of thymidine and hypoxanthine required for growth of *hdrB* backgrounds in Hv-CA is not sufficient to support a proper cell shape (i.e., cell size and shape closer to DS2). The supplementation of thymidine and hypoxanthine to the medium is enough for the cells to grow (Supplementary Figure S3F) but most likely not sufficient to support normal cellular morphology thus cells tend to grow larger and more elongated and irregular as compared to DS2 for example. To the opposite, the complementation of *hdrB* by the plasmidic copy might better restore the prototrophy through endogenous biosynthesis that may be more efficient than uptake from the medium, and therefore most likely better restore the cell physiology. However, the complementation of the strains with plasmid carrying *hdrB* did not reduce the proportion of rods as one could have expected but to the opposite, it increased further the proportion of rods and intermediate cells in the *hdrB* background suggesting a combinatorial effect of the plasmid presence and the auxotrophy on cell shape in this background. Remarkably, for the H26 ($\Delta pyrE2$) and H53 ($\Delta pyrE2\Delta trpA$) strains that became merodiploids for *hdrB* by the transformation of pTA1392 ($p_{\text{fix}}\text{-pyrE2}::\text{hdrB}$), a greater proportion of rods and intermediate cells could be observed compared to the same strains transformed by other plasmids without the *hdrB* marker. Suggesting that in this case, the cumulative expression of the chromosomal and plasmid copies of *hdrB* does induce rod formation even more than other plasmids in these backgrounds. This suggests that the over expression of the *hdrB* gene severely induces the formation of rods additionally to the sole effect of the plasmid presence. This could also suggest that the expression of *hdrB* from the p_{fix} promoter on the plasmid could be too high as a stimulation of rod formation is observed too in strains H98 and H729 transformed with a plasmid. By extrapolation, we could hypothesize that the increase of rods percentage observed for the $\Delta hdrB$ strains grown in Hv-CA could be attributed to an over activation of the *hdrB* downstream pathways by addition of thymidine and hypoxanthine rather suggesting that the concentration of the additives could be too high.

HdrB encodes for a dihydrofolate reductase involved in the tetrahydrofolate biosynthesis which in turn participates in many reactions for the synthesis of major cellular components such as methionine, purine, pyrimidine and glycine. Such enzymes are found universally and often essential to cells, which makes its activity a common target for various anti-cancerous and antibacterial treatments (Hawser et al., 2006; Chawla et al., 2021). In *H. volcanii*, the deletion of *hdrB* is not lethal but causes a thymidine auxotrophy and deletion mutants require the addition of thymidine and hypoxanthine to grow in Hv-CA (Ortenberg et al., 2000; Falb et al., 2008). In bacteria, the level of thymidine has been shown to greatly affect the cells leading eventually to death under thymidine starvation conditions (Ahmad et al., 1998). It has also been shown in *E. coli* that both an under- and oversupply of thymidine in a Thy^- affect the activity of the ribonucleoside

diphosphate reductase (RNR) which in turn unbalances the pools of dNTPs and affect the replication velocity (Molina et al., 1998; Zhu et al., 2017). Additionally, a lower titration of RNR did increase the size of the cells by three times and some strains were shown to form filament under thymidine starvation (Ahmad et al., 1998; Zhu et al., 2017).

So based on these studies it is possible that in *H. volcanii*, the deregulation of the tetrahydrofolate pathway and downstream branches by either deletion of *hdrB*, the misuse of non-required *hdrb* marker on the plasmid or the under or over supplying of thymidine could have major effects on *H. volcanii* nucleotide metabolism inducing growth and cell shape deregulation. Additionally, the misregulation of the dihydrofolate reductase expression could have a direct impact on *H. volcanii* DNA replication. So far the existence of a clear cell cycle in *H. volcanii* and the coupling of DNA replication, cell division and chromosome segregation mechanisms are unknown but a potential link between replication, growth and therefore cell shape is not excluded.

Our current study does not allow for disentangling the origin of the phenotypes observed in $\Delta\textit{hdrB}$ mutant, but it does show that the *hdrB* deletion background is not suitable for experiments about shape and that the use of non-required *hdrb* selection marker on a plasmid should be strictly avoided. pTA1392 and pTA962 are widely used in the field and this even in non $\Delta\textit{hdrB}$ backgrounds (see examples in Supplementary Table S2) so we hope our study would convince the community of changing this practice. This would necessitate the community to develop more vectors free from the *hdrb* selection marker. However, our results show an intricate link between *H. volcanii* physiology and metabolism with cell shape that would be interesting to investigate further. In addition, the narrow period of rod development seen in batch culture seems to be sensitive to the constant availability of fresh media. Indeed, *H. volcanii* grown in the microfluidics system CellASIC ONIX remain rod, suggesting that the maintenance of rod is sensitive to either the impoverishment or the accumulation of a compound that could trigger the transition to plate shape in batch culture (de Silva et al., 2021).

As of why the sole presence of a plasmid is inducing rod formation, our study could not give an answer, but it is probable that the presence of the plasmid is a burden by itself changing the cell metabolism and affecting cell size and shape. All the studies referred to and done in this work are based upon plasmids containing the *H. volcanii* endogenous pHV2 plasmid origin (Charlebois et al., 1987; Wendoloski et al., 2001; Hartman et al., 2010). Proven to be a useful genetic tool, we however lack insights into the molecular mechanism of pHV2 origin.

As an alternative and for studies where cell shape is critical, we would highly recommend avoiding using a plasmid and favor the genomic integration for complementation or for expression of fluorescently tagged proteins for example. Gene deletions could be made by the replacement with the selection marker to restore the cells prototrophy without the necessity of a plasmidic marker. Additionally, the expression level of the selection cassette into the chromosome might resemble better the endogenous gene expression, or other promoters could be tested to fine tune the appropriate level of expression of the different selection markers to minimize impacts on cell shape. However, a plasmid free system would require the development of a genetic system that allows for the excision of the selection cassette from the genome in case of the construction of

multi locus mutant strains or complementation experiments for example. A strategy that would be worth developing to move the field forward.

Aside from considerations about genetic backgrounds and plasmid usage, we would like to highlight that our study was conducted in Hv-CA medium that was not supplemented by trace elements. Previous studies (Duggin et al., 2015; de Silva et al., 2021) showed that the addition of trace elements is crucial both in Hv-CA and Hv-YPG for the proper development of cell shape. To the contrary we observe regular cell shape in Hv-CA without trace elements. However, it is most likely that sufficient trace elements are present in the media. This could be due to the water used for media preparation in our team that is deionized water and not from an ultrapure purification system as it is the case for our colleagues. And as shown before, some trace elements might be present in sufficient amount in media preparation reagents depending on their degree of purity (de Silva et al., 2021). In addition to the natural plasticity of *H. volcanii* cell shape, such differences could render difficult the comparison and the reproduction of results between laboratories, it is certain that the research community would benefit from a homogenization of culture conditions.

Overall, we hope our study raised awareness to the community about the impact of the current genetic system used in *H. volcanii* on cell shape and on the importance of cautiousness in choosing a genetic background for a given study. We believe that with common efforts, new standards can be achieved by the development of new methods such as a plasmid free system, new selection markers and more standardized culture conditions for moving the haloarchaea cell biology field forward.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

MP: Conceptualization, Data curation, Formal analysis, Methodology, Visualization, Writing – original draft, Writing – review & editing, Investigation, Validation. ID: Conceptualization, Funding acquisition, Supervision, Writing – review & editing. S-VA: Conceptualization, Funding acquisition, Supervision, Writing – review & editing. SI: Conceptualization, Data curation, Formal analysis, Methodology, Visualization, Writing – original draft, Writing – review & editing, Supervision, Validation, Investigation, Project administration.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. SI was supported by a Momentum grant from the VW Science Foundation (grant number 94933) granted to S-VA. MP was supported by grant AL1206/4-3 by the German Science Foundation (DFG). ID was supported by the Australian Research Council (DP160101076 and FT160100010).

Acknowledgments

The authors acknowledge discussions with colleagues of the Albers and Duggin laboratories. We acknowledge support by the Open Access Publication Fund of the University of Freiburg.

Conflict of interest

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.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

References

- Abdul Halim, M. F., Karch, K. R., Zhou, Y., Haft, D. H., Garcia, B. A., and Pohlschroder, M. (2016). Permuting the PGF signature motif blocks both archaeosortase-dependent C-terminal cleavage and prenyl lipid attachment for the *Haloferax volcanii* S-layer glycoprotein. *J. Bacteriol.* 198, 808–815. doi: 10.1128/JB.00849-15
- Abdul-Halim, M. F., Schulze, S., DiLucido, A., Pfeiffer, F., Bisson Filho, A. W., and Pohlschroder, M. (2020). Lipid anchoring of archaeosortase substrates and midcell growth in haloarchaea. *mBio* 11, e00349–e00320. doi: 10.1128/mBio.00349-20
- Ahmad, S. I., Kirk, S. H., and Eisenstark, A. (1998). Thymine metabolism and thymineless death in prokaryotes and eukaryotes. *Annu. Rev. Microbiol.* 52, 591–625. doi: 10.1146/annurev.micro.52.1.591
- Allers, T., Ngo, H.-P., Mevarech, M., and Lloyd, R. G. (2004). Development of additional selectable markers for the halophilic archaeon *Haloferax volcanii* based on the *leuB* and *trpA* genes. *Appl. Environ. Microbiol.* 70, 943–953. doi: 10.1128/AEM.70.2.943-953.2004
- Bisson-Filho, A. W., Zheng, J., and Garner, E. (2018). Archaeal imaging: leading the hunt for new discoveries. *Mol. Biol. Cell* 29, 1675–1681. doi: 10.1091/mbc.E17-10-0603
- Breuert, S., Allers, T., Spohn, G., and Soppa, J. (2006). Regulated polyploidy in halophilic archaea. *PLoS One* 1:e92. doi: 10.1371/journal.pone.0000092
- Charlebois, R. L., Lam, W. L., Cline, S. W., and Doolittle, W. F. (1987). Characterization of pHV2 from *Haloferax volcanii* and its use in demonstrating transformation of an archaeobacterium. *Proc. Natl. Acad. Sci. U.S.A.* 84, 8530–8534. doi: 10.1073/pnas.84.23.8530
- Chawla, P., Teli, G., Gill, R. K., and Narang, R. K. (2021). An insight into synthetic strategies and recent developments of dihydrofolate reductase inhibitors. *ChemistrySelect* 6, 12101–12145. doi: 10.1002/slct.202102555
- Colavin, A., Shi, H., and Huang, K. C. (2018). RodZ modulates geometric localization of the bacterial actin MreB to regulate cell shape. *Nat. Commun.* 9:1280. doi: 10.1038/s41467-018-03633-x
- de Silva, R. T., Abdul-Halim, M. F., Pittrich, D. A., Brown, H. J., Pohlschroder, M., and Duggin, I. G. (2021). Improved growth and morphological plasticity of *Haloferax volcanii*. *Microbiology* 167:001012. doi: 10.1099/mic.0.001012
- Ducet, A., Quardokus, E. M., and Brun, Y. V. (2016). MicrobeJ, a tool for high throughput bacterial cell detection and quantitative analysis. *Nat. Microbiol.* 1, 1–7. doi: 10.1038/nmicrobiol.2016.77
- Duggin, I. G., Aylett, C. H. S., Walsh, J. C., Michie, K. A., Wang, Q., Turnbull, L., et al. (2015). CetZ tubulin-like proteins control archaeal cell shape. *Nature* 519, 362–365. doi: 10.1038/nature13983
- Eun, Y.-J., Ho, P.-Y., Kim, M., LaRussa, S., Robert, L., Renner, L. D., et al. (2018). Archaeal cells share common size control with bacteria despite noisier growth and division. *Nat. Microbiol.* 3, 148–154. doi: 10.1038/s41564-017-0082-6
- Falb, M., Müller, K., Königsmaier, L., Oberwinkler, T., Horn, P., von Gronau, S., et al. (2008). Metabolism of halophilic archaea. *Extremophiles* 12, 177–196. doi: 10.1007/s00792-008-0138-x
- Haque, R. U., Paradisi, F., and Allers, T. (2020). *Haloferax volcanii* for biotechnology applications: challenges, current state and perspectives. *Appl. Microbiol. Biotechnol.* 104, 1371–1382. doi: 10.1007/s00253-019-10314-2
- Hartman, A. L., Norais, C., Badger, J. H., Delmas, S., Haldenby, S., Madupu, R., et al. (2010). The complete genome sequence of *Haloferax volcanii* DS2, a model archaeon. *PLoS One* 5:e9605. doi: 10.1371/journal.pone.0009605
- Hawser, S., Locicuro, S., and Islam, K. (2006). Dihydrofolate reductase inhibitors as antibacterial agents. *Biochem. Pharmacol.* 71, 941–948. doi: 10.1016/j.bcp.2005.10.052
- Ithurbide, S., Gribaldo, S., Albers, S. V., and Pende, N. (2022). Spotlight on FtsZ-based cell division in archaea. *Trends Microbiol.* 30, 665–678. doi: 10.1016/j.tim.2022.01.005
- Li, Z., Kinoshita, Y., Rodriguez-Franco, M., Nußbaum, P., Braun, F., Delpech, F., et al. (2019). Positioning of the motility machinery in halophilic Archaea. *mBio* 10:e00377. doi: 10.1128/mBio.00377-19
- Liao, Y., Ithurbide, S., Evenhuis, C., Löwe, J., and Duggin, I. G. (2021). Cell division in the archaeon *Haloferax volcanii* relies on two FtsZ proteins with distinct functions in division ring assembly and constriction. *Nat. Microbiol.* 6, 594–605. doi: 10.1038/s41564-021-00894-z
- Molina, F., Jiménez-Sánchez, A., and Guzmán, E. C. (1998). Determining the optimal thymidine concentration for growing *Thy⁻ Escherichia coli* strains. *J. Bacteriol.* 180, 2992–2994. doi: 10.1128/JB.180.11.2992-2994.1998
- Mullakhanbhai, M. F., and Larsen, H. (1975). *Haloferax volcanii* spec. Nov., a dead sea halobacterium with a moderate salt requirement. *Arch. Microbiol.* 104, 207–214. doi: 10.1007/BF00447326
- Nußbaum, P., Gerstner, M., Dingethal, M., Erb, C., and Albers, S. V. (2020). Archaeal SepF is essential for cell division in *Haloferax volcanii*. *Microbiology* 12:3469. doi: 10.1101/2020.10.06.327809
- Nußbaum, P., Gerstner, M., Dingethal, M., Erb, C., and Albers, S. V. (2021). The archaeal protein SepF is essential for cell division in *Haloferax volcanii*. *Nat. Commun.* 12:3469. doi: 10.1038/s41467-021-23686-9
- Oldewurtel, E. R., Kitahara, Y., and van Teeffelen, S. (2021). Robust surface-to-mass coupling and turgor-dependent cell width determine bacterial dry-mass density. *Proc. Natl. Acad. Sci. U.S.A.* 118:e2021416118. doi: 10.1073/pnas.2021416118
- Ortenberg, R., Rozenblatt-Rosen, O., and Mevarech, M. (2000). The extremely halophilic archaeon *Haloferax volcanii* has two very different dihydrofolate reductases. *Mol. Microbiol.* 35, 1493–1505. doi: 10.1046/j.1365-2958.2000.01815.x
- Pérez-Arnaiz, P., Dattani, A., Smith, V., and Allers, T. (2020). *Haloferax volcanii*—a model archaeon for studying DNA replication and repair. *Open Biol.* 10:200293. doi: 10.1098/rsob.200293
- Pohlschroder, M., and Schulze, S. (2019). *Haloferax volcanii*. *Trends Microbiol.* 27, 86–87. doi: 10.1016/j.tim.2018.10.004
- Schiller, H., Kouassi, J., Hong, Y., Rados, T., Kwak, J., DiLucido, A., et al. (2023). Identification and characterization of structural and regulatory cell-shape determinants in *Haloferax volcanii*. *bioRxiv*. Available at: <https://doi.org/10.1101/2023.03.05.531186>. [Epub ahead of print].
- Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., et al. (2012). Fiji: an open-source platform for biological-image analysis. *Nat. Methods* 9, 676–682. doi: 10.1038/nmeth.2019
- Schwarzer, S., Rodriguez-Franco, M., Oksanen, H. M., and Quax, T. E. F. (2021). Growth phase dependent cell shape of haloarcula. *Microorganisms* 9:231. doi: 10.3390/microorganisms9020231
- Sezonov, G., Joseleau-Petit, D., and D'Ari, R. (2007). *Escherichia coli* physiology in Luria-Bertani broth. *J. Bacteriol.* 189, 8746–8749. doi: 10.1128/JB.01368-07
- Shi, H., Colavin, A., Bigos, M., Tropini, C., Monds, R. D., and Huang, K. C. (2017). Deep phenotypic mapping of bacterial cytoskeletal mutants reveals physiological robustness to cell size. *Curr. Biol.* 27, 3419–3429.e4. doi: 10.1016/j.cub.2017.09.065

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2023.1270665/full#supplementary-material>

- Shi, H., Hu, Y., Odermatt, P. D., Gonzalez, C. G., Zhang, L., Elias, J. E., et al. (2021). Precise regulation of the relative rates of surface area and volume synthesis in bacterial cells growing in dynamic environments. *Nat. Commun.* 12:1975. doi: 10.1038/s41467-021-22092-5
- Tittes, C., Schwarzer, S., Pfeiffer, F., Dyll-Smith, M., Rodriguez-Franco, M., Oksanen, H. M., et al. (2021). Cellular and genomic properties of *Haloferax gibbonsii* LR2-5, the host of euryarchaeal virus HFTV1. *Front. Microbiol.* 12:625599. doi: 10.3389/fmicb.2021.625599
- van Wolferen, M., Pulschen, A. A., Baum, B., Gribaldo, S., and Albers, S. V. (2022). The cell biology of archaea. *Nat. Microbiol.* 7, 1744–1755. doi: 10.1038/s41564-022-01215-8
- Walsh, J. C., Angstmann, C. N., Bisson-Filho, A. W., Garner, E. C., Duggin, I. G., and Curmi, P. M. G. (2019). Division plane placement in pleomorphic archaea is dynamically coupled to cell shape. *Mol. Microbiol.* 112, 785–799. doi: 10.1111/mmi.14316
- Wendoloski, D., Ferrer, C., and Dyll-Smith, M. L. (2001). A new simvastatin (mevinolin)-resistance marker from *Haloarcula hispanica* and a new *Haloferax volcanii* strain cured of plasmid pHV2. The GenBank accession number for the sequence reported in this paper is AF123438. *Microbiology* 147, 959–964. doi: 10.1099/00221287-147-4-959
- Westfall, C. S., and Levin, P. A. (2017). Bacterial cell size: multifactorial and multifaceted. *Annu. Rev. Microbiol.* 71, 499–517. doi: 10.1146/annurev-micro-090816-093803
- Zhu, M., Dai, X., Guo, W., Ge, Z., Yang, M., Wang, H., et al. (2017). Manipulating the bacterial cell cycle and cell size by titrating the expression of ribonucleotide reductase. *mBio* 8:e01741. doi: 10.1128/mbio.01741-17



OPEN ACCESS

EDITED BY

Michel Geovanni Santiago-Martínez,
University of Connecticut, United States

REVIEWED BY

Changyi Zhang,
University of Illinois at Urbana-Champaign,
United States
C. Martin Lawrence,
Montana State University, United States

*CORRESPONDENCE

Sonja-Verena Albers
✉ sonja.albers@biologie.uni-freiburg.de

RECEIVED 07 July 2023

ACCEPTED 11 September 2023

PUBLISHED 02 October 2023

CITATION

Bost J, Recalde A, Waßmer B, Wagner A,
Siebers B and Albers S-V (2023) Application of
the endogenous CRISPR-Cas type I-D system
for genetic engineering in the
thermoacidophilic archaeon *Sulfolobus*
acidocaldarius. *Front. Microbiol.* 14:1254891.
doi: 10.3389/fmicb.2023.1254891

COPYRIGHT

© 2023 Bost, Recalde, Waßmer, Wagner,
Siebers and Albers. This is an open-access
article distributed under the terms of the
[Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/).
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.

Application of the endogenous CRISPR-Cas type I-D system for genetic engineering in the thermoacidophilic archaeon *Sulfolobus acidocaldarius*

Jan Bost¹, Alejandra Recalde¹, Bianca Waßmer¹,
Alexander Wagner², Bettina Siebers² and Sonja-Verena Albers^{1*}

¹Molecular Biology of Archaea, Microbiology, Faculty of Biology, University of Freiburg, Freiburg, Germany, ²Molecular Enzyme Technology and Biochemistry (MEB), Environmental Microbiology and Biotechnology (EMB), Centre for Water and Environmental Research (CWE), Faculty of Chemistry, University of Duisburg-Essen, Essen, Germany

CRISPR (clustered regularly interspaced short palindromic repeats)-Cas systems are widely distributed among bacteria and archaea. In this study, we demonstrate the successful utilization of the type I-D CRISPR-Cas system for genetic engineering in the thermoacidophilic archaeon *Sulfolobus acidocaldarius*. Given its extreme growth conditions characterized by a temperature of 75°C and pH 3, an uracil auxotrophic selection system was previously established, providing a basis for our investigations. We developed a novel plasmid specifically designed for genome editing, which incorporates a mini-CRISPR array that can be induced using xylose, resulting in targeted DNA cleavage. Additionally, we integrated a gene encoding the β -galactosidase of *Saccharolobus solfataricus* into the plasmid, enabling blue-white screening and facilitating the mutant screening process. Through the introduction of donor DNA containing genomic modifications into the plasmid, we successfully generated deletion mutants and point mutations in the genome of *S. acidocaldarius*. Exploiting the PAM (protospacer adjacent motif) dependence of type I systems, we experimentally confirmed the functionality of three different PAMs (CCA, GTA, and TCA) through a self-targeting assessment assay and the gene deletion of *upsE*. Our findings elucidate the application of the endogenous Type I-D CRISPR-Cas system for genetic engineering in *S. acidocaldarius*, thus expanding its genetic toolbox.

KEYWORDS

archaea, genetic tools, deletion mutant, genetic engineering, type I-D CRISPR system, protospacer adjacent motif

1. Introduction

The CRISPR-Cas system is an RNA-guided prokaryotic defense system to protect bacterial and archaeal cells from foreign DNA, such as virus invasion or conjugative plasmids (Barrangou et al., 2007; Brouns et al., 2008; Hale et al., 2009; Garneau et al., 2010; Marraffini and Sontheimer, 2010; Westra et al., 2012; Elmore et al., 2015). This self-defense mechanism consists of various steps (McGinn and Marraffini, 2019). First, the infected cell acquires a piece of foreign DNA and incorporates it into its own genome between specific clustered regularly interspaced short palindromic repeats (CRISPR). The integrated sequences are

called spacers, which function as a memory from past survived infections (Fineran and Charpentier, 2012). Depending on the species, there are several of these clusters, which are accompanied by specific genes encoding for CRISPR-associated (Cas) proteins. CRISPR systems are divided into 2 classes, 6 types, and 33 subtypes and several variants, according to the properties of the Cas proteins (Makarova et al., 2015). Class 1 systems (type I, III, and IV) consist of a ribonucleoprotein (RNP) effector complex that is composed of several Cas proteins and bound crRNA (CRISPR RNA) during interference (Liu and Doudna, 2020). In comparison, class 2 systems (type II, V, and VI) only utilize one multidomain Cas protein, which interacts with crRNA for interference (Makarova et al., 2020; Nidhi et al., 2021). Approximately 47% of analyzed bacterial and archaeal genomes contain CRISPR systems, which, however, are much more prevalent in archaea (87%) than in bacteria (50%). Type I systems are the most dominant form of CRISPR systems, present in 64 and 60% of archaea and bacteria, respectively (Makarova et al., 2011, 2013). Type I and II systems interfere with invading DNA (Sinkunas et al., 2013), whereas type III systems, for example, interact with DNA and RNA in a transcriptional-dependent fashion (Samai et al., 2015).

In Sulfolobales, most CRISPR systems include type I-A, I-D, type III-B, and III-D (Vestergaard et al., 2014). Most research studies regarding CRISPR systems in Sulfolobales were performed in *Saccharolobus solfataricus* and *Sulfolobus islandicus*, showing the roles of different Cas proteins during CRISPR activity and the necessity of a protospacer adjacent motif (PAM) for type I systems (Peng et al., 2013), properties of protospacer and crRNA for interference (Manica et al., 2013; Mousaei et al., 2016), and the degradation properties of the type I-D system (Lin et al., 2020). For a more detailed insight into the different aspects of CRISPR-Cas systems in Sulfolobales, we refer to the reviews of the study mentioned in the reference (Garrett et al., 2011, 2015; Cannone et al., 2013; Manica and Schleper, 2013; Zhang and White, 2013; Zink et al., 2020).

After exploring various aspects of CRISPR-Cas systems in Sulfolobales, the endogenous CRISPR type I and III systems in *S. islandicus* were utilized for genome editing (Li et al., 2016). In this approach, the endogenous CRISPR-Cas system was primed using a specific spacer incorporated into a mini-CRISPR array integrated into a plasmid. By targeting a wild-type sequence, the CRISPR system induced DNA damage at the desired genomic site. Notably, the plasmid itself was not targeted due to the absence of a PAM sequence. The resulting DNA damage was subsequently repaired by cellular DNA repair mechanisms, thereby enabling the introduction of genetic modifications through homologous recombination with a provided repair fragment (Yang et al., 2020).

Furthermore, Schleper et al. demonstrated the potential of CRISPR systems in Sulfolobales by utilizing the type III system for RNA interference assays in *Sa. solfataricus*. Their studies highlighted the versatility and applicability of CRISPR systems in this context (Zebec et al., 2016; Zink et al., 2021).

S. acidocaldarius exhibits CRISPR-Cas type I-D and III-D systems (Figure 1) (Makarova et al., 2020). The type I-D system is a unique type I system, as it does not have the Cas8 scaffold protein, but the Cas10 protein, which originates from a type III system, leads to the hypothesis of an evolutionary link between

type I and III systems (Lin et al., 2020). The cluster contains the signature protein from the type I system, Cas3' (helicase domain), without the endonuclease domain (Cas3''), as well as a Cas10-like large subunit, Cas10d from the type III system, together with Csc1 (Cas5) and the backbone Csc2 (Cas7) (Makarova et al., 2011). These Cas genes are associated with CRISPR clusters with different numbers of spacer sequences throughout the genome, with the type I-D genes associated with cluster 1 and cluster 2 (Figure 1).

To distinguish the foreign DNA from inherent DNA, type I and II systems need specific motifs that can be found next to the targeting sites (protospacer), called protospacer adjacent motif (PAM) (Mojica et al., 2009; Fischer et al., 2012). These PAMs differ in length and sequence between species and are located next to the protospacer at the 5' end in type I systems and 3' end in type II systems (Gleditsch et al., 2019). These sequences can be discovered by analyzing the adjacent motifs of the protospacers of previous infections in the endogenous CRISPR clusters via bioinformatics tools. It is shown in *S. islandicus* LAL14/1 that the type I-D system is able to cleave dsDNA using the PAM GTN, similar to other type I systems, and also ssDNA by a ruler-like mechanism that is similar to type III systems without the need for a PAM (Lin et al., 2020). Previously, Lillestøl et al. proposed CCN, GTN, and TCN as PAMs for different Sulfolobales species (Lillestøl et al., 2009).

In this study, we used the endogenous type I CRISPR system of *S. acidocaldarius* to generate deletion mutants and introduce point mutations in previously characterized genes. Using this as proof of concept, we were able to expand the genetic toolbox of this thermoacidophilic model organism.

2. Materials and methods

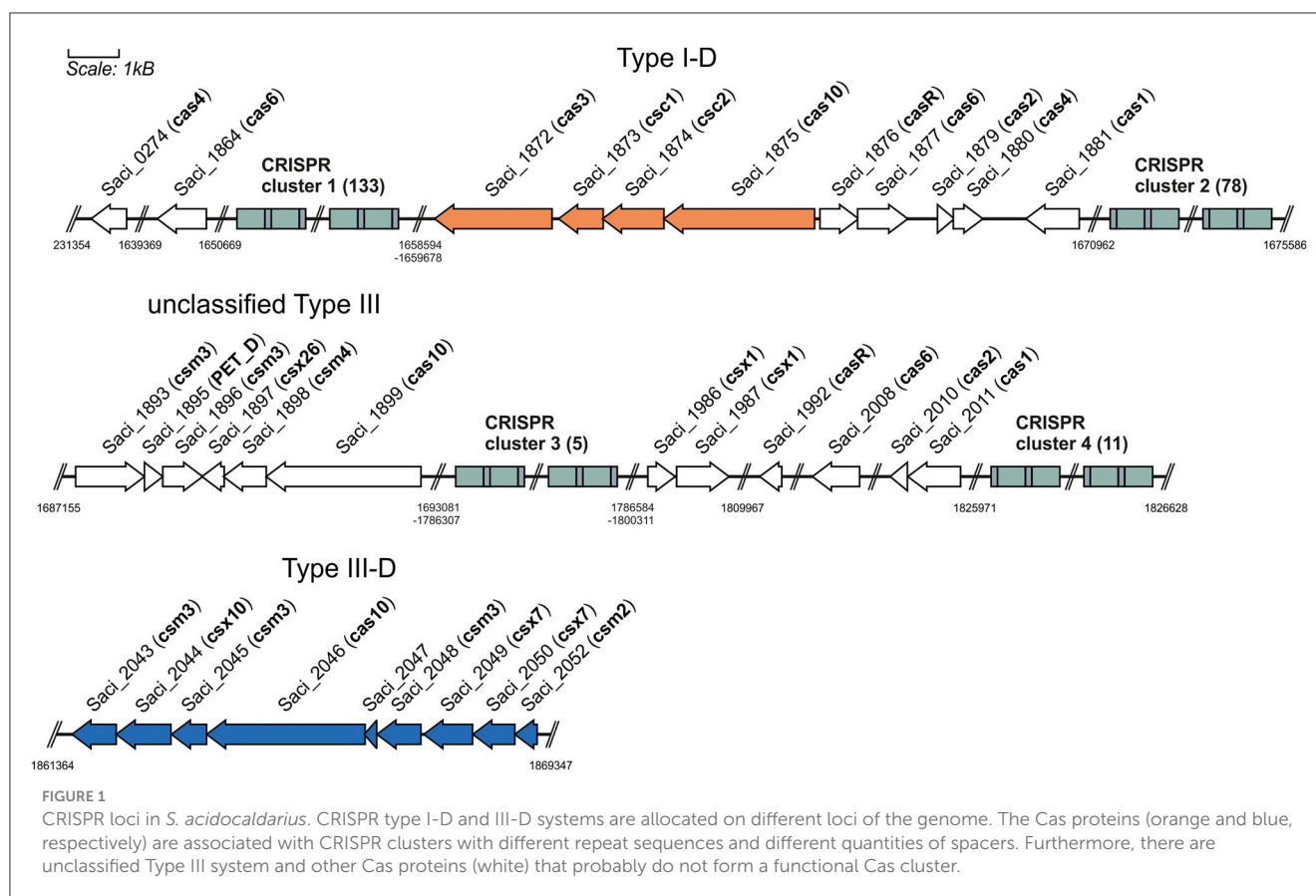
2.1. Strains and growth conditions

For all experiments, *Sulfolobus acidocaldarius* MW001 was used as the background strain and aerobically grown in Brock standard medium (Brock et al., 1972) supplemented with 0.1% (w/v) N-Z-Amine and 0.2% (w/v) dextrin at pH of 3 and 75°C. For inducing protein expression from plasmids containing a xylose-inducible promoter, dextrin was replaced with 0.2% xylose. For complementation of the uracil auxotrophy, 20 µg/ml of uracil was added to the medium.

To cultivate *S. acidocaldarius* on a plate, Brock medium that was concentrated two times supplemented with 6 mM CaCl₂, 20 mM MgCl₂, 0.2% (w/v) N-Z-Amine, and 0.4% (w/v) dextrin was mixed in equal amount with freshly boiled 1.4% (w/v) Gelrite (Carl Roth, Karlsruhe, German). Cultures on plates were incubated for 5–7 days at 75°C in humidified containers. To remove plasmids from cells, they were streaked out on the second selection plates, which contained an additional 10 µg/ml uracil and 100 µg/ml 5-FOA.

2.2. Competent *S. acidocaldarius* MW001 cells

S. acidocaldarius strain MW001 was grown in 50 ml of Brock medium supplemented with NZ-Amine, dextrin, and uracil. Upon



reaching an optical density (OD_{600}) of 0.5–0.7, a portion of the culture was transferred to 50 ml of fresh medium and harvested at an optical density OD_{600} of 0.2–0.3. The culture was cooled down on ice and then centrifuged for 15 min at $4000 \times g$ at 4°C and washed three times with 30 ml and one last time with 1 ml of ice-cold 20 mM sucrose. The cells were resuspended in 20 mM sucrose to reach a final theoretical optical density of 20 and divided into a portion of 50 μl and immediately used for transformation or stored at -80°C without freezing in liquid nitrogen.

2.3. Transformation of competent *S. acidocaldarius* MW001 cells

To prevent restriction by the *SuaI* restriction system, plasmids were methylated prior to their transformation into *S. acidocaldarius*. For this purpose, *E. coli* ER1821 cells, containing the additional plasmid pM.EsaBC4I (New England Biolabs, Frankfurt am Main, Germany), were transformed with the obtained plasmids. The methylated plasmids were, then, purified and electroporated into competent MW001 cells, using a Gene Pulser Xcell (Bio-Rad, München, Germany) with a constant time protocol at 1.5 kV, 25 μF , 600 Ω in 1-mm cuvettes. Cells were regenerated for 30 min at 75°C in 450 μl Brock medium without pH adjustment. To recover the transformants, 50 μl of cell suspension was inoculated in 5 ml of Brock medium supplemented with 0.2% (w/v) xylose and 0.1%

(w/v) dextrin and incubated for 2 days at 75°C . From these cultures, 150 μl were plated per selection plate.

2.4. Construction of plasmids

To obtain a base vector for using the endogenous CRISPR-Cas system, an expression plasmid named pSVA13134 was designed. To that end, *lacI/lacZ* was cloned into the *NcoI/NotI* cloning site of pSVAXylFX-Stop (van der Kolk et al., 2020) using primers 11670/11671 for amplifying the plasmid and 11672/11673 for amplifying *lacI/lacZ* (Table 1). The primers contained overlapping regions at each end, which consisted of the cluster 1 repeat (GTAATAACGACAAGAACTAAAAC) and *SapI* restriction site. In between the repeats was the *lacI/lacZ* cluster. Cloning was performed using *in vivo* assembly resulting in pSVAXylFX-CRISPR. The *lacS* (*ss03019*) gene of *Sa. solfataricus*, encoding a β -galactosidase, was also added using *NdeI* and *NheI* with the primers 11642/11643 resulting in pSVA13133.

Primers 12042/12049 were used to delete additional *ApaI* and *XhoI* sites and 12050/12051 to move an *ApaI* restriction site to another position, using T4 PNK cloning, to expand the usable multiple cloning site of the resulting plasmid pSVA13134.

Repair fragments for homologous recombination, containing genetic modifications, were then cloned into the *ApaI/NdeI* restriction sites using the primers, as shown in Table 1. For that, an upstream and a downstream fragment of *upsE* were amplified

TABLE 1 Primer list.

Primer	Sequence (5'-3')	Purpose
Primers for the CRISPR base vector		
11642	GACTGCTAGCCCGCGGCTAATTAATAATACTA	fwd to amplify P_{mal} <i>lacS</i> and terminator
11643	GATCCATATGCCGAATCTAATGAAAATGAGA	rev to amplify P_{mal} <i>lacS</i> and terminator
11670	CCATGGTACGTATTATCTTATCATTC	fwd to linearize pSVAxylFX-Stop
11671	GCCCGCGGCTAATTAATAATAC	rev to linearize pSVAxylFX-Stop
11672	GAATGATAAGATAATACGTACCATGGGTAATAACGACAAGAACTAAAA CTGAAGAGCGCGCCAATACG	fwd repeat cluster 1 and <i>lacI</i>
11673	GTATTATTAATAGCCCGCGGCTAGCTCGAGGTCGACGTTTGTGTTCTT GTCGTATTACTGAAGAGCGACGTCTTAATGCGC	rev repeat cluster 1 and <i>lacZ</i>
12042	AGAAAGTGGTCCCTTACTCTAGTGCCTGTC	fwd to remove <i>XhoI/ApaI</i>
12049	CAAGTCTCACTATACCAATGAG	rev to remove <i>XhoI/ApaI</i>
12050	AAATCTACCGTTGTCAATTTTA	fwd to introduce <i>ApaI</i>
12051	TTCAGTAGGGCCCGTGTGAAAGCGGCCG	rev to introduce <i>ApaI</i>
Primers for US/DS repair fragments		
12904	GTACATCCATATGAACATTTACGAGAATATTTATTACGCTAAGG	fwd US/DS for Δ <i>upsE</i>
12905	AGAATGGGCCCTTAATCTATCTTAAGCGAAACG	rev US/DS for Δ <i>upsE</i>
12918	GTACATCCATATGCGAGATTACTCCGTTATGTTAG	fwd US/DS for <i>upsE</i> Walker A motif change
12919	AGAATGGGCCAGTTTCACTCCACATCTAC	rev US/DS for <i>upsE</i> Walker A motif change
12922	ATTGGGTCCAACGGGATCTGGAGCTACTACATTATTAAAC	fwd overlap US/DS <i>upsE</i> Walker A motif change K232A
12923	AAAGCGTTTAATAATGTAGTAGTCCAGATCCCGTTGGAC	rev overlap US/DS <i>upsE</i> Walker A motif change K232A
Primers for self-targeting assay		
13606	TATGCTCTTCAAACAGAAAATATCTCAAGGAGGGCGAGGAAGTAT GCGAAAGGTAAGAAGAGCAAT	protospacer cluster 1 spacer 1 ctr w/o PAM
13607	ATTGCTCTTCTTACCTTTTCGCATACTTCCTCGCCCTCCTTGAGAT ATTTTCTGTTTGAAGAGCATA	protospacer cluster 1 spacer 1 ctr w/o PAM
13608	TATGCTCTTCAAACcCaAGAAAATATCTCAAGGAGGGCGAGGAAG TATGCGAAAGGTAAGAAGAGCAAT	protospacer cluster 1 spacer 1 with CCA PAM
13609	ATTGCTCTTCTTACCTTTTCGCATACTTCCTCGCCCTCCTTGAGAT ATTTTCTTggGTTTGAAGAGCATA	protospacer cluster 1 spacer 1 with CCA PAM
13610	TATGCTCTTCAAACtCaAGAAAATATCTCAAGGAGGGCGAGGAAGT ATGCGAAAGGTAAGAAGAGCAAT	protospacer cluster 1 spacer 1 with TCA PAM
13611	ATTGCTCTTCTTACCTTTTCGCATACTTCCTCGCCCTCCTTGAGAT ATTTTCTTgaGTTTGAAGAGCATA	protospacer cluster 1 spacer 1 with TCA PAM
13612	TATGCTCTTCAAACgtaAGAAAATATCTCAAGGAGGGCGAGGAAGTA TGCGAAAGGTAAGAAGAGCAAT	protospacer cluster 1 spacer 1 with GTA PAM
13613	ATTGCTCTTCTTACCTTTTCGCATACTTCCTCGCCCTCCTTGAGATATTT TCTtacGTTTGAAGAGCATA	protospacer cluster 1 spacer 1 with GTA PAM
Primers for genetic manipulation		
11554	TATGCTCTTCAAACCTTAAACCTCTGAACATTCTGGAAGTTATCAAT TCCTGTAAGAAGAGCAAT	fwd spacer targeting <i>upsE</i> , CCA PAM
11555	ATTGCTCTTCTTACAGGAATTGATAACTTCCAGAATGTTTACAGAGGT TTTAAAGTTTGAAGAGCATA	rev spacer targeting <i>upsE</i> , CCA PAM
12900	TATGCTCTTCAAACACGGGATCTGGAAAACTACATTATTAAACGC TTTACGTAAGAAGAGCAAT	fwd spacer targeting <i>upsE</i> Walker A motif change K232A, CCA PAM
12901	ATTGCTCTTCTTACGTAAAGCGTTTAATAATGTAGTTTTCAGAT CCCGTGTGTTGAAGAGCATA	rev spacer targeting <i>upsE</i> Walker A motif change K232A, CCA PAM
12912	TATGCTCTTCAAACATCCCGGTAAAGAGATTCTTTAGATATAGT CGCTGCGTAAGAAGAGCAAT	fwd spacer targeting <i>upsE</i> , GTA PAM

(Continued)

TABLE 1 (Continued)

Primer	Sequence (5'-3')	Purpose
12913	ATTGCTCTTCTTACGCAGCGACTATATCTAAAGAAATCTCTTTA CCGGGATGTTTGAAGAGCATA	rev spacer targeting <i>upsE</i> , GTA PAM
12914	TATGCTCTTCAAACCTTGCCCGAGGGTCATAGGGTAGCAGCGACTAT ATCTAGTAAGAAGAGCAAT	fwd spacer targeting <i>upsE</i> , TCA PAM
12915	ATTGCTCTTCTTACTAGATATAGTCGCTGCTACCCTATGACCCTC GGGCAAGTTTGAAGAGCATA	rev spacer targeting <i>upsE</i> , TCA PAM

from genomic *S. acidocaldarius* DNA, and both PCR products ligated, using overlap extension PCR, resulting in pSVA13271 (for *upsE* KO) and pSVA13280 (Walker A mutation K232A) (Table 2). Spacers for targeted CRISPR activity were generated by annealing the forward and reverse primers at 98°C for 10 s followed by 50°C for 10 s. The primer contained *SapI* restriction sites, parts of the cluster 1 repeat, and target sequence for the CRISPR system.

For the self-assessment protocol, the protospacer of the first spacer of CRISPR cluster 1 (AGAAAATATCTCAAGGAGGGCGAGGAAGTATGCGAAAG) was cloned into the *SapI* restriction site using FX cloning with primers 13606/13607 for the non-PAM control and primers 13608-13613 for the different tested PAMs CCA, GTA, and TCA, respectively (Table 1).

2.5. Blue-white screening and colony PCR for genotype analysis

Potential candidates were first selected by blue-white screening, spraying X-Gal (25mg/ml in DMF) diluted in a 1:4 ratio with dextrin [20% (w/v)] on visible colonies. Afterward, the plates were incubated for up to 4 h until colonies turned blue. To verify the genotype of potential mutants, blue colonies were lysed for 2 min in 10 μ l 0.2 M NaOH. To prevent DNA denaturation, 60 μ l of 0.2 M Tris (pH 7.8) was added, as well as 60 μ l of ddH₂O added to dilute the sample. After vortexing, 1 μ l was used as a template for a 20 μ l PCR reaction using the Phusion polymerase. MW001 DNA was used as a wild-type control. For analysis of deletion mutants, plasmid DNA was used as a negative control to ensure that the signal was due to genetic alteration, not from plasmid amplification. After the analysis of gel electrophoresis, PCR products were sequenced (Eurofins Genomics Europe).

2.6. Ultraviolet aggregation assay

UV treatment was carried out following the protocol described by Fröls et al. (2008). In total, 10 ml of culture with an optical density (OD) of 0.2 to 0.3 was exposed to 75 J/m² of UV light at 254 nm using a UV Crosslinker device (Spectroline, Westbury, NY). The cultures were, then, incubated at 75°C for 3 h. To determine the number of aggregated cells after UV exposure, the cell culture was diluted to OD 0.2, and 5 μ l spotted onto

a microscope slide coated with a thin layer of 1% (w/v) agarose in Brock minimal medium. After drying the cell suspension, a coverslip was added, and pictures were taken in three fields per sample under a phase contrast microscope. The number of free and aggregated cells (≥ 3) was counted using the ImageJ cell counter (NIH, Bethesda, MD).

3. Results and discussion

3.1. The CRISPR-Cas base vector for genetic manipulations

Wagner et al. established a genetic toolbox for *S. acidocaldarius* in 2012 based on a uracil auxotrophic strain MW001 in combination with a number of plasmids usable for the construction of deletion mutants, or mutants in which genes were either mutated genomically or tags were added to the gene of interest (Wagner et al., 2012). Moreover, using this system, genes were placed ectopically into the genome for genetic modifications (Wagner et al., 2012). To this end, plasmids are integrated into the genomic DNA after transformation *via* homologous recombination and can be excised after specific selection using 5-FOA and uracil, leading to alteration of the genomic DNA.

For the usage of the CRISPR-Cas system for genetic engineering, we designed plasmids based on the expression vector pSVAXylFX-CRISPR, which replicates in *S. acidocaldarius* does not integrate into the genome and is based on the plasmid pRN1 (Berkner et al., 2007). It derives from pSVAXylFX-Stop (van der Kolk et al., 2020) and contains 2 CRISPR repeats of cluster 1 (GTAATAACGACAAGAACTAAAAC), which are downstream of a D-xylose-inducible promoter P_{xyl}/P_{saci_1938} . Additionally, *lacS_{Sso}* from *Sa. solfataricus* was integrated into pSVAXylFX-CRISPR, to allow for blue-white screening in *S. acidocaldarius* using X-gal and verify successful transformation. The final vector pSVA13134, which was the base for all plasmids used for genetic manipulations (Figure 2A), also contains a multiple cloning site that is suitable for inserting the repair fragment. The spacer/target sequence can be ordered as a primer pair and cloned into pSVA13134 by restriction with *SapI* (Figure 2B). Spacer primers are designed by searching for a 37 nt protospacer sequence in the target area, which needs to be flanked by a PAM at the 5' end (Figure 2C). Selection of positive *E. coli* clones is accomplished with blue-white screening because of the presence of the *lacI/lacZ* cassette in between the *SapI* restriction sites (Figure 2A) (Geertsma, 2013).

Brock medium were tested, containing different carbon sources (D-xylose, sucrose, and dextrin). As the plasmid mini-CRISPR array is under the control of a D-xylose-inducible promoter $P_{xyl}/P_{sac_i_{1938}}$, induction of the CRISPR array is tested on plates, as well as in liquid medium for the 1–3 days of recovery step. Ultimately, positive genetically modified colonies only formed after 2 days of induction in Brock-NZ-Amine-D-xylose and plating on the first selection plates (Brock-NZ-Amine-dextrin) (Figure 3). No other combination yielded any positive colonies on plates.

In general, transformation in *S. acidocaldarius* always yields a certain amount of false positive colonies on the first selection plates, probably due to uracil cross-feeding from lysed cells. Therefore, transformants were diluted in a 1:100 ratio after transformation (50 μ l transformants in 5 ml inducing medium for 2 days to prevent a high amount of background colonies, resulting in a ratio, that allowed for consistent colony formation).

3.3. The endogenous CRISPR-Cas type I-D system targets plasmid in self-targeting assessment assay

For *S. acidocaldarius*, *Sa. solfataricus*, and *S. islandicus*, several PAM sequences have been predicted bioinformatically by analyzing the targets of the protospacer sequences of CRISPR arrays from different Sulfolobales (Lillestøl et al., 2009). Three commonly found motifs were 5'-CCN-3', 5'-GTN-3', and 5'-TCN-3'. To verify the activity of the endogenous CRISPR type I-D system in *S. acidocaldarius*, a plasmid self-targeting assessment was performed. To that end, a plasmid containing a target sequence of the endogenous CRISPR-Cas system was transformed into *S. acidocaldarius* MW001. In the case of a functional CRISPR-Cas system and PAM sequence, the CRISPR-Cas RNP complex is able to cleave the plasmid, which also harbors a selection cassette, leading to the formation of fewer colonies compared with a non-target control (Figure 4A). Therefore, the protospacer sequence corresponding to the last acquired spacer of the CRISPR cluster 1 (AGAAAATATCTCAAGGAGGGCGAGGAAGTATGCGAAAG) was used as a target (Figure 4B). The sequence was flanked at the 5'-end by the PAMs CCA, GTA, or TCA in the CRISPR expression cassette, which was designed to mimic the sequence and arrangement of the native type I-D array in *S. acidocaldarius*. As a negative control, no PAM was inserted at the 5'-end of the target, which was just flanked by the native cluster 1 repeat sequence (AAC) (Figure 4A).

We showed that the presence of the target sequence for the first spacer of cluster 1 totally abolished the presence of the plasmid when using TCA as PAM. After blue-white screening to verify the presence of the plasmid, there was nearly full clearance for all tested PAMs with ~1 colony per 200 ng plasmid DNA for CCA and GTA PAMs (Figure 4C). The no PAM control yielded, on average, 303 colonies, showing a clearance effect for the used PAMs.

The self-target assessment indicated a functional CRISPR-type I-D system in *S. acidocaldarius*. The expression of endogenous Cas proteins under the native promoter system was sufficient

to generate full clearance of transformed plasmids. Through this method, PAMs can easily be tested, as demonstrated previously, e.g., in *Pyrococcus furiosus* (Elmore et al., 2015). Notably, Lillestøl predicted CCN as a functional PAM for *Sa. solfataricus* and *S. islandicus* as no valid PAM was predicted for *Sulfolobus acidocaldarius* (Lillestøl et al., 2009). Our results demonstrated that the three PAMs published by Lillestøl et al. have a similar impact on the clearing of targeted DNA.

3.4. The endogenous CRISPR-Cas type I-D system can be used for genetic engineering

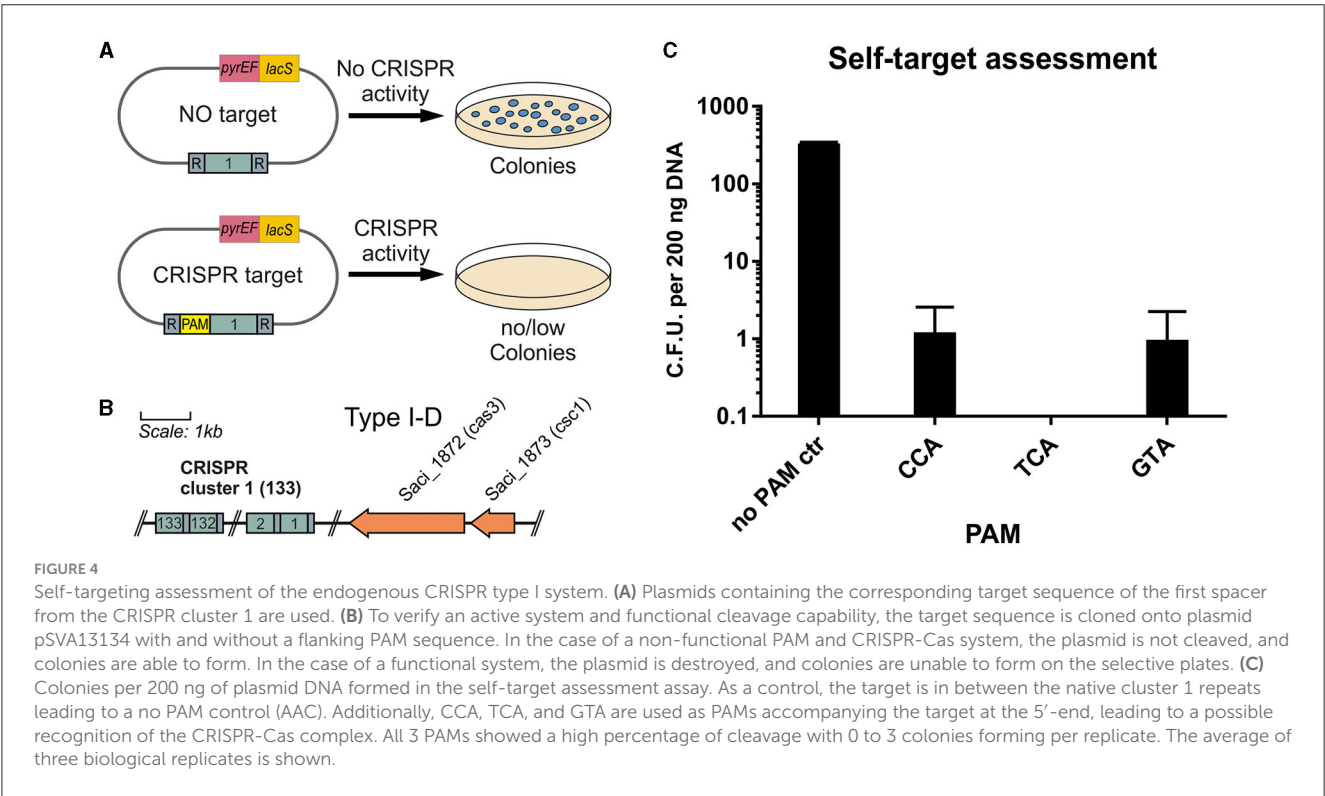
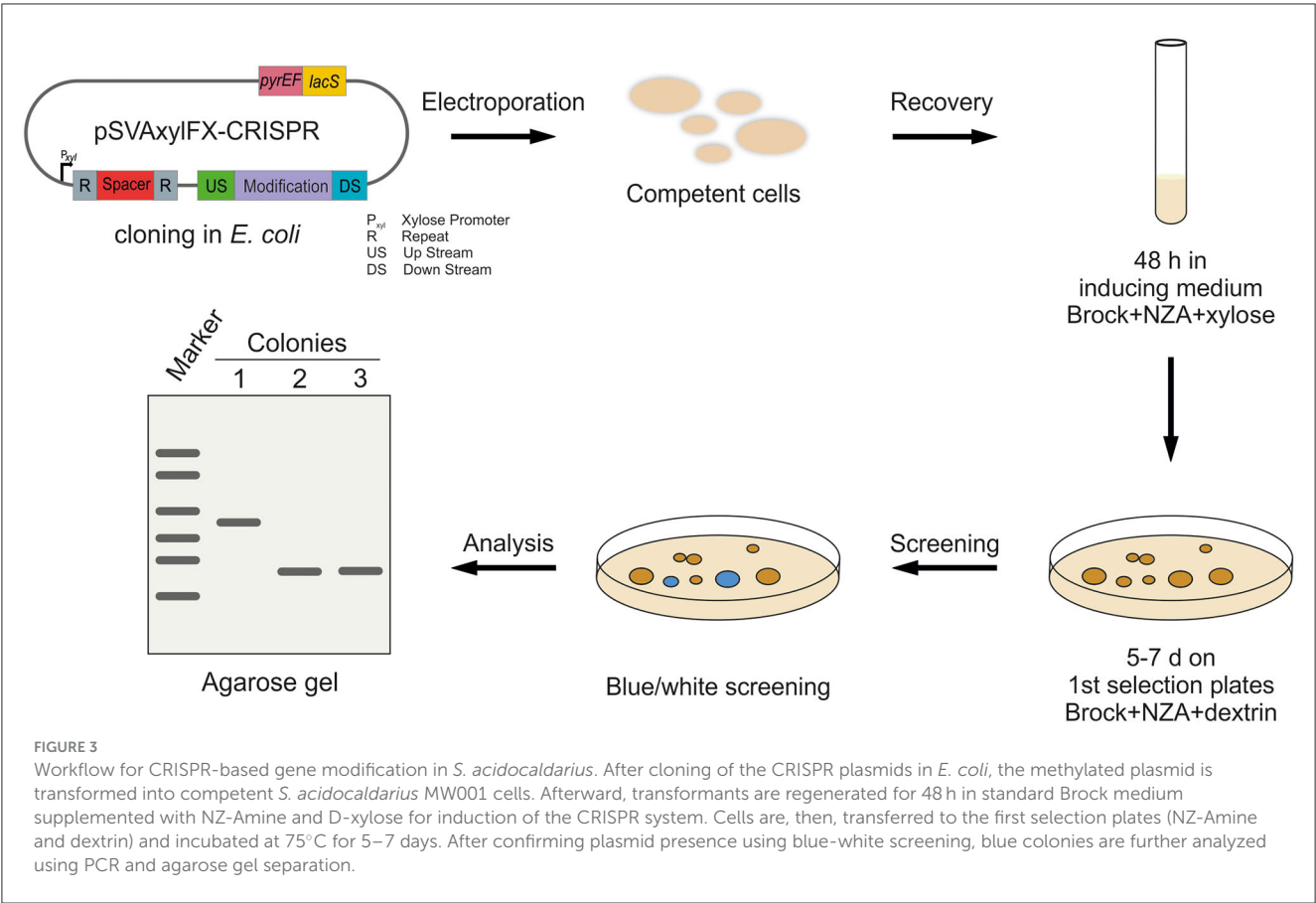
3.4.1. Deletion of *upsE*-proof viability of CRISPR-Cas as a genetic tool

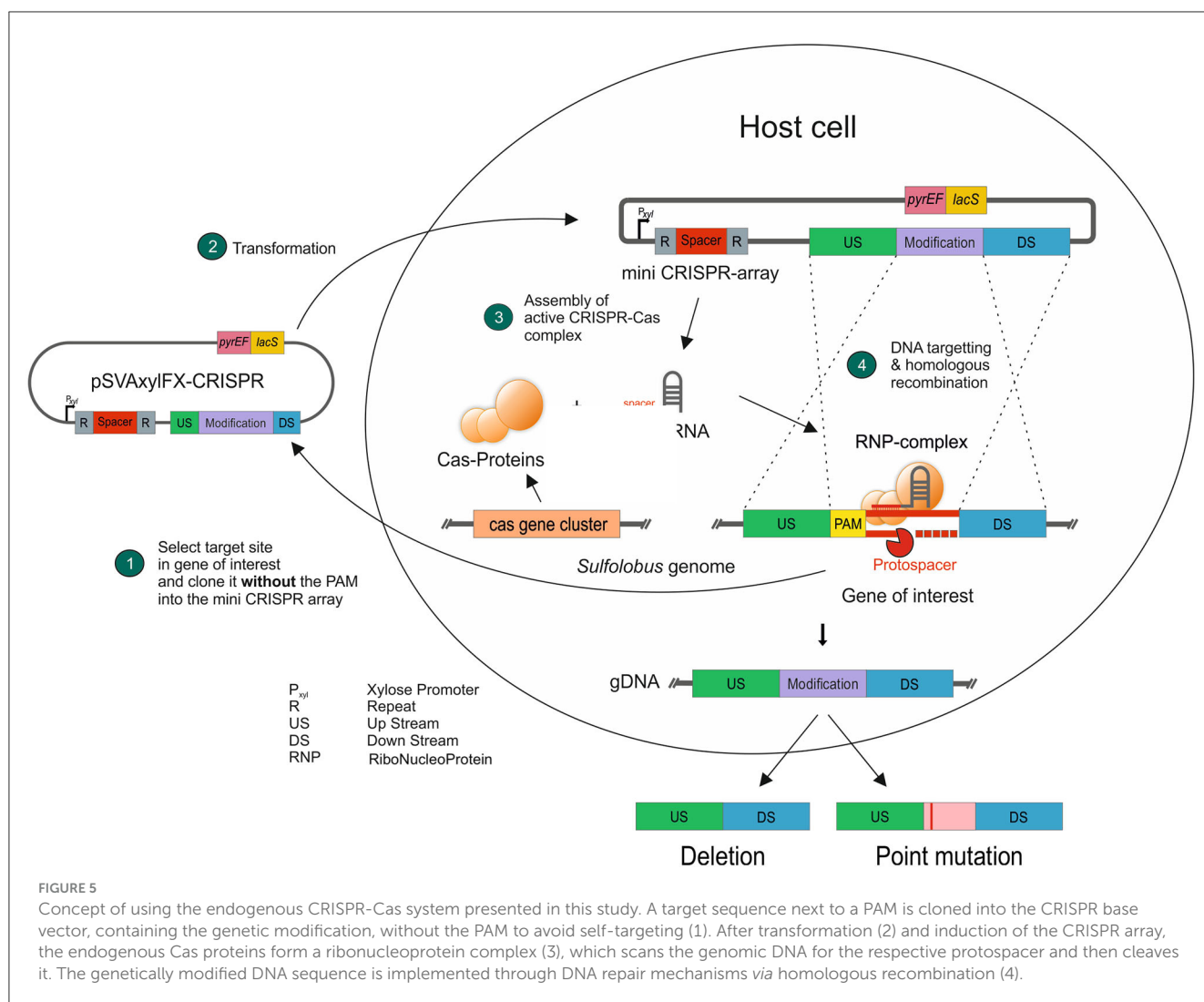
After showing the activity of the CRISPR system and functional PAMs through the self-targeting assessment, we used *upsE* (*saci_1494*) as a target to test whether we can obtain a gene deletion using the system. *UpsE* encodes for the UV pili assembly ATPase (Fröls et al., 2008; Ajon et al., 2011; Wagner et al., 2012; van Wolferen et al., 2013). We chose it as a target as the successful deletion can be additionally verified via UV aggregation assays (Fröls et al., 2008). Therefore, we screened for PAM sequences in the sequence of *upsE* on the sense strand and used 37 nt downstream of it as a protospacer (Figure 5). The spacer sequence was cloned in between the CRISPR repeats of cluster 1 in the artificial CRISPR array on pSVA13271. This plasmid harbors a repair fragment for homologous recombination, which consists of the 500 bp upstream and 500 bp downstream of *upsE* and is derived from pSVA13134. We tested three different PAMs CCA, GTA, and TCA using the plasmids pSVA13272, pSVA13273, and pSVA13274, respectively. For transformation, 200 ng of plasmid was transformed, and the cells were incubated for 48 h in Brock medium with NZ-Amine and 0.2% (w/v) D-xylose. D-xylose induces the transcription of the CRISPR array on the plasmid, due to a D-xylose-inducible promoter, leading to the production of crRNA (CRISPR RNA), which then forms an RNP complex with the endogenous CRISPR-Cas proteins (Figure 5). After the formation of colonies on the first selection plates, initial screening of the presence of the plasmid was performed using X-Gal (5-bromo-4-chloro-indolyl- β -D-galactopyranoside) as the gene for the β galactosidase (*lacS*) is encoded on the plasmid. PCR analysis and subsequent sequencing confirmed that all blue colonies were indeed clones in which the *upsE* gene was deleted. There was no difference in either of the used PAMs CCA, GTA, and TCA (Figure 6A).

We were also able to cure the cells of the plasmid after verifying the genotype by putting the cells on the second selection plates containing 5-FOA, which was metabolized to a cytotoxic compound by the *pyrEF* gene, forcing the plasmid out of the cell (Wagner et al., 2012).

3.4.2. Alteration of walker a motif of *upsE*

To further expand other possibilities for CRISPR-based genome editing, we tried to introduce a point mutation in the Walker A motif of *upsE*. In the case of successful genome alteration, the





newly generated mutant should not be able to form aggregates upon UV induction similar to the deletion mutant. Therefore, we wanted to mutate a lysine residue at position 232 to an alanine (K232A) (Figure 6B), to abolish the ATPase function (delToro et al., 2016). For this, the target protospacer sequence needs to be spanning over the mutational site so that the crRNA only hybridizes with the WT sequence while abolishing targeting the mutation (Figure 6B). For *upsE*, a PAM CCA is localized 12 nt upstream of the target site, which is used. The modification was, then, put into a 1 kb repair fragment into the MCS of pSVA13134, containing the previously described point mutation. After following our established protocol, we were able to generate a Walker A mutant after the first transformation, named *S. acidocaldarius* MW1304, by only screening five clones, showing very high efficiency.

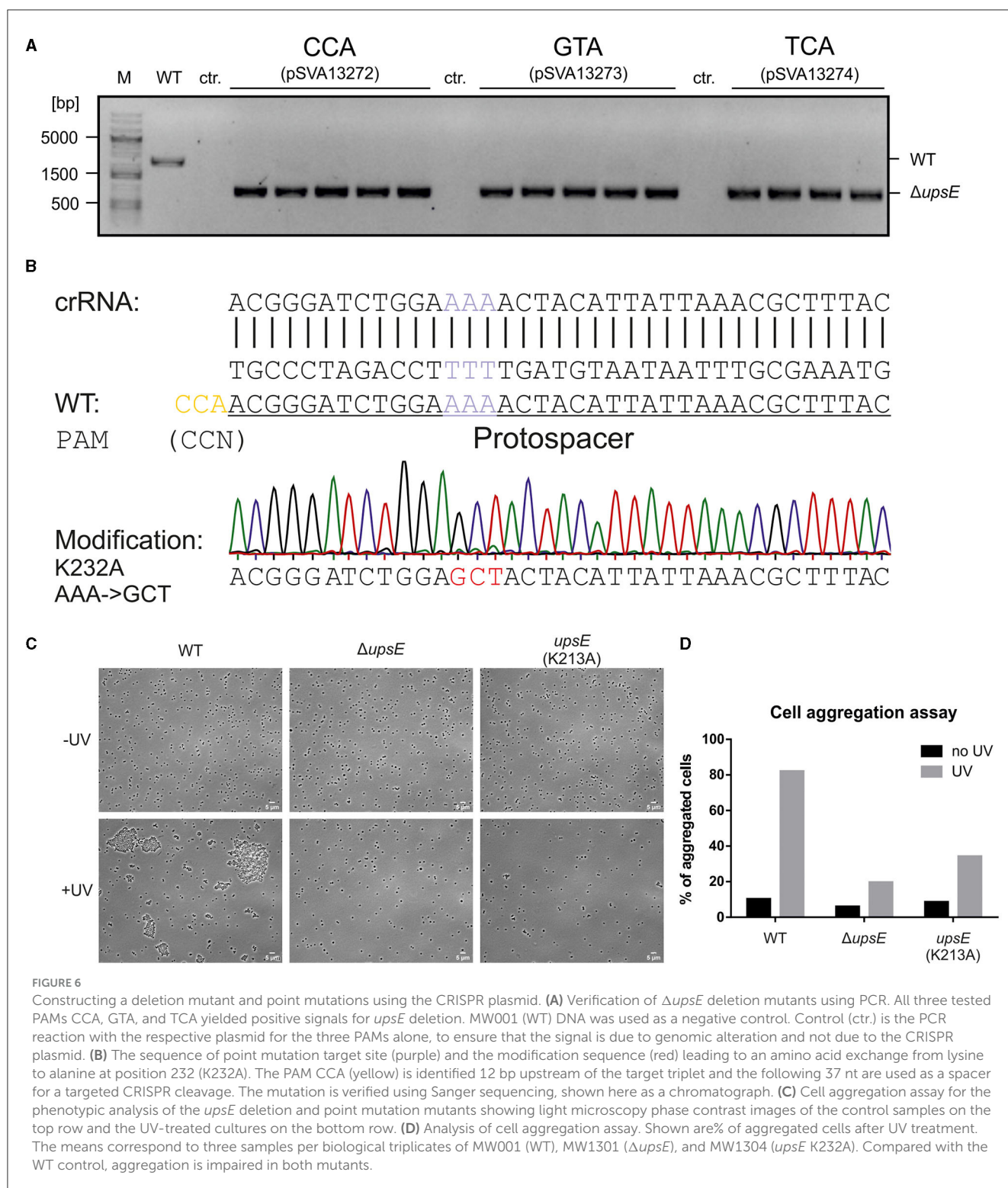
To verify the genetic edition, a UV aggregation assay was performed, showing impaired aggregation for both $\Delta upsE$ MW1301 and the Walker A mutant MW1304 (Figure 6C), as previously described (Fröls et al., 2008; Ajon et al., 2011; Wagner et al., 2012; van Wolferen et al., 2013). In contrast to the deletion mutant, there is still some aggregation of the Walker A mutant,

but much less compared with the MW001 control, showing that the obtained mutants behave as expected (Figure 6D).

4. Conclusion

In this study, we have successfully demonstrated the utility of the endogenous type I CRISPR system in *S. acidocaldarius* as a versatile genetic tool for generating gene deletion mutants and introducing single codon changes within the *S. acidocaldarius* genome, similar to the previous findings in *S. islandicus* (Li et al., 2016). By testing various potential protospacer adjacent motifs (PAMs), we have expanded the range of available PAM sites, offering more options for targeted genetic modifications.

Although the cloning process for our method involves an additional step compared with the well-established “pop-in/pop-out” approach (Wagner et al., 2012), the identification of desired mutants can be confirmed on the first selection plate. Consequently, the acquisition of mutants in *S. acidocaldarius* is significantly accelerated compared with the previous methods. This expedited process is attributed not only to the increased



yield of mutants, thereby reducing the screening period, but also to the fact that mutant screening is accomplished during the initial selection plates, ~10 days post-transformation. In contrast, the previous method necessitated screening after the second selection, which occurred ~16 days after transformation. It is worth noting that the second selection is still required to eliminate the plasmid from the cells, but this

step is performed after confirming the genotype of the colony of interest.

These advancements in the application of the endogenous type I-D CRISPR system in *S. acidocaldarius* offer a significant improvement in both efficiency and speed, thus facilitating genetic manipulations and expanding the genetic engineering capabilities in this organism.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

JB: Investigation, Methodology, Project administration, Validation, Visualization, Writing—original draft, Writing—review and editing. AR: Investigation, Methodology, Project administration, Supervision, Writing—review and editing. BW: Investigation, Writing—review and editing. AW: Conceptualization, Investigation, Writing—review and editing. BS: Funding acquisition, Resources, Supervision, Writing—review and editing. S-VA: Conceptualization, Funding acquisition, Project administration, Resources, Supervision, Writing—review and editing.

Funding

JB and AR were supported by the BMBF 031B0848C. AW was supported by the BMBF 031L0078A and 031B0848A. S-VA and

BS acknowledge the funding of the HotAcidFACTORY *Sulfolobus acidocaldarius* as novel thermoacidophilic bio-factory project within the BMBF funding initiative Mikrobielle Biofabriken für die industrielle Bioökonomie—Neuartige Plattformorganismen für innovative Produkte und nachhaltige Bioprozesse. We acknowledge support by the Open Access Publication Fund of the University of Freiburg.

Conflict of interest

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.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Ajon, M., Fröls, S., van Wolferen, M., Stoecker, K., Teichmann, D., Driessen, A. J. M., et al. (2011). UV-inducible DNA exchange in hyperthermophilic archaea mediated by type IV pili. *Mol. Microbiol.* 82, 807–817. doi: 10.1111/j.1365-2958.2011.07861.x
- Albers, S.-V., and Driessen, A. J. M. (2008). Conditions for gene disruption by homologous recombination of exogenous DNA into the *Sulfolobus solfataricus* genome. *Archaea* 2, 145–149. doi: 10.1155/2008/948014
- Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., et al. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. *Science* 315, 1709–1712. doi: 10.1126/science.1138140
- Berkner, S., Grogan, D., Albers, S.-V., and Lipps, G. (2007). Small multicopy, non-integrative shuttle vectors based on the plasmid pRN1 for *Sulfolobus acidocaldarius* and *Sulfolobus solfataricus*, model organisms of the (cren-) archaea. *Nucleic Acids Res.* 35, e88. doi: 10.1093/nar/gkm449
- Brock, T. D., Brock, K. M., Belly, R. T., and Weiss, R. L. (1972). *Sulfolobus*: a new genus of sulfur-oxidizing bacteria living at low pH and high temperature. *Archiv. Mikrobiol.* 84, 54–68.
- Brouns, S. J. J., Jore, M. M., Lundgren, M., Westra, E. R., Slijkhuys, R. J. H., Snijders, A. P. L., et al. (2008). Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* 321, 960–964. doi: 10.1126/science.1159689
- Cannone, G., Webber-Birungi, M., and Spagnolo, L. (2013). Electron microscopy studies of Type III CRISPR machines in *Sulfolobus solfataricus*. *Biochem. Soc. Transact.* 41, 1427–1430. doi: 10.1042/BST20130166
- delToro, D., Ortiz, D., Ordyan, M., Sipky, J., Oh, C.-S., Keller, N., et al. (2016). Walker-A motif acts to coordinate ATP Hydrolysis with motor output in viral DNA packaging. *J. Mol. Biol.* 428, 2709–2729. doi: 10.1016/j.jmb.2016.04.029
- Elmore, J., Deighan, T., Westpheling, J., Terns, R. M., and Terns, M. P. (2015). DNA targeting by the type I-G and type I-A CRISPR-Cas systems of *pyrococcus furiosus*. *Nucleic Acids Res.* 43, 10353–10363. doi: 10.1093/nar/gkv1140
- Fineran, P. C., and Charpentier, E. (2012). Memory of viral infections by CRISPR-Cas adaptive immune systems: acquisition of new information. *Virology* 434, 202–209. doi: 10.1016/j.virol.2012.10.003
- Fischer, S., Maier, L.-K., Stoll, B., Brendel, J., Fischer, E., Pfeiffer, F., et al. (2012). An archaeal immune system can detect multiple protospacer adjacent motifs (PAMs) to target invader DNA. *J. Biologic. Chem.* 287, 33351–33363. doi: 10.1074/jbc.M112.377002
- Fröls, S., Ajon, M., Wagner, M., Teichmann, D., Zolghadr, B., Folea, M., et al. (2008). UV-inducible cellular aggregation of the hyperthermophilic archaeon *Sulfolobus solfataricus* is mediated by pili formation. *Mol. Microbiol.* 70, 938–952. doi: 10.1111/j.1365-2958.2008.06459.x
- Garneau, J. E., Dupuis, M.-È., Villion, M., Romero, D. A., Barrangou, R., Boyaval, P., et al. (2010). The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* 468, 67–71. doi: 10.1038/nature09523
- Garrett, R. A., Shah, S. A., Erdmann, S., Liu, G., Mousaei, M., León-Sobrinho, C., et al. (2015). CRISPR-cas adaptive immune systems of the *Sulfolobales*: unravelling their complexity and diversity. *Life* 5, 783–817. doi: 10.3390/life5010783
- Garrett, R. A., Shah, S. A., Vestergaard, G., Deng, L., Gudbergdottir, S., Kenchappa, C. S., et al. (2011). CRISPR-based immune systems of the *Sulfolobales*: complexity and diversity. *Biochem. Soc. Transact.* 39, 51–57. doi: 10.1042/BST0390051
- Geertsma, E. R. (2013). FX cloning: a versatile high-throughput cloning system for characterization of enzyme variants. *Methods Mol. Biol.* 978, 133–148. doi: 10.1007/978-1-62703-293-3_10
- Gleditsch, D., Pausch, P., Müller-Esparza, H., Özcan, A., Guo, X., Bange, G., et al. (2019). PAM identification by CRISPR-Cas effector complexes: diversified mechanisms and structures. *RNA Biol.* 16, 504–517. doi: 10.1080/15476286.2018.1504546
- Hale, C. R., Zhao, P., Olson, S., Duff, M. O., Graveley, B. R., Wells, L., et al. (2009). RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. *Cell* 139:945–956. doi: 10.1016/j.cell.2009.07.040
- Li, Y., Pan, S., Zhang, Y., Ren, M., Feng, M., Peng, N., et al. (2016). Harnessing Type I and Type III CRISPR-Cas systems for genome editing. *Nucleic Acids Res.* 44, e34. doi: 10.1093/nar/gkv1044
- Lillestøl, R. K., Shah, S. A., Brügger, K., Redder, P., Phan, H., Christiansen, J., et al. (2009). CRISPR families of the crenarchaeal genus *Sulfolobus*: bidirectional transcription and dynamic properties. *Mol. Microbiol.* 72, 259–272. doi: 10.1111/j.1365-2958.2009.06641.x
- Lin, J., Fuglsang, A., Kjeldsen, A. L., Sun, K., Bhoobalan-Chitty, Y., and Peng, X. (2020). DNA targeting by subtype I-D CRISPR-Cas shows type I and type III features. *Nucleic Acids Res.* 48, 10470–10478. doi: 10.1093/nar/gkaa749

- Liu, T. Y., and Doudna, J. A. (2020). Chemistry of Class 1 CRISPR-Cas effectors: binding, editing, and regulation. *J. Biologic. Chemistr.* 295, 14473–14487. doi: 10.1074/jbc.REV120.007034
- Makarova, K. S., Haft, D. H., Barrangou, R., Brouns, S. J. J., Charpentier, E., Horvath, P., et al. (2011). Evolution and classification of the CRISPR-Cas systems. *Nat. Rev. Microbiol.* 9, 467–477. doi: 10.1038/nrmicro2577
- Makarova, K. S., Wolf, Y. I., Alkhnbashi, O. S., Costa, F., Shah, S. A., Saunders, S. J., et al. (2015). An updated evolutionary classification of CRISPR-Cas systems. *Nat. Rev. Microbiol.* 13, 722–736. doi: 10.1038/nrmicro3569
- Makarova, K. S., Wolf, Y. I., Iranzo, J., Shmakov, S. A., Alkhnbashi, O. S., Brouns, S. J. J., et al. (2020). Evolutionary classification of CRISPR-Cas systems: a burst of class 2 and derived variants. *Nat. Rev. Microbiology* 18, 67–83. doi: 10.1038/s41579-019-0299-x
- Makarova, K. S., Wolf, Y. I., and Koonin, E. V. (2013). Comparative genomics of defense systems in archaea and bacteria. *Nucleic Acids Res.* 41, 4360–4377. doi: 10.1093/nar/gkt157
- Manica, A., and Schleper, C. (2013). CRISPR-mediated defense mechanisms in the hyperthermophilic archaeal genus *Sulfolobus*. *RNA biology* 10, 671–678. doi: 10.4161/rna.24154
- Manica, A., Zebec, Z., Steinkellner, J., and Schleper, C. (2013). Unexpectedly broad target recognition of the CRISPR-mediated virus defence system in the archaeon *Sulfolobus solfataricus*. *Nucleic Acids Res.* 41, 10509–10517. doi: 10.1093/nar/gkt767
- Marraffini, L. A., and Sontheimer, E. J. (2010). CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. *Nat. Rev. Genetics* 11, 181–190. doi: 10.1038/nrg2749
- McGinn, J., and Marraffini, L. A. (2019). Molecular mechanisms of CRISPR-Cas spacer acquisition. *Nat. Rev. Microbiology* 17, 7–12. doi: 10.1038/s41579-018-0071-7
- Mojica, F. J. M., Díez-Villaseñor, C., García-Martínez, J., and Almendros, C. (2009). Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* 155, 733–740. doi: 10.1099/mic.0.023960-0
- Mousaei, M., Deng, L., She, Q., and Garrett, R. A. (2016). Major and minor crRNA annealing sites facilitate low stringency DNA protospacer binding prior to Type I-A CRISPR-Cas interference in *sulfolobus*. *RNA Biol.* 13, 1166–1173. doi: 10.1080/15476286.2016.1229735
- Nidhi, S., Anand, U., Oleksak, P., Tripathi, P., Lal, J. A., Thomas, G., et al. (2021). Novel CRISPR-Cas systems: an updated review of the current achievements, applications, and future research perspectives. *Int. J. Mol. Sci.* 22, 7327. doi: 10.3390/ijms22073327
- Peng, W., Li, H., Hallström, S., Peng, N., Liang, Y. X., and She, Q. (2013). Genetic determinants of PAM-dependent DNA targeting and pre-crRNA processing in *Sulfolobus islandicus*. *RNA Biol.* 10, 738–748. doi: 10.4161/rna.23798
- Samai, P., Pyenson, N., Jiang, W., Goldberg, G. W., Hatoum-Aslan, A., and Marraffini, L. A. (2015). Co-transcriptional DNA and RNA cleavage during Type III CRISPR-Cas immunity. *Cell* 161, 1164–1174. doi: 10.1016/j.cell.2015.04.027
- Sinkunas, T., Gasiunas, G., Waghmare, S. P., Dickman, M. J., Barrangou, R., Horvath, P., et al. (2013). In vitro reconstitution of Cascade-mediated CRISPR immunity in *Streptococcus thermophilus*. *EMBO J.* 32, 385–394. doi: 10.1038/emboj.2012.352
- van der Kolk, N., Wagner, A., Wagner, M., Waßmer, B., Siebers, B., and Albers, S.-V. (2020). Identification of XylR, the activator of arabinose/xylose inducible regulon in *sulfolobus acidocaldarius* and its application for homologous protein expression. *Front. Microbiol.* 11, 1066. doi: 10.3389/fmicb.2020.01066
- van Wolferen, M., Ajon, M., Driessen, A. J. M., and Albers, S.-V. (2013). Molecular analysis of the UV-inducible pili operon from *Sulfolobus acidocaldarius*. *Microbiol. Open* 2, 928–937. doi: 10.1002/mbo3.128
- Vestergaard, G., Garrett, R. A., and Shah, S. A. (2014). CRISPR adaptive immune systems of archaea. *RNA Biol.* 11, 156–167. doi: 10.4161/rna.27990
- Wagner, M., van Wolferen, M., Wagner, A., Lassak, K., Meyer, B. H., Reimann, J., et al. (2012). Versatile genetic tool box for the crenarchaeote *sulfolobus acidocaldarius*. *Front. Microbiol.* 3, 214. doi: 10.3389/fmicb.2012.00214
- Westra, E. R., van Erp, P. B. G., Künne, T., Wong, S. P., Staals, R. H. J., Seegers, C. L. C., et al. (2012). CRISPR immunity relies on the consecutive binding and degradation of negatively supercoiled invader DNA by Cascade and Cas3. *Mol. Cell* 46, 595–605. doi: 10.1016/j.molcel.2012.03.018
- Yang, H., Ren, S., Yu, S., Pan, H., Li, T., Ge, S., et al. (2020). Methods favoring homology-directed repair choice in response to CRISPR/Cas9 induced-double strand breaks. *Int. J. Mol. Sci.* 21, 8461. doi: 10.3390/ijms21186461
- Zebec, Z., Zink, I. A., Kerou, M., and Schleper, C. (2016). Efficient CRISPR-mediated post-transcriptional gene silencing in a hyperthermophilic archaeon using multiplexed crRNA expression. *G3* 6, 3161–3168. doi: 10.1534/g3.116.032482
- Zhang, J., and White, M. F. (2013). Hot and crispy: CRISPR-Cas systems in the hyperthermophile *Sulfolobus solfataricus*. *Biochem. Soc. Transact.* 41, 1422–1426. doi: 10.1042/BST20130031
- Zink, I. A., Fouqueau, T., Tarrason Risa, G., Werner, F., Baum, B., Bläsi, U., et al. (2021). Comparative CRISPR type III-based knockdown of essential genes in hyperthermophilic *Sulfolobales* and the evasion of lethal gene silencing. *RNA Biol.* 18, 421–434. doi: 10.1080/15476286.2020.1813411
- Zink, I. A., Wimmer, E., and Schleper, C. (2020). Heavily Armed Ancestors: CRISPR immunity and applications in archaea with a comparative analysis of CRISPR types in *sulfolobales*. *Biomolecules* 10, 11523. doi: 10.3390/biom10111523



OPEN ACCESS

EDITED BY

Marleen van Wolferen,
University of Freiburg, Germany

REVIEWED BY

Tamara Basta,
UMR9198 Institut de Biologie Intégrative de la
Cellule (I2BC), France
Nuno Machado,
Oxford Nanopore Technologies,
United Kingdom
Shamphavi Sivabalasarma,
University of Freiburg, Germany,
in collaboration with reviewer NM

*CORRESPONDENCE

Dina Grohmann
✉ dina.grohmann@ur.de
Felix Grünberger
✉ felix.gruenberger@ur.de

[†]These authors have contributed equally to this work and share first authorship

RECEIVED 16 June 2023

ACCEPTED 23 October 2023

PUBLISHED 09 November 2023

CITATION

Stöckl R, Nißl L, Reichelt R, Rachel R,
Grohmann D and Grünberger F (2023) The
transcriptional regulator EarA and intergenic
terminator sequences modulate archaeallation
in *Pyrococcus furiosus*.
Front. Microbiol. 14:1241399.
doi: 10.3389/fmicb.2023.1241399

COPYRIGHT

© 2023 Stöckl, Nißl, Reichelt, Rachel,
Grohmann and Grünberger. This is an open-
access article distributed under the terms of
the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/)
(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 transcriptional regulator EarA and intergenic terminator sequences modulate archaeallation in *Pyrococcus furiosus*

Richard Stöckl^{1†}, Laura Nißl^{1†}, Robert Reichelt¹, Reinhard Rachel²,
Dina Grohmann^{1*} and Felix Grünberger^{1*}

¹Institute of Microbiology and Archaea Centre, Faculty for Biology and Preclinical Medicine, University of Regensburg, Regensburg, Germany, ²Centre for Electron Microscopy, Faculty for Biology and Preclinical Medicine, University of Regensburg, Regensburg, Germany

The regulation of archaeallation, the formation of archaeal-specific cell appendages called archaella, is crucial for the motility, adhesion, and survival of archaeal organisms. Although the heavily archaeallated and highly motile *Pyrococcus furiosus* is a key model organism for understanding the production and function of archaella in Euryarchaea, the transcriptional regulation of archaellum assembly is so far unknown. Here we show that the transcription factor EarA is the master regulator of the archaellum (*arl*) operon transcription, which is further modulated by intergenic transcription termination signals. *EarA* deletion or overexpression strains demonstrate that EarA is essential for archaeallation in *P. furiosus* and governs the degree of archaeallation. Providing a single-molecule update on the transcriptional landscape of the *arl* operon in *P. furiosus*, we identify sequence motifs for EarA binding upstream of the *arl* operon and intergenic terminator sequences as critical elements for fine-tuning the expression of the multicistronic *arl* cluster. Furthermore, transcriptome re-analysis across different *Thermococcales* species demonstrated a heterogeneous production of major archaellins, suggesting a more diverse composition of archaella than previously recognized. Overall, our study provides novel insights into the transcriptional regulation of archaeallation and highlights the essential role of EarA in *Pyrococcus furiosus*. These findings advance our understanding of the mechanisms governing archaeallation and have implications for the functional diversity of archaella.

KEYWORDS

Archaea, archaellum, transcriptomics, EarA, *Thermococcales*, single-molecule sequencing, transcriptional regulator

1. Introduction

Extracellular locomotive structures are ubiquitous across all domains of life and known as cilia or flagella in Eukaryotes, flagella in Bacteria, and archaella in Archaea (Khan and Scholey, 2018; Beeby et al., 2020). Though analogous in function to flagella, the archaellum is distinct in its evolutionary origin, genetic makeup, and the structural organization of its motility machinery (Jarrell et al., 2021; Nuno De Sousa Machado et al., 2022). Sharing similarities to the bacterial Type IV pili system, the archaellum distinguishes itself by its unique assembly and ATP-dependent rotation mechanism, in stark contrast to bacterial flagella, which operate based on ion gradients (Albers and Jarrell, 2015; Poweleit et al., 2016). Notably, in some archaea, the

function of archaella extends beyond swimming and these cell appendages contribute to surface adhesion and biofilm formation through intra- and interspecies cell–cell contacts. For instance, in the hyperthermophilic model archaeon *Pyrococcus furiosus*, archaella have been demonstrated to facilitate the attachment of cells to black smoker material taken from their native habitat, indicating their role in colonizing extreme environments (Näther et al., 2006; Schopf et al., 2008; Wirth, 2017; Wirth et al., 2018).

The complex architecture of the archaellum encompasses multiple components, including filament proteins (ArlA and/or ArlB), stator elements (ArlF, ArlG), proteins potentially needed for motor function (ArlC, ArlD, ArlE, ArlX), ATPase and ATPase-modulating proteins (ArlI, ArlH), and platform proteins (ArlJ) [for an overview see (Jarrell et al., 2021)]. Most of these components are encoded within the archaellum (*arl*) gene cluster. Arl gene clusters can be divided into the highly conserved Arl1 or Arl2 cluster that differ in their composition and phylogeny (Desmond et al., 2007). However, genes encoding associated proteins, such as enzymes for N-glycosylation of archaellins and chemotaxis system components, may be located in different regions of the genome (Schlesner et al., 2009; Jarrell et al., 2014).

To fine tune the expression of *arl* genes during different environmental conditions, diverse transcriptional regulation mechanisms have been discovered in Crenarchaeota (also known as Thermoproteota) (Jarrell et al., 2021; Oren and Garrity, 2021). In *Sulfolobus*, transcription factor (TF)-based regulation has been extensively studied, revealing several factors that control archaellum formation. Transcription factors ArnA and ArnB play key roles in this process (Reimann et al., 2012). Other regulators that either impair (ArnC, ArnS, Abf1, ArnR, ArnR1) or enhance (ArnD, PP2A) motility have also been identified (Lassak et al., 2013; Haurat et al., 2017; Hoffmann et al., 2017; Li et al., 2017). Notably, these proteins are absent in Euryarchaeota (also known as Methanobacteriota), where only a single regulator, the euryarchaeal archaellum regulator (EarA), was recently identified (Ding et al., 2016, 2017b).

This factor is widely distributed in the Euryarchaeota and has been shown to be indispensable for transcription of the *arl* operon in *Methanococcus maripaludis*, which encodes most proteins involved in archaellum formation (Ding et al., 2016, 2017b). Interestingly, spontaneous mutation in the promoter region of an EarA knockout strain restored archaellum functionality (Ding et al., 2017a). Although the molecular mechanism was not revealed yet, it led to the hypothesis that EarA might aid in recruiting transcription factor B to a weak B recognition element (BRE) in the wild-type *arl* promoter, but may not be essential when the *arl* promoter has a strong BRE (Ding et al., 2017a). While a deletion mutant could be restored when EarA homologs from other archaea were heterologously expressed in *M. maripaludis*, this was not possible with the EarA variant from *Pyrococcus furiosus*, which could be attributed to lower sequence identity or codon usage issues (Ding et al., 2017b). This is particularly intriguing because among Euryarchaeota, *Pyrococcus* has served as a key model system to further our understanding of the archaellum machinery at the functional, structural, and regulatory levels (Näther et al., 2006; Näther-Schindler et al., 2014; Lewis et al., 2015; Daum et al., 2017; Chaudhury et al., 2018).

Given this discrepancy, our study investigated the role of the euryarchaeal archaellum regulator EarA in *P. furiosus*, finding that EarA is also essential for archaellation in this hyperthermophilic organism. Additionally, we present an updated single-molecule perspective of the transcriptional landscape of the *arl* operon. Finally, we contextualized our findings by integrating transcriptomics data

from across the *Thermococcales* and examining the potential influence of sequence elements on the expression of the *arl* operon. This includes the presence of EarA binding sequences upstream of the *arl* promoter, as well as intergenic terminator sequences. Together these data suggest that while the transcription factor EarA is the key regulatory player of archaellum expression, intergenic sequences help in fine-tuning other components of the *arl* operon in *Thermococcales*.

2. Materials and methods

2.1. Strains, plasmids, primers, and antibodies

All strains, plasmids, primers, and antibodies used in this study are listed in [Supplementary Table 1](#).

2.2. Construction of vectors and transformation of *Pyrococcus furiosus*

For the construction of *P. furiosus* strains $\Delta earA$, $\Delta earA$ recovery and $\Delta earA$ overexpression, the modified genetic system for *P. furiosus* DSM3638 based on selection via agmatine auxotrophy as described in Grünberger et al. (2021) was used.

For markerless disruption of the *P. furiosus earA* gene, the upstream and downstream flanking regions of the *P. furiosus earA* gene including the promoter region were amplified using the primer pairs 0340upAscFW/0340upFusRW and 0340doFW/0340doNotIRW ([Supplementary Table 1](#)). Both PCR products were used as a template for a single-overlap extension PCR and the resulting PCR product was cloned into the modified pMUR47 vector containing a two-gene resistance cassette via the AscI and NotI restriction sites (Grünberger et al., 2021). Successful and correct insertion of the PCR fragment into the vector was verified by DNA sequencing resulting in the plasmid pMUR443.

Circular plasmid DNA pMUR443 and strain MURPf37 were used for transformation and selection was carried out in 1/2 SME-starch liquid medium without agmatine sulfate and inosine+guanosine (I + G) at 85°C for 12h (Grünberger et al., 2021). Pure cultures of the intermediate mutant MURPf76_i were obtained by plating the cells on solidified medium. The integration of the plasmid into the genome by single cross-over was verified by analyzing corresponding PCR products. For the counter selection cells were plated on solidified medium containing 50 μ M 6-methylpurine and 8 mM agmatine sulfate to induce a second homologous recombination step to recycle the selection marker and to eliminate integrated plasmid sequences. The genotype of the final mutant (MURPf76) was confirmed by PCR and cells had to be grown in the presence of 8 mM agmatine sulfate and 8 mM I + G.

For EarA recovery and EarA overexpression in trans, a shuttle vector system for *P. furiosus* was used. Based on plasmid pMUR310 (Grünberger et al., 2021), an overexpression shuttle vector system was constructed that allows constitutive expression of a protein of interest via the *gdh* promoter (PCR amplified from *T. kodakarensis* genomic DNA using primers RPA_pYS13_GA_F/pYS13_GA-R) and the *hypA1* gene terminator (PCR amplified from *P. furiosus* genomic DNA using primers PYS14_GA_F/pYS13_RPA_GA_R) regions. Both regions were fused upstream and downstream of the gene of interest as described in

Waage et al. (2010) and cloned into pMUR310 via the EcoRV restriction site. One of these overexpression plasmids was used for PCR amplification of the plasmid backbone using primers pYS14Exp gdh_F /pYS14_GA_F. The *earA* gene was PCR amplified from *P. furiosus* genomic DNA using primers flaaktivGap914F/ flaaktivgap914r. Both PCR products were assembled using Gibson Assembly (Gibson et al., 2009) resulting in the plasmid pMUR444. Next, the Strep-tag® II sequence was inserted at the C-terminus via Phusion® High-Fidelity DNA Polymerase mutagenesis PCR (New England Biolabs) using the primers pYS14_GA_F/ 5' phosphorylated flaakstrp14R to amplify the whole plasmid. The PCR product was ligated using T4 ligase (New England Biolabs) and transformed into *E. coli* DH5 α cells (New England Biolabs), resulting in pMUR445. The final construct including the correct insertion of the Strep-tag® II sequence was verified by sequencing (Microsynth). Finally, to construct a shuttle vector expressing EarA under its native promoter the pMUR444 backbone including the *earA* gene, but without *gdh* promoter, was PCR amplified using primers flaaktivpromGap914F/GAp14bbminpromR. The *earA* promoter region was PCR amplified from *P. furiosus* genomic DNA using primers p14fusPF0340prom/034upRW1. Both PCR products were assembled using Gibson Assembly (Gibson et al., 2009) resulting in the plasmid pMUR458. The correct exchange of promoter regions was verified by sequencing (Microsynth).

1 μ g of the circular plasmids pMUR310, pMUR445 and pMUR458 were transformed into MURPf76 as described previously (Waage et al., 2010; Kreuzer et al., 2013; Grünberger et al., 2021). Selection was carried out in 1/2 SME liquid medium without agmatine sulfate and I + G at 85°C for 12 h. Pure cultures of the mutants MURPf79 (Δ *earA* strain containing pMUR310), MURPf77 (Δ *earA* recovery strain containing pMUR458) and MURPf78 (*earA* overexpression strain containing pMUR445) were obtained by plating the cells on solidified medium. Plasmid stability was verified by re-transformation into *E. coli* DH5 α and DNA sequencing of purified plasmids (Microsynth). Final *P. furiosus* mutants could be grown in 1/2 SME medium (Fiala and Stetter, 1986) supplemented with 0.1% starch, 0.1% peptone and 0.1% yeast extract but without agmatine sulfate and I+G supplementation (Grünberger et al., 2021).

2.3. Analysis of swimming behavior

Pyrococcus furiosus strains were grown at 85°C in 1/2 SME medium (Fiala and Stetter, 1986) supplemented with 0.1% starch, 0.1% peptone and 0.1% yeast extract to a cell density of $\sim 1 \times 10^6$ cells per mL and swimming experiments were performed using a temperature gradient forming device (TGFD) at 100°C as described in Mora et al. (2014). Movies were recorded and analyzed as described previously (Herzog and Wirth, 2012; Mora et al., 2014).

2.4. Western blot analysis

Pyrococcus furiosus strains were grown at 85°C in 1/2 SME medium (Fiala and Stetter, 1986) supplemented with 0.1% starch, 0.1% peptone and 0.1% yeast extract to a cell density of $\sim 1 \times 10^8$ cells per mL and harvested by centrifugation at 20,500g (4°C). The cell pellets from 20 mL cultures were resuspended in 1x PBS buffer and the cell suspensions were sonicated for 50 s using a Bandelin Electronic™

Sonopuls™ HD 3400 homogenizer (cycle: 80%, power 80%). After centrifugation for 30 min at 16,200g (4°C), total protein concentrations of the cell extracts were quantified using the Qubit™ Protein Assay Kit 3.0 Fluorometer (ThermoFisher). 2 μ g of total protein extracts from each *P. furiosus* strain were separated using SDS-PAGE. Western blots were performed as described previously (Waage et al., 2010).

2.5. PCR test

Pyrococcus furiosus strains were grown at 85°C in 1/2 SME medium (Fiala and Stetter, 1986) supplemented with 0.1% starch, 0.1% peptone and 0.1% yeast extract to a cell density of $\sim 1 \times 10^8$ cells per mL and harvested by centrifugation at 20,500g (4°C). Genomic DNAs were isolated using the ReliaPrep™ gDNA Tissue Miniprep System (Promega) following the Mouse tail protocol. The region of the *arlD* gene was PCR amplified using the primers Fladup_100fw/ Fladdo_100rw, and the region of the *earA* gene locus was PCR amplified using the primers 0340do_113rw/0340up_100fw, respectively. PCR products were separated by agarose gel electrophoresis.

2.6. Transmission electron microscopy

Pyrococcus furiosus cells grown to the exponential phase were taken up in a syringe and carefully filtered through a polyamide/nylon filter (pore size 0.2 μ m). The concentrated cell film was resuspended in 20–40 μ L fresh media and chemically fixed with 1% (v/v) glutaraldehyde (final concentration) at room temperature. 5 μ L of the fixed cell suspension were applied onto copper grids (400-mesh; G2400C, Plano) coated with a 10 nm carbon film (Carbon Coater 208 Turbo, Cressington), washed twice by blotting with distilled water, and the samples were subsequently shadowed with ~ 1.0 nm Pt/C (15° angle; CFE-50, Cressington). Transmission electron micrographs were recorded on a CM12 transmission electron microscope (FEI) operated at 120 keV and fitted with a slow-scan CCD camera (TEM 0124; TVIPS) using EMMENU v4.0 (TVIPS).

2.7. Gene calling and re-annotation

All currently available *Thermococcales* RefSeq reference genomes with a completeness of scaffold or better were acquired via the NCBI genome page (compare Supplementary Table 2).¹

The genomes were annotated via the arCOG definitions as described elsewhere (Dombrowski et al., 2020). In short, CDS sites and protein sequences predicted with Prokka (v. 1.14.6) (Seemann, 2014) were searched against arCOG² (Makarova et al., 2015) Hidden Markov Models (HMM) prepared with hmmbuild using hmmsearch (HMMER v. 3.3.2; Nov 2020).³ The best hit was assigned as annotation based on bit score then e-value, after an e-value cutoff $\leq 1e-3$ (see Supplementary Table 2).

1 https://www.ncbi.nlm.nih.gov/data-hub/genome/?taxon=2258&reference_only=true&assembly_level=1%3A3; accessed 2023-05-12.

2 <https://ftp.ncbi.nlm.nih.gov/pub/wolf/COGs/arCOG/zip.aliar14.tgz>

3 <http://hmmer.org/>

2.8. Re-analysis of short-read RNA-seq data

For coverage analysis of the *arl* gene cluster the following RNA-seq datasets were reanalyzed: Illumina data from *Pyrococcus furiosus* DSM3638 exponential growth phase at the optimum growth temperature of 95°C (ERR11200497, ERR11200498, ERR11200499, ERR11200500) (Grünberger et al., 2023), *Palaeococcus pacificus* DY20341 cells from exponential growth phase with S⁰ addition (SRR10749090) (Zeng et al., 2020), *Thermococcus onnurineus* NA1 cells from exponential phase in CO-supplied media (SRR1702360, SRR1702361) (Lee et al., 2015), *Thermococcus litoralis* DSM5473 cells from exponential phase (SRR12486532, SRR12486546, SRR12486547) (Liang et al., 2021) and *Thermococcus kodakarensis* KOD1 cells from exponential phase (ERR6384078, ERR6384079, ERR6384080, ERR6384081) (Villain et al., 2021).

Sequencing reads in fastq format were filtered and trimmed using fastp (v. 0.23.3) to remove low-quality reads and adapter sequences (`--cut_front`, `--cut_tail`, `--qualified_quality_phred 30`, `--length_required 30`) using parameters selective for single-end or paired-end reads (Chen et al., 2018). Subsequently, trimmed reads were aligned to the respective NCBI reference genomes (*P. furiosus*: NZ_CP023154.1, *P. pacificus*: NZ_CP006019.1, *T. onnurineus*: NC_011529.1, *T. litoralis*: NC_022084.1, *T. kodakarensis*: NC_006624.1) using Bowtie2 (v.2.5.1) with default parameters (Langmead and Salzberg, 2012). The resulting sequence alignment files (SAM) were converted to binary mapping format (BAM) and sorted using samtools (v.1.17) (Li et al., 2009). Additionally, position-specific coverage files were generated using samtools depth (`-a -j`), including reads with deletions in the coverage computation. Downstream normalization and visualization was performed using the Tidyverse in R (Wickham et al., 2019). Briefly, coverage for each position was first normalized by counts per million (CPM), before calculating the mean value of all replicates.

For calculation of transcripts per million (TPM) values of the protein-coding genes, featureCounts (RSubread package v. 2.10.5) was used to calculate the count matrix based on custom GFF files containing all *arl* genes detected using HMM as described above (Liao et al., 2019). Next, TPM values were calculated by dividing the number of reads mapping to each gene by the gene length in kilobases, then dividing the resulting reads per kilobase (RPK) values by the sum of all RPK values in the sample, and finally multiplying the quotient by one million. For visualization, the z-score was calculated for all genes in the *arl* gene cluster region, including the two upstream genes.

2.9. Re-analysis of Nanopore RNA-seq data

Nanopore data from two biological replicates of *Pyrococcus furiosus* DSM3638 cells from exponential growth phase grown at the optimum growth temperature of 95°C were reanalyzed (ENA-project: PRJEB61177, runs: ERR11203080, ERR11203081) (Grünberger et al., 2023). Generation of the dataset, including detailed information about cell growth, RNA isolation, RNA treatment, library preparation, sequencing following the PCR-cDNA barcoding kit protocol (SQK-PCB109) from Oxford Nanopore Technologies (ONT) and data analysis is described in Grünberger et al. (2023). Briefly, basecalling and demultiplexing was performed using guppy (v. 6.4.2 + 97a7f06) in

high-accuracy mode. Full-length sequenced reads were detected, strand-oriented and trimmed using pychopper (v.2.7.2)⁴ with autotuned cutoffs and the edlib backend for identifying the custom VNprimer(5'-ACTTGCTGTCGCTCTATCTTCATTGATGGTGCC TACAG-3'). Mapping to the *P. furiosus* DSM 3638 genome (NCBI:NZ_CP023154.1) was performed using minimap2 (v. 2.24-r1122) with standard parameters suggested for aligning Nanopore genomic reads (`-ax map-ont`) (Li, 2018, 2021). Soft or hard clipped (>5 bases) reads were removed using samclip (v. 0.4.0), and SAM files were converted to sorted BAM files using samtools (v. 1.16.1) (Li et al., 2009). Coverage files generation including samtools depth and CPM normalization in R was performed as described for the short-read datasets. Single-read tracks of unspliced reads (`njunc=0`) were plotted using ggplot2 `geom_segment` from read start to end.

2.10. Sequence analysis and motif identification

Annotations were filtered using R and the Tidyverse packages (v. 2.0.0) (Wickham et al., 2019). Genomic sequences annotated as EarA were extracted and translated using seqkit (v. 2.4.0) (Shen et al., 2016). EarA protein sequences were aligned using ClustalO (v. 1.2.3) (Sievers and Higgins, 2014) with default parameters and visualized using ggmsa (v. 1.4.0) (Zhou et al., 2022).

The genomic locations that matched the “FlaB” HMM were annotated as “arlB” and assigned a rank, with the most-upstream location designated as “arlB0” to be consistent with the annotation used in *Pyrococcus furiosus*. Please note that this deviates from the database annotation of other strains that start with “arlB1.” Genomic regions located 150 nt upstream of *arlB0* and 50 nt downstream of all *arlB* genes were defined using plyranges (v. 1.18.0) (Lee et al., 2019). The sequences corresponding to these regions were extracted using seqkit subseq.

Motif analysis was conducted on the sequences upstream of the *arlB0* genes using MEME (v. 5.5.2; `--mod anr --dna --w 8`) (Bailey et al., 2015). The identified motifs and their respective locations were visualized using the ggmotif (v. 0.2.1) (Li et al., 2022), ggseqlogo (v. 0.1) (Wagih, 2017), memes (v. 1.6.0), and Tidyverse packages.

The enrichment of “T” bases in the genomic regions downstream of the *arlB* genes was calculated as the log₂ fold change of the letter frequency in the 50 nt downstream region, obtained using bedtools nuc (v. 2.31.0) (Quinlan and Hall, 2010), to the average letter frequency in the whole genome retrieved using the ‘letterFrequency (“T”)’ function from the Biostrings package (v. 2.66.0).

3. Results

3.1. The euryarchaeal archaeellum regulator EarA is essential for archaeation in *Pyrococcus furiosus*

To elucidate the *in vivo* role of the regulator EarA, we took advantage of the existing genetic system of *P. furiosus* based on

⁴ <https://github.com/epi2me-labs/pychopper>

agmatine auxotrophy (parental wildtype) (Waage et al., 2010) and constructed a mutant strain ($\Delta earA$) in which the *earA* gene locus including the corresponding promoter region was deleted. Gene rescue strains expressed EarA *in trans* either under the control of its native promoter ($\Delta earA$ recovery) or under the control of the strong, constitutive *gdh* promoter ($\Delta earA$ overexpression) (Micorescu et al., 2008). Successful deletion of the genomic *earA* locus was confirmed via PCR (Supplementary Figure 1A). Additionally, we assessed the recovery and overexpression of EarA levels within the $\Delta earA$ background by quantifying the production of archaeellins using Western blot analysis (Supplementary Figures 1B,C).

The archaeellation of these strains was visualized by transmission electron microscopy of Platinum/Carbon shadowed samples (Figures 1A–D; Supplementary Figure 2). The *earA* gene deletion led primarily to cells devoid of archaeella, whereas overexpression resulted in heavily archaeellated *P. furiosus* cells (Figures 1A–D). These outcomes mirrored the observed swimming behaviors, which indicated that no or few archaeella in the deletion mutant are not sufficient to allow the cells to swim normally (data not shown).

To systematically evaluate the observed initial differences in the extent of archaeellation, we pursued further analysis by classifying the archaeellation level based on the number of visible archaeella of approximately 70–90 cells for each strain used in this study as none (0), low (1–10), mid (11–40) and high (>40) (Figure 1E; Supplementary Figure 2A). The increased EarA protein levels observed in the $\Delta earA$ overexpression strain indeed correlated with a significantly higher *in vivo* archaeellation level clearly surpassing both the parental wildtype and the recovery strain (Figure 1F; Supplementary Figure 2B). These results underscore EarA as a pivotal regulator of archaeellation in *P. furiosus*, suggesting that the cellular levels of this transcriptional regulator directly govern the degree of archaeellation.

3.2. A single-molecule update on the transcriptional regulation of the *arl* gene cluster in *Pyrococcus furiosus*

To explore the potential hierarchic nature of EarA-induced transcriptional regulation, we analyzed the transcriptome landscape of the *arl* gene cluster in *Pyrococcus furiosus*. Therefore, we re-analyzed both short- and long-read RNA-seq data in addition to other database information, including transcription start and termination sites (Figure 2; Grünberger et al., 2019, 2023). The Arl1-type cluster is characterized by three archaeellin genes (*arlB0*, *arlB1*, *arlB2*) followed by the *arl* accessory genes *arlC*, *arlD*, *arlF*, *arlG*, *arlH*, *arlI*, *arlJ* and a hypothetical gene (Näther-Schindler et al., 2014; Jarrell et al., 2021). According to the DOOR2 database annotation, two genes preceding the *arl* genes, namely the transcriptional regulator EarA and the postulated methyltransferase *fam*, are part of the operon (Figures 2A,B). In contrast, prediction based on short-read RNA-seq data from mixed conditions suggests that the two genes are transcribed as an extra single unit (Figure 2A; Mao et al., 2014; Grünberger et al., 2019).

This was further supported by coverage re-analysis of short-read RNA-seq recently obtained from cells grown to exponential phase and cells recovered at the optimal temperature of 95°C following an extended cold shock at 4°C (Figures 2C,D; Grünberger et al., 2023).

Cells recovered from such conditions have been demonstrated to respond swiftly by swimming, which was confirmed by the induction of the *arl* gene cluster (Mora et al., 2014). Interestingly, we noted a considerable upregulation of the EarA regulator (\log_2 fold change: 2.3), initiating the activation of the *arl* genes. The level of activation decreased as the distance from the primary transcription start site increased.

To achieve a higher resolution of the transcriptional output, we re-analyzed long-read single-molecule data from samples cultivated under optimal growth conditions (Grünberger et al., 2023). Previous transcriptional start site mapping using differential RNA sequencing of Terminator-Exonuclease (TEX) treated RNA revealed that the primary start site is situated upstream of *arlB0*, the essential major archaeellin protein (Grünberger et al., 2019). Indeed, the majority of the reads from long-read sequencing originate from this position. Furthermore, long-read coverage suggests differential expression of multiple transcription units, prompting us to refine and update the transcription model based on RT-PCR analysis (Figures 2E–H; Näther-Schindler et al., 2014). In alignment with a previous study and structural data, our analysis verified that *arlB0* is the primary archaeellin transcript, particularly during early exponential growth (Näther-Schindler et al., 2014; Daum et al., 2017). However, *arlB0* can be detected as co-transcript with downstream genes but also as a single transcript. Notably, we discovered several shortened units derived from the primary *arl* operon transcript (Figure 2H). These units exhibited a subset of reads originating from the upstream position of *arlB0*. Moreover, we detected transcripts with random downstream read starts that could not be attributed to promoter elements, suggesting the occurrence of 5' processing events. Our findings revealed that transcripts spanning from *arlB0* to *arlJ* were not detected, in agreement with the previous study (Näther-Schindler et al., 2014). Importantly, the absence of enriched 5' ends in our dataset, including the TEX-enriched deep-sequenced RNA-seq data from cells under various conditions, strongly suggests that all transcripts likely originate from *arlB0*. The hierarchical gene expression changes observed in the cold-shocked recovery cells further support this conclusion.

To validate these findings, we performed additional analysis on enriched 3' ends and conducted a terminator sequence analysis (Figures 2I–K). This analysis confirmed the observed coverage profiles and highlighted the presence of poly(T)-sequences after *arlB0*, *arlB2*, and *arlD* that may serve as intergenic transcriptional terminators.

3.3. Investigation of transcriptome data and sequence features in *Thermococcales* reveal a heterogeneous production of major archaeellins

We expanded our study by reanalyzing RNA-seq data from various organisms within the order *Thermococcales* including *Palaeococcus pacificus*, *Thermococcus onnurineus*, *Thermococcus kodakarensis*, and *Thermococcus litoralis* to understand the transcriptional characteristics of the *arl* gene cluster (Lee et al., 2015; Zeng et al., 2020; Liang et al., 2021; Villain et al., 2021; Grünberger et al., 2023). While the overall Arl1-like operon organization is well conserved among these organisms, they exhibited variations in the number of *arlB* genes, ranging from two

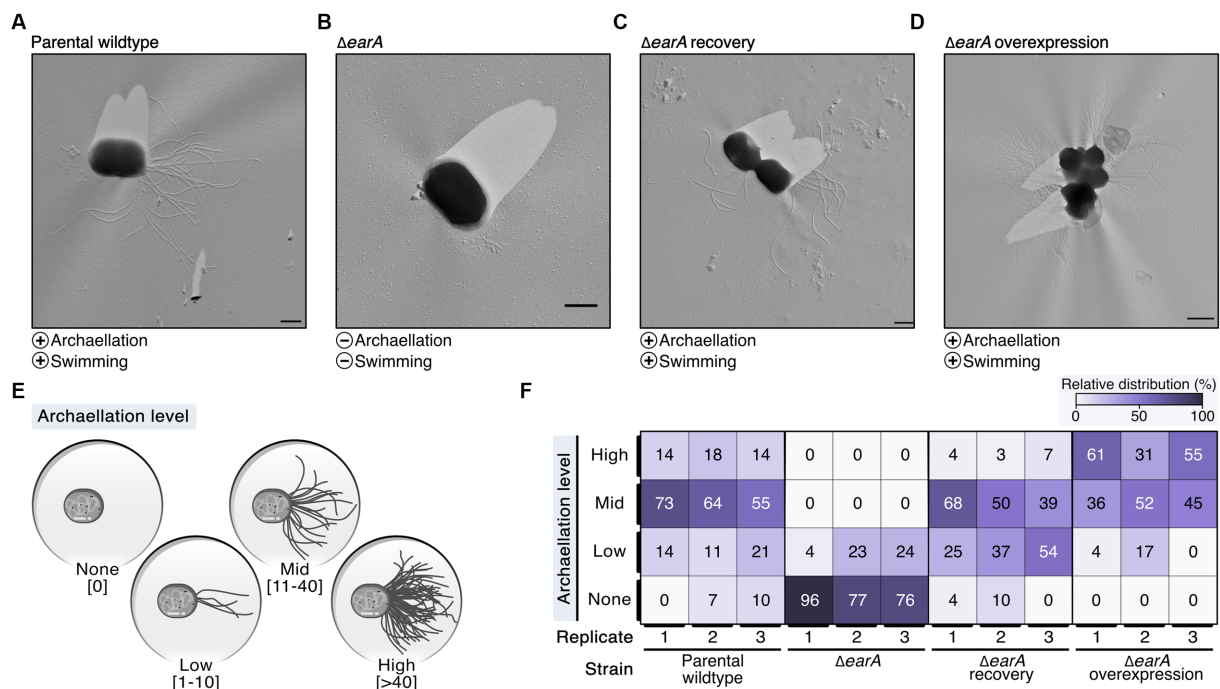


FIGURE 1

Electron microscopy evaluation of archaella formation in different genetic variants of *Pyrococcus furiosus* illustrates the central role of EarA.

(A) Transmission electron microscopy (TEM) micrographs of representative cells observed in the parental wildtype, (B) $\Delta earA$ knockout strain, (C) $\Delta earA$ recovery (native promoter) and (D) $\Delta earA$ overexpression strain (*gdh* promoter). The presence and absence of archaella and swimming behavior are denoted by plus and minus symbols, respectively. Scale bar: 0.5 μ m. (E) Systematic classification of the archaeallation level based on the number of observable archaella originating from one cell. (F) Color-coded representation (from 0% in white to 100% in violet) of the relative distribution of the four archaeallation levels across the strains under investigation.

to five gene variants (Figure 3A). Broadening the genomic analysis, in all but six of the 33 *Thermococcales* reference genomes currently available in the NCBI genomes page, arCOG Hidden Markov Model (HMM) matches for the genes encoding for EarA and Arl proteins were found (Supplementary Figure 3; Supplementary Table 2). The transcriptional regulator EarA is highly conserved in *Thermococcales* and mostly oriented in the same orientation as the *arlB* cluster, with some exceptions (Supplementary Figures 3, 4; Ding et al., 2017b). Note, that the NCBI annotations for the six reference genomes without HMM matches do not contain archaellin annotations and three of these organisms were described as “non-motile” or “not-flagellated” (Canganella et al., 1998; Grote et al., 1999; Hensley et al., 2014).

Comparing the coverage profiles of *Pyrococcus* with the other strains, we consistently observed a higher number of transcripts in the *arlB* region compared to the downstream accessory *arl* genes (Figures 3A,B). This finding indicates that under exponential growth conditions, *arlB* transcripts are the most abundant in all *Thermococcales*. In contrast to *Pyrococcus*, there were notable differences in the transcriptional output of the other strains. Notably, *arlB0* was not consistently the major archaellin based on expression levels. We observed some variability of the transcript levels, with either *arlB1* or *arlB2* as the most highly expressed gene or a balanced distribution between the *arlB* genes (Figures 2A,B). This balance was particularly evident for the five tandem co-transcribed archaellin genes in *T. kodakarensis*, aligning with previous Northern Blot results (Nagahisa et al., 1999). Also, the transcript counts of the regulator

earA were significantly higher or equal to the *arlB* expression levels in the *Thermococcus* strains, suggesting it is transcribed as a separate unit.

Since previous analysis showed that EarA binds to a 6bp consensus sequence (TACATA) in the promoter region of the first archaellin gene in *Methanococcus*, we performed sequence analysis in all *Thermococcales* (Ding et al., 2017b). We found that the *arl* loci in all organisms examined were preceded by archaea-typical BRE and TATA elements, which were uniformly located at a similar distance from the start codon of the first *arl* gene (Figures 3C,D). Specifically, the *arlB0* gene in *P. furiosus* featured a 30 nt long 5' UTR (Grünberger et al., 2019).

Moreover, most cases showed a motif A(C/G)(C/G)TACA(T/C) at a fixed distance of nine nucleotides upstream of the BRE element, often repeating up to three times adjacently. We observed three distinct patterns, with either one (e.g., *P. horikoshii*), two (*P. furiosus*), or three EarA recognition sites (*T. kodakarensis*). It is noteworthy that the motifs in all *Pyrococcus* strains were shifted to either one window upstream or downstream compared to those in most of the *Thermococcus* strains. The absence of this motif in the upstream regions of all other *arl* genes supports the notion that EarA specifically binds upstream of the *arlB0* gene, thereby regulating the full gene cluster.

As a second layer of transcriptional regulation, we observed internal poly(T) sequences in *P. furiosus* (Figures 2I–K). Investigating whether this also applies to other *Thermococcales* and whether the strongest potential terminator sequence is positioned after the major archaellin, we calculated the enrichment of poly(T) sequences located

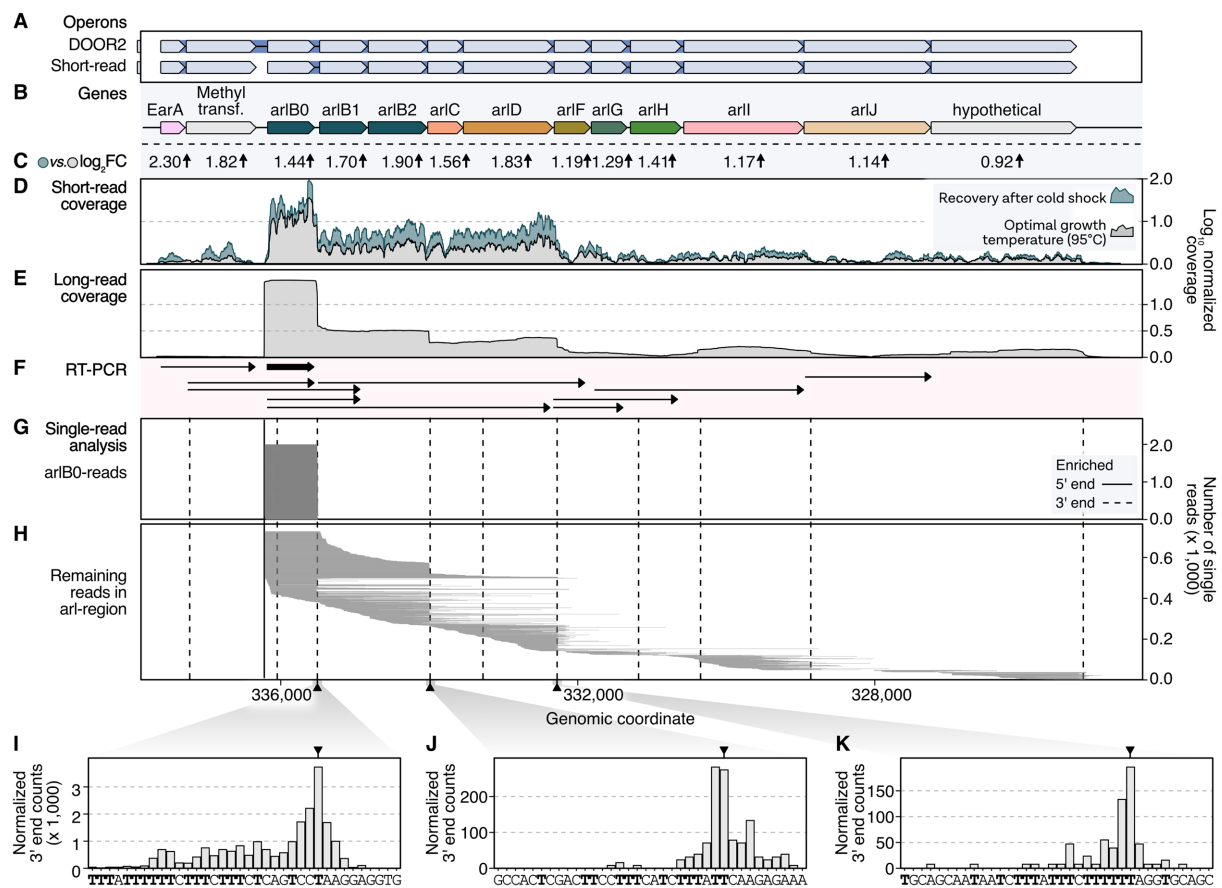


FIGURE 2

Long-read sequencing analysis provides high-resolution insights into transcriptional regulation of the *arl* gene cluster in *P. furiosus*. (A) Annotated operons in the DOOR2 database and as suggested by ANNOgesic using short-read RNA sequencing of mixed growth conditions (Mao et al., 2014; Grünberger et al., 2019). (B) Gene annotation of the *arl* cluster, including the transcriptional regulator EarA and a postulated methyltransferase upstream of the *arlB0* gene. (C) Log₂ fold changes obtained from differential gene expression analysis of samples recovered at 95°C after prolonged cold shock at 4°C (24h) compared to cells in exponential growth phase (Grünberger et al., 2023). (D) Mean coverage profiles normalized by CPM per position of these conditions are shown in gray (exponential phase) and blue (recovery) (Grünberger et al., 2023). (E) Mean long-read sequencing coverage profile (Nanopore SQK-PCB109) of cells in exponential growth phase. (F) RT-PCR results from Näther-Schindler et al. (2014). Detected transcripts are visualized as arrows. *ArIB0* was detected as the major archaeellin in *P. furiosus* and is highlighted in bold. (G) Single-read plotting of reads mapping to *arl* region, split by reads mapping to *arlB0* and (H) all other reads. Enriched primary 5' ends from short-read sequencing of TEX-treated RNAs (Grünberger et al., 2019) and primary 3' ends obtained from a recent Term-seq dataset (Grünberger et al., 2023) are shown as solid and dashed vertical lines, respectively. (I) Histograms of CPM-normalized 3' end positions in three regions [downstream of *arlB0*: (I) downstream of *arlB2*: (J) downstream of *arlD*: (K)] that showed significant enrichment of 3' ends using short-read Term-seq downstream of coding genes (Grünberger et al., 2023).

50 bases downstream of each *arlB* gene. Our results aligned with the overall expression levels, showing the highest T enrichment after the most abundant *arlB* gene (Figure 3E). The analysis across all *Thermococcales* confirmed our coverage analysis and indicated that, despite the *arl* gene organization, the most upstream *arl* gene is not always the most abundant gene, supporting the hypothesis that different ArlB might get incorporated into final archaeellum filaments.

4. Discussion

In this study, we highlighted the crucial role of the euryarchaeal archaeellum regulator EarA for archaeellation within *P. furiosus*. Using TEM analysis of different genetic variants, we gathered *in vivo* evidence that the expression levels of EarA directly correlate with the level of archaeellation. Given that the production of each protein

component of the archaeellum needs to be carefully controlled, we investigated how this control might be achieved on the transcriptional level (Chaban et al., 2007).

Our findings provide experimental support for prior speculations about EarA-based activation of transcription. Based on the occurrence of a spontaneous promoter mutation it was postulated that EarA might aid in the recruitment of TFB to a weak BRE element, marked by 7–8 consecutive adenines, in the *arl* gene cluster's promoter (Ding et al., 2017a). This hypothesis is supported by our motif analysis which reveals a similarly weak BRE sequence present in all *Thermococcales* preceding the first *arl* gene. Notably, the approximated spacing of 9–11 base pairs between the first upstream located EarA binding site and the BRE implies a direct interaction of EarA with TFB. This interaction is reminiscent of the activation pattern observed for the potent transcriptional activator 2 (Ptr2), which is presumed to interact directly with TBP to recruit it

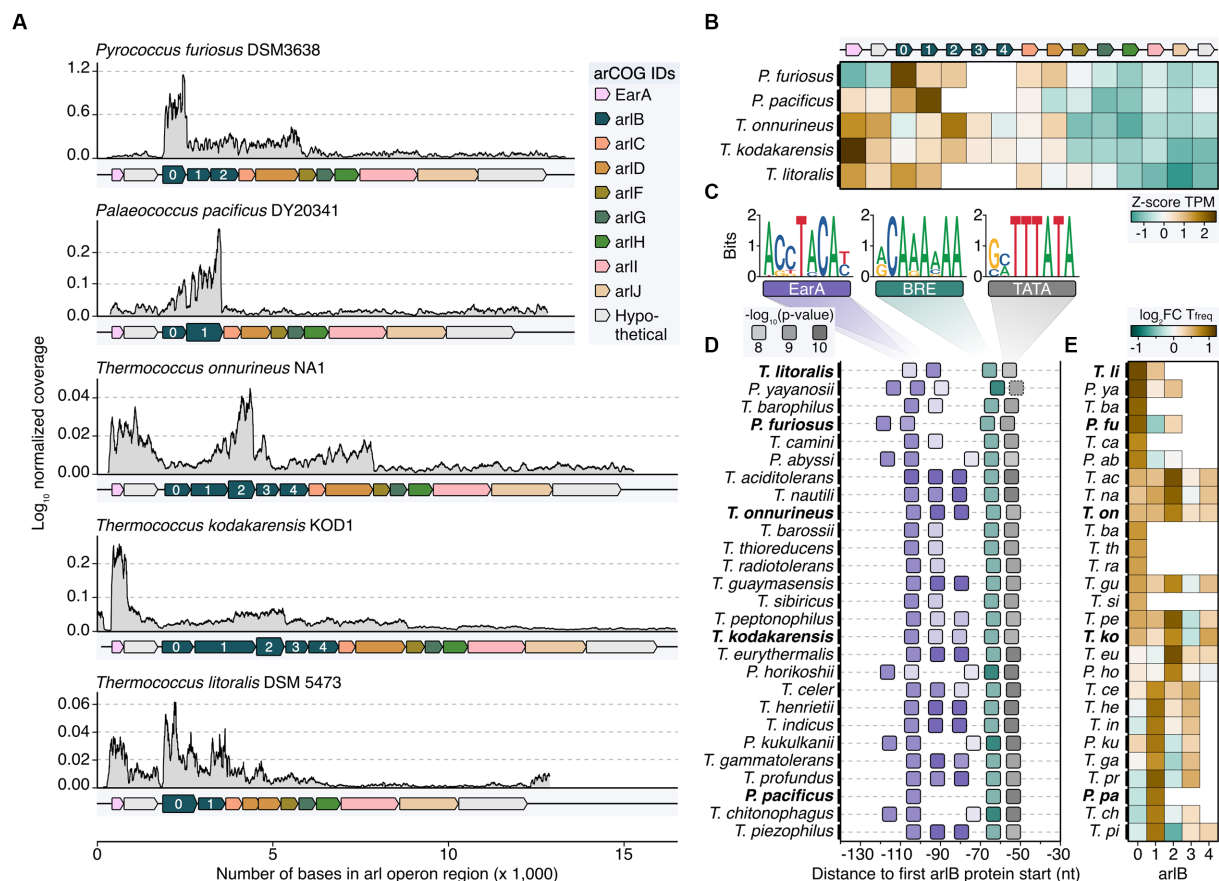


FIGURE 3

Transcriptome and sequence feature analysis reveal heterogeneous expression profiles of major archaeellins in *Thermococcales*. (A) Normalized coverage profiles from short-read sequencing of archaeal cells grown until exponential growth phase. Note that throughout the manuscript the most upstream *arl* gene was renamed *arlB0* in line with the current annotation in *P. furiosus*. Highest expressed *arlB* transcripts are highlighted in bold. (B) Z-score normalized TPM values (brown: high TPM, green: low TPM) for all genes shown in panel (A). (C) Enriched motifs potentially recognized by EarA (pink), TFB (BRE, green) and TBP (TATA, gray) identified using MEME search (Bailey et al., 2009) of the sequence region 150 nt upstream of the first *arl* gene. (D) Position of detected motifs (colored according to the description in C) are shown in a window from position -30 to -140 from the first *arlB* gene start. P-values are indicated by transparency of the boxes. Organisms shown in panels (A,B) are highlighted with bold text. The dashed box outline of the TATA box indicates that the element was not found by MEME but added after manual inspection. (E) Enrichment of the base "T" 50 nt downstream of the respective *arlB* gene over genomic "T" frequency. Boxes are filled according to the \log_2 fold enrichment in brown (high "T"-frequency) or green (low "T"-frequency).

to an imperfect TATA-box (Ouhammouch et al., 2005). Until now, only one archaeal transcriptional regulator, TFB recruitment factor 1 (TFB-RF1), that mediates TFB-recruitment, has been biochemically characterized in depth (Ochs et al., 2012). For both TFB-RF1 and EarA, it remains elusive which internal and/or external cellular stimuli impact their activities as transcriptional regulators (Reichelt et al., 2018). This aspect warrants more detailed examination, especially in organisms that encode both, the archaeellum machinery and chemotaxis systems, such as *T. kodakarensis* (Briegel et al., 2017). To address this, strains with targeted knockouts of EarA could be generated, followed by a comprehensive transcriptome analysis to assess global changes in gene expression and identify genes directly regulated by EarA. Additionally, a systematic approach of individually knocking out *arl* or chemotaxis genes, followed by phenotypic and transcriptomic analyses, would enable the identification of downstream effects and potentially the construction of a regulatory network, contributing to a better understanding of its cellular functions.

In contrast to transcriptional regulation at the level of initiation, relatively little is known about fine-tuning gene expression using terminator sequences, particularly in operons. In Archaea, poly(T) sequences have been shown to mediate intrinsic termination, as well as the recruitment of the recently identified termination factor aCPSF1, which is involved in 3' end formation (Maier and Marchfelder, 2019; Sanders et al., 2020; Yue et al., 2020; Li et al., 2021; Grünberger et al., 2023). Our findings suggest that operon-internal poly(T) sequences likely function as intergenic transcriptional terminators for transcripts originating from *arlB0*. This type of terminator-based regulation and the role of readthrough in archaeella accessory gene production have already been observed in *S. acidocaldarius* and appear to be important in *Thermococcales* as well (Lassak et al., 2012). Conceptually, this represents an elegant mechanism for the regulation of the multicistronic *arl* cluster allowing the fine-tuning of high-demand core components and lower-demand accessory genes (Thomas and Jarrell, 2001). However, extrapolating our findings to the final incorporation of the major

archaellin into the archaellum is not straightforward, due to the potential influences of post-transcriptional and post-translational regulation. Intriguingly, our single-molecule analysis of transcripts in *P. furiosus* showed that mRNA stability could indeed play a role, potentially explaining why the most upstream gene is not always the highest expressed.

In the case of *Thermococcus kodakarensis*, RNA-seq data compare favorably with a quantitative proteomics study. The latter disclosed similar proportions (0.4–0.25%) of total cell protein for *arlB0*, *arlB1*, and *arlB2*, hinting at a heterogeneous composition of the archaellum filament (Briegel et al., 2017). Although previous data were interpreted such that, like in Crenarchaeota, euryarchaeal archaella mainly consist of one type of archaellin only, emerging evidence contradicts this (Jarrell et al., 2021). Examples are the heterodimers of *arlB1* and *arlB2* observed in *Halorubrum lacusprofundi* (Pyatibratov et al., 2020) and *Methanocaldococcus villosus* (Gambelli et al., 2022). Whether this heterogeneity correlates with a structural organization of the archaellum analogous to the hook-, filament-, and cap-structures in the bacterial flagellum, or any type of multifunctionality of the archaellum, such as adhesion to cells or a variety of surfaces, warrants future investigation (Näther et al., 2006; Schopf et al., 2008; Weiner et al., 2012). Another intriguing prospect is exploring how the expression of a specific *ArlB* variant, presumably the principal archaellin, can be post-transcriptionally regulated if its encoding gene is not the most upstream *arlB* gene in the *arl* operon (*P. pacificus*, *T. onnurineus*).

In conclusion, our research uncovers new facets of the transcriptional regulation of the *arl* gene cluster in *P. furiosus* and establishes EarA as crucial transcriptional regulator for archaellation in *Pyrococcus furiosus*. Interestingly, the observed variation in *arlB* gene number and expression among *Thermococcales* indicates a more diverse production and assembly of archaellins than previously thought. Furthermore, our findings concerning the promoter motifs and intergenic poly(T)-sequences advance our understanding of the transcriptional regulation and potential mechanisms impacting the integration of *arl* gene products into archaellum filaments.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found at: <https://www.ncbi.nlm.nih.gov/sra/?term=SRR10749090> (*Palaeococcus pacificus* DY20341); <https://www.ncbi.nlm.nih.gov/sra/?term=SRR1304620> (*Pyrococcus abyssi*); <https://www.ncbi.nlm.nih.gov/sra/?term=ERR11200497> (*Pyrococcus furiosus*); <https://www.ncbi.nlm.nih.gov/sra/?term=ERR11200498> (*Pyrococcus furiosus*); <https://www.ncbi.nlm.nih.gov/sra/?term=ERR11200499> (*Pyrococcus furiosus*); <https://www.ncbi.nlm.nih.gov/sra/?term=ERR11200500> (*Pyrococcus furiosus*); <https://www.ncbi.nlm.nih.gov/sra/?term=ERR6384078> (*Thermococcus kodakarensis* KOD1); <https://www.ncbi.nlm.nih.gov/sra/?term=ERR6384079> (*Thermococcus kodakarensis* KOD1); <https://www.ncbi.nlm.nih.gov/sra/?term=ERR6384080> (*Thermococcus kodakarensis* KOD1); <https://www.ncbi.nlm.nih.gov/sra/?term=ERR6384081> (*Thermococcus kodakarensis* KOD1); <https://www.ncbi.nlm.nih.gov/sra/?term=ERR6384082> (*Thermococcus kodakarensis* KOD1); <https://www.ncbi.nlm.nih.gov/sra/?term=SRR12486532> (*Thermococcus litoralis* DSM 5473); <https://www.ncbi.nlm.nih.gov/sra/?term=SRR12486547> (*Thermococcus litoralis* DSM 5473); <https://www.ncbi.nlm.nih.gov/sra/?term=SRR12486546> (*Thermococcus litoralis* DSM 5473); <https://www.ncbi.nlm.nih.gov/sra/?term=SRR1702360> (*Thermococcus onnurineus* NA1); <https://www.ncbi.nlm.nih.gov/sra/?term=SRR1702361> (*Thermococcus onnurineus* NA1).

[sra/?term=SRR1702360](https://www.ncbi.nlm.nih.gov/sra/?term=SRR1702360) (*Thermococcus onnurineus* NA1); <https://www.ncbi.nlm.nih.gov/sra/?term=SRR1702361> (*Thermococcus onnurineus* NA1).

Author contributions

RS performed bioinformatic analysis of the data, with contributions from FG. LN performed analysis of swimming behavior and TEM imaging. RRe performed construction of vectors and genetic manipulation of *P. furiosus* cells. RRe, RRa, and DG supervised this work. DG acquired funding to conduct the work. FG prepared the figures, with contribution of all authors. RS, RRe, DG, and FG wrote the manuscript. All authors contributed to manuscript revision and approval of the submitted version.

Funding

Work in the Grohmann lab was supported by the Deutsche Forschungsgemeinschaft (DFG funding scheme SFB960 TPA7 to DG).

Acknowledgments

We thank all the members of the Grohmann lab for fruitful discussions. Also, we would like to thank all current and previous members of the “archaeal transcription group” (especially Winfried Hausner, Zubeir El Ahmad, and Martin Fenk) for establishment of the genetic system and sequencing protocols and Annett Bellack for providing antibodies. This manuscript has previously appeared online in a preprint, which can be accessed through the DOI: <https://doi.org/10.1101/2023.06.15.545098>.

Conflict of interest

DG is co-founder of Microbify GmbH. However, there are no commercial interests by the company, or any financial support granted by Microbify GmbH.

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.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2023.1241399/full#supplementary-material>

References

- Albers, S.-V., and Jarrell, K. F. (2015). The archaeellum: how archaea swim. *Front. Microbiol.* 6:23. doi: 10.3389/fmicb.2015.00023
- 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
- Bailey, T. L., Johnson, J., Grant, C. E., and Noble, W. S. (2015). The MEME suite. *Nucleic Acids Res.* 43, W39–W49. doi: 10.1093/nar/gkv416
- Beeby, M., Ferreira, J. L., Tripp, P., Albers, S.-V., and Mitchell, D. R. (2020). Propulsive nanomachines: the convergent evolution of archaea, flagella and cilia. *FEMS Microbiol. Rev.* 44, 253–304. doi: 10.1093/femsre/fuaa006
- Briegleb, A., Oikonomou, C. M., Chang, Y., Kjær, A., Huang, A. N., Kim, K. W., et al. (2017). Morphology of the archaeal motor and associated cytoplasmic cone in *Thermococcus kodakarensis*. *EMBO Rep.* 18, 1660–1670. doi: 10.15252/embr.201744070
- Canganella, F., Jones, W. J., Gambacorta, A., and Antranikian, G. (1998). *Thermococcus guaymasensis* sp. nov. and *Thermococcus aggregans* sp. nov., two novel thermophilic archaea isolated from the Guaymas Basin hydrothermal vent site. *Int. J. Syst. Bacteriol.* 48, 1181–1185. doi: 10.1099/00207713-48-4-1181
- Chaban, B., Ng, S. Y. M., Kanbe, M., Saltzman, I., Nimmo, G., Aizawa, S.-I., et al. (2007). Systematic deletion analyses of the flagella operon identify several genes essential for proper assembly and function of flagella in the archaeon, *Methanococcus maripaludis*. *Mol. Microbiol.* 66, 596–609. doi: 10.1111/j.1365-2958.2007.05913.x
- Chaudhury, P., Van Der Does, C., and Albers, S.-V. (2018). Characterization of the ATPase FlaI of the motor complex of the *Pyrococcus furiosus* archaeellum and its interactions between the ATP-binding protein FlaH. *PeerJ* 6:e4984. doi: 10.7717/peerj.4984
- Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34, i884–i890. doi: 10.1093/bioinformatics/bty560
- Daum, B., Vonck, J., Bellack, A., Chaudhury, P., Reichelt, R., Albers, S.-V., et al. (2017). Structure and in situ organisation of the *Pyrococcus furiosus* archaeellum machinery. *Elife* 6:e27470. doi: 10.7554/eLife.27470
- Desmond, E., Brochier-Armanet, C., and Gribaldo, S. (2007). Phylogenomics of the archaeal flagellum: rare horizontal gene transfer in a unique motility structure. *BMC Evol. Biol.* 7:106. doi: 10.1186/1471-2148-7-106
- Ding, Y., Berezuk, A., Khursigara, C. M., and Jarrell, K. F. (2017a). Bypassing the need for the transcriptional activator EarA through a spontaneous deletion in the BRE portion of the fla operon promoter in *Methanococcus maripaludis*. *Front. Microbiol.* 8:1329. doi: 10.3389/fmicb.2017.01329
- Ding, Y., Berezuk, A., Khursigara, C. M., and Jarrell, K. F. (2017b). Phylogenetic distribution of the euryarchaeal archaeellum regulator EarA and complementation of a *Methanococcus maripaludis* ΔearA mutant with heterologous earA homologues. *Microbiology* 163, 804–815. doi: 10.1099/mic.0.000464
- Ding, Y., Nash, J., Berezuk, A., Khursigara, C. M., Langelaan, D. N., Smith, S. P., et al. (2016). Identification of the first transcriptional activator of an archaeellum operon in a euryarchaeon. *Mol. Microbiol.* 102, 54–70. doi: 10.1111/mmi.13444
- Dombrowski, N., Williams, T. A., Sun, J., Woodcroft, B. J., Lee, J.-H., Minh, B. Q., et al. (2020). Undinarchaeota illuminate DPANN phylogeny and the impact of gene transfer on archaeal evolution. *Nat. Commun.* 11:3939. doi: 10.1038/s41467-020-17408-w
- Fiala, G., and Stetter, K. O. (1986). *Pyrococcus furiosus* sp. nov. represents a novel genus of marine heterotrophic archaeobacteria growing optimally at 100°C. *Arch. Microbiol.* 145, 56–61. doi: 10.1007/BF00413027
- Gambelli, L., Isupov, M. N., Connors, R., McLaren, M., Bellack, A., Gold, V., et al. (2022). An archaeellum filament composed of two alternating subunits. *Nat. Commun.* 13:710. doi: 10.1038/s41467-022-28337-1
- Gibson, D. G., Young, L., Chuang, R.-Y., Venter, J. C., Hutchison, C. A., and Smith, H. O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* 6, 343–345. doi: 10.1038/nmeth.1318
- Grote, R., Li, L., Tamaoka, J., Kato, C., Horikoshi, K., and Antranikian, G. (1999). *Thermococcus siculi* sp. nov., a novel hyperthermophilic archaeon isolated from a deep-sea hydrothermal vent at the Mid-Okinawa trough. *Extremophiles* 3, 55–62. doi: 10.1007/s007920050099
- Grünberger, F., Reichelt, R., Bunk, B., Spröer, C., Overmann, J., Rachel, R., et al. (2019). Next generation DNA-Seq and differential RNA-Seq allow re-annotation of the *Pyrococcus furiosus* DSM 3638 genome and provide insights into archaeal antisense transcription. *Front. Microbiol.* 10:1603. doi: 10.3389/fmicb.2019.01603
- Grünberger, F., Reichelt, R., Waage, I., Ned, V., Bronner, K., Kaljanac, M., et al. (2021). CopR, a global regulator of transcription to maintain copper homeostasis in *Pyrococcus furiosus*. *Front. Microbiol.* 11:613532. doi: 10.3389/fmicb.2020.613532
- Grünberger, F., Schmid, G., El Ahmad, Z., Fenk, M., Vogl, K., Reichelt, R., et al. (2023). Uncovering the temporal dynamics and regulatory networks of thermal stress response in a hyperthermophile using transcriptomics and proteomics. *mBio*. e02174–23. doi: 10.1128/mbio.02174-23
- Haurat, M. F., Figueiredo, A. S., Hoffmann, L., Li, L., Herr, K., Wilson, A. J., et al. (2017). ArnS, a kinase involved in starvation-induced archaeellum expression: starvation-induced archaeellum expression. *Mol. Microbiol.* 103, 181–194. doi: 10.1111/mmi.13550
- Hensley, S. A., Jung, J.-H., Park, C.-S., and Holden, J. F. (2014). *Thermococcus paralvinellae* sp. nov. and *Thermococcus cleffensis* sp. nov. of hyperthermophilic heterotrophs from deep-sea hydrothermal vents. *Int. J. Syst. Evol. Microbiol.* 64, 3655–3659. doi: 10.1099/ijss.0.066100-0
- Herzog, B., and Wirth, R. (2012). Swimming behavior of selected species of Archaea. *Appl. Environ. Microbiol.* 78, 1670–1674. doi: 10.1128/AEM.06723-11
- Hoffmann, L., Schummer, A., Reimann, J., Haurat, M. F., Wilson, A. J., Beeby, M., et al. (2017). Expanding the archaeellum regulatory network - the eukaryotic protein kinases ArnC and ArnD influence motility of *Sulfolobus acidocaldarius*. *Microbiol. Open* 6:e00414. doi: 10.1002/mbo3.414
- Jarrell, K. F., Albers, S.-V., and Machado, J. N. D. S. (2021). A comprehensive history of motility and archaeellum in archaea. *FEMS Microbes* 2:xtab002. doi: 10.1093/femsmc/xtab002
- Jarrell, K. F., Ding, Y., Meyer, B. H., Albers, S.-V., Kaminski, L., and Eichler, J. (2014). N-linked glycosylation in Archaea: a structural, functional, and genetic analysis. *Microbiol. Mol. Biol. Rev.* 78, 304–341. doi: 10.1128/MMBR.00052-13
- Khan, S., and Schöley, J. M. (2018). Assembly, functions and evolution of archaea, flagella and cilia. *Curr. Biol.* 28, R278–R292. doi: 10.1016/j.cub.2018.01.085
- Kreuzer, M., Schmutzler, K., Waage, I., Thomm, M., and Hausner, W. (2013). Genetic engineering of *Pyrococcus furiosus* to use chitin as a carbon source. *BMC Biotechnol.* 13:19. doi: 10.1186/1472-6750-13-9
- Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923
- Lassak, K., Neiner, T., Ghosh, A., Klingl, A., Wirth, R., and Albers, S.-V. (2012). Molecular analysis of the crenarchaeal flagellum. *Mol. Microbiol.* 83, 110–124. doi: 10.1111/j.1365-2958.2011.07916.x
- Lassak, K., Peeters, E., Wróbel, S., and Albers, S.-V. (2013). The one-component system ArnR: a membrane-bound activator of the crenarchaeal archaeellum: ArnA activates archaea expression. *Mol. Microbiol.* 88, 125–139. doi: 10.1111/mmi.12173
- Lee, S., Cook, D., and Lawrence, M. (2019). plyranges: a grammar of genomic data transformation. *Genome Biol.* 20:4. doi: 10.1186/s13059-018-1597-8
- Lee, S. H., Kim, M.-S., Jung, H. C., Lee, J., Lee, J.-H., Lee, H. S., et al. (2015). Screening of a novel strong promoter by RNA sequencing and its application to H₂ production in a hyperthermophilic archaeon. *Appl. Microbiol. Biotechnol.* 99, 4085–4092. doi: 10.1007/s00253-015-6444-1
- Lewis, D. L., Notey, J. S., Chandrayan, S. K., Loder, A. J., Lipscomb, G. L., Adams, M. W. W., et al. (2015). A mutant ('lab strain') of the hyperthermophilic archaeon *Pyrococcus furiosus*, lacking flagella, has unusual growth physiology. *Extremophiles* 19, 269–281. doi: 10.1007/s00792-014-0712-3
- Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34, 3094–3100. doi: 10.1093/bioinformatics/bty191
- Li, H. (2021). New strategies to improve minimap2 alignment accuracy. *Bioinformatics* 37, 4572–4574. doi: 10.1093/bioinformatics/btab705
- Li, L., Banerjee, A., Bischof, L. F., Maklad, H. R., Hoffmann, L., Henche, A.-L., et al. (2017). Wing phosphorylation is a major functional determinant of the Lrs14-type biofilm and motility regulator AbfR1 in *Sulfolobus acidocaldarius*. *Mol. Microbiol.* 105, 777–793. doi: 10.1111/mmi.13735
- 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
- Li, X., Ma, L., Mei, X., Liu, Y., and Huang, H. (2022). gg motif: an R package for the extraction and visualization of motifs from MEME software. *PLoS One* 17:e0276979. doi: 10.1371/journal.pone.0276979
- Li, J., Yue, L., Li, Z., Zhang, W., Zhang, B., Zhao, F., et al. (2021). aCPSF1 cooperates with terminator U-tract to dictate archaeal transcription termination efficacy. *Elife* 10:e70464. doi: 10.7554/eLife.70464
- Liang, R., Robb, F. T., and Onstott, T. C. (2021). Aspartic acid racemization and repair in the survival and recovery of hyperthermophiles after prolonged starvation at high temperature. *FEMS Microbiol. Ecol.* 97:fiab112. doi: 10.1093/femsec/fiab112
- Liao, Y., Smyth, G. K., and Shi, W. (2019). The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads. *Nucleic Acids Res.* 47:e47. doi: 10.1093/nar/gkz114
- Maier, L. K., and Marchfelder, A. (2019). It's all about the T: transcription termination in Archaea. *Biochem. Soc. Trans.* 47, 461–468. doi: 10.1042/BST20180557
- Makarova, K. S., Wolf, Y. I., and Koonin, E. V. (2015). Archaeal clusters of orthologous genes (arCOGs): an update and application for analysis of shared features between *Thermococcales*, *Methanococcales*, and *Methanobacteriales*. *Life* 5, 818–840. doi: 10.3390/life5010818
- Mao, X., Ma, Q., Zhou, C., Chen, X., Zhang, H., Yang, J., et al. (2014). DOOR 2.0: presenting operons and their functions through dynamic and integrated views. *Nucleic Acids Res.* 42, D654–D659. doi: 10.1093/nar/gkt1048
- Micorescu, M., Grünberg, S., Franke, A., Cramer, P., Thomm, M., and Bartlett, M. (2008). Archaeal transcription: function of an alternative transcription factor B from *Pyrococcus furiosus*. *J. Bacteriol.* 190, 157–167. doi: 10.1128/jb.01498-07

- Mora, M., Bellack, A., Ugele, M., Hopf, J., and Wirth, R. (2014). The temperature gradient-forming device, an accessory unit for normal light microscopes to study the biology of hyperthermophilic microorganisms. *Appl. Environ. Microbiol.* 80, 4764–4770. doi: 10.1128/AEM.00984-14
- Nagahisa, K., Ezaki, S., Fujiwara, S., Imanaka, T., and Takagi, M. (1999). Sequence and transcriptional studies of five clustered flagellin genes from hyperthermophilic archaeon *Pyrococcus kodakaraensis* KOD1. *FEMS Microbiol. Lett.* 178, 183–190. doi: 10.1111/j.1574-6968.1999.tb13776.x
- Näther, D. J., Rachel, R., Wanner, G., and Wirth, R. (2006). Flagella of *Pyrococcus furiosus*: multifunctional organelles, made for swimming, adhesion to various surfaces, and cell-cell contacts. *J. Bacteriol.* 188, 6915–6923. doi: 10.1128/JB.00527-06
- Näther-Schindler, D. J., Schopf, S., Bellack, A., Rachel, R., and Wirth, R. (2014). *Pyrococcus furiosus* flagella: biochemical and transcriptional analyses identify the newly detected flaB0 gene to encode the major flagellin. *Front. Microbiol.* 5:695. doi: 10.3389/fmicb.2014.00695
- Nuno de Sousa Machado, J., Albers, S.-V., and Daum, B. (2022). Towards elucidating the rotary mechanism of the archaeum machinery. *Front. Microbiol.* 13:848597. doi: 10.3389/fmicb.2022.848597
- Ochs, S. M., Thumann, S., Richau, R., Weirauch, M. T., Lowe, T. M., Thomm, M., et al. (2012). Activation of archaeal transcription mediated by recruitment of transcription factor B. *J. Biol. Chem.* 287, 18863–18871. doi: 10.1074/jbc.M112.365742
- Oren, A., and Garrity, G. M. (2021). Valid publication of the names of forty-two phyla of prokaryotes. *Int. J. Syst. Evol. Microbiol.* 71. doi: 10.1099/ijsem.0.005056
- Ouhammouch, M., Langham, G. E., Hausner, W., Simpson, A. J., El-Sayed, N. M. A., and Geiduschek, E. P. (2005). Promoter architecture and response to a positive regulator of archaeal transcription: architecture of an archaeal UAS. *Mol. Microbiol.* 56, 625–637. doi: 10.1111/j.1365-2958.2005.04563.x
- Poweleit, N., Ge, P., Nguyen, H. H., Loo, R. R. O., Gunsalus, R. P., and Zhou, Z. H. (2016). CryoEM structure of the *Methanospirillum hungatei* archaeum reveals structural features distinct from the bacterial flagellum and type IV pilus. *Nat. Microbiol.* 2:16222. doi: 10.1038/nmicrobiol.2016.222
- Pyatibratov, M. G., Syutkin, A. S., Quax, T. E. F., Melnik, T. N., Papke, R. T., Gogarten, J. P., et al. (2020). Interaction of two strongly divergent archaealins stabilizes the structure of the *Haloarubrum archaeum*. *Microbiol. Open* 9:e1047. doi: 10.1002/mbo3.1047
- 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
- Reichelt, R., Ruperti, K. M. A., Kreuzer, M., Dextl, S., Thomm, M., and Hausner, W. (2018). The transcriptional regulator TFB-RF1 activates transcription of a putative ABC transporter in *Pyrococcus furiosus*. *Front. Microbiol.* 9:838. doi: 10.3389/fmicb.2018.00838
- Reimann, J., Lassak, K., Khadouma, S., Ettema, T. J. G., Yang, N., Driessen, A. J. M., et al. (2012). Regulation of archaeal expression by the FHA and von Willebrand domain-containing proteins ArnA and ArnB in *Sulfolobus acidocaldarius*: ArnA and ArnB regulate archaeal expression in *S. acidocaldarius*. *Mol. Microbiol.* 86, 24–36. doi: 10.1111/j.1365-2958.2012.08186.x
- Sanders, T. J., Wenck, B. R., Selan, J. N., Barker, M. P., Trimmer, S. A., Walker, J. E., et al. (2020). FttA is a CPSF73 homologue that terminates transcription in Archaea. *Nat. Microbiol.* 5, 545–553. doi: 10.1038/s41564-020-0667-3
- Schlesner, M., Miller, A., Streif, S., Staudinger, W. F., Müller, J., Scheffer, B., et al. (2009). Identification of Archaea-specific chemotaxis proteins which interact with the flagellar apparatus. *BMC Microbiol.* 9:56. doi: 10.1186/1471-2180-9-56
- Schopf, S., Wanner, G., Rachel, R., and Wirth, R. (2008). An archaeal bi-species biofilm formed by *Pyrococcus furiosus* and *Methanopyrus kandleri*. *Arch. Microbiol.* 190, 371–377. doi: 10.1007/s00203-008-0371-9
- Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30, 2068–2069. doi: 10.1093/bioinformatics/btu153
- Shen, W., Le, S., Li, Y., and Hu, F. (2016). SeqKit: a cross-platform and ultrafast toolkit for FASTA/Q file manipulation. *PLoS One* 11:e0163962. doi: 10.1371/journal.pone.0163962
- Sievers, F., and Higgins, D. G. (2014). Clustal Omega. *Curr. Protoc. Bioinforma.* 48, 3.13.1–3.13.16. doi: 10.1002/0471250953.bi0313s48
- Thomas, N. A., and Jarrell, K. F. (2001). Characterization of flagellum gene families of methanogenic archaea and localization of novel flagellum accessory proteins. *J. Bacteriol.* 183, 7154–7164. doi: 10.1128/JB.183.24.7154-7164.2001
- Villain, P., da Cunha, V., Villain, E., Forterre, P., Oberto, J., Catchpole, R., et al. (2021). The hyperthermophilic archaeon *Thermococcus kodakarensis* is resistant to pervasive negative supercoiling activity of DNA gyrase. *Nucleic Acids Res.* 49, 12332–12347. doi: 10.1093/nar/gkab869
- Waage, I., Schmid, G., Thumann, S., Thomm, M., and Hausner, W. (2010). Shuttle vector-based transformation system for *Pyrococcus furiosus*. *Appl. Environ. Microbiol.* 76, 3308–3313. doi: 10.1128/AEM.01951-09
- Wagih, O. (2017). ggseqlogo: a versatile R package for drawing sequence logos. *Bioinformatics* 33, 3645–3647. doi: 10.1093/bioinformatics/btx469
- Weiner, A., Schopf, S., Wanner, G., Probst, A., and Wirth, R. (2012). Positive, neutral and negative interactions in cocultures between *Pyrococcus furiosus* and different methanogenic Archaea. *Microbiol. Insights* 5, 1–10. doi: 10.4137/MBI.S8516
- Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., et al. (2019). Welcome to the Tidyverse. *J. Open Source Softw.* 4:1686. doi: 10.21105/joss.01686
- Wirth, R. (2017). Colonization of black smokers by hyperthermophilic microorganisms. *Trends Microbiol.* 25, 92–99. doi: 10.1016/j.tim.2016.11.002
- Wirth, R., Luckner, M., and Wanner, G. (2018). Validation of a hypothesis: colonization of black smokers by hyperthermophilic microorganisms. *Front. Microbiol.* 9:524. doi: 10.3389/fmicb.2018.00524
- Yue, L., Li, J., Zhang, B., Qi, L., Li, Z., Zhao, F., et al. (2020). The conserved ribonuclease aCPSF1 triggers genome-wide transcription termination of Archaea via a 3'-end cleavage mode. *Nucleic Acids Res.* 48, 9589–9605. doi: 10.1093/nar/gkaa702
- Zeng, X., Zhang, X., and Shao, Z. (2020). Metabolic adaptation to sulfur of hyperthermophilic palaeococcus pacificus DY20341T from Deep-Sea hydrothermal sediments. *Int. J. Mol. Sci.* 21:368. doi: 10.3390/ijms21010368
- Zhou, L., Feng, T., Xu, S., Gao, F., Lam, T. T., Wang, Q., et al. (2022). ggmsa: a visual exploration tool for multiple sequence alignment and associated data. *Brief. Bioinform.* 23:bbac222. doi: 10.1093/bib/bbac222



OPEN ACCESS

EDITED BY

Michel Geovanni Santiago-Martinez,
University of Connecticut, United States

REVIEWED BY

Nahui Olin Medina-Chavez,
University of Minnesota Twin Cities,
United States
Marika Palmer,
University of Nevada, Las Vegas, United States

*CORRESPONDENCE

Andreas Brune
✉ brune@mpi-marburg.mpg.de

†PRESENT ADDRESS

James O. Nonoh,
Department of Biomedical Sciences
and Technology, School of Public Health,
Maseno University, Kisumu, Kenya

‡These authors share first authorship

RECEIVED 22 August 2023

ACCEPTED 13 October 2023

PUBLISHED 15 November 2023

CITATION

Protasov E, Nonoh JO, Kästle Silva JM,
Mies US, Hervé V, Dietrich C, Lang K,
Mikulski L, Platt K, Poehlein A, Köhler-Ramm T,
Miambi E, Boga HI, Feldewert C, Ngugi DK,
Plarre R, Sillam-Dussès D, Šobotník J, Daniel R
and Brune A (2023) Diversity and taxonomic
revision of methanogens and other archaea
in the intestinal tract of terrestrial arthropods.
Front. Microbiol. 14:1281628.
doi: 10.3389/fmicb.2023.1281628

COPYRIGHT

© 2023 Protasov, Nonoh, Kästle Silva, Mies,
Hervé, Dietrich, Lang, Mikulski, Platt, Poehlein,
Köhler-Ramm, Miambi, Boga, Feldewert, Ngugi,
Plarre, Sillam-Dussès, Šobotník, Daniel and
Brune. 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.

Diversity and taxonomic revision of methanogens and other archaea in the intestinal tract of terrestrial arthropods

Evgenii Protasov^{1†}, James O. Nonoh^{1†}, Joana M. Kästle Silva^{1†},
Undine S. Mies¹, Vincent Hervé¹, Carsten Dietrich¹,
Kristina Lang¹, Lena Mikulski¹, Katja Platt¹, Anja Poehlein²,
Tim Köhler-Ramm¹, Edouard Miambi³, Hamadi I. Boga¹,
Christopher Feldewert¹, David K. Ngugi¹, Rudy Plarre⁴,
David Sillam-Dussès⁵, Jan Šobotník⁶, Rolf Daniel² and
Andreas Brune^{1*}

¹Research Group Insect Gut Microbiology and Symbiosis, Max Planck Institute for Terrestrial Microbiology, Marburg, Germany, ²Genomic and Applied Microbiology and Göttingen Genomics Laboratory, Institute of Microbiology and Genetics, Georg August University of Göttingen, Göttingen, Germany, ³Evolutionary Ecology Department, Institute of Ecology and Environmental Sciences of Paris (iEES-Paris), University of Paris-Est Créteil (UPEC), Créteil, France, ⁴Bundesanstalt für Materialforschung und -prüfung, Berlin, Germany, ⁵Laboratory of Experimental and Comparative Ethology (LEEC), UR 4443, Université Sorbonne Paris Nord, Villetaneuse, France, ⁶Faculty of Tropical AgriSciences, Czech University of Life Sciences, Prague, Czechia

Methane emission by terrestrial invertebrates is restricted to millipedes, termites, cockroaches, and scarab beetles. The arthropod-associated archaea known to date belong to the orders *Methanobacteriales*, *Methanomassiliicoccales*, *Methanomicrobiales*, and *Methanosarcinales*, and in a few cases also to non-methanogenic *Nitrososphaerales* and *Bathyarchaeales*. However, all major host groups are severely undersampled, and the taxonomy of existing lineages is not well developed. Full-length 16S rRNA gene sequences and genomes of arthropod-associated archaea are scarce, reference databases lack resolution, and the names of many taxa are either not validly published or under-classified and require revision. Here, we investigated the diversity of archaea in a wide range of methane-emitting arthropods, combining phylogenomic analysis of isolates and metagenome-assembled genomes (MAGs) with amplicon sequencing of full-length 16S rRNA genes. Our results allowed us to describe numerous new species in hitherto undescribed taxa among the orders *Methanobacteriales* (*Methanacia*, *Methanarmilla*, *Methanobaculum*, *Methanobinarius*, *Methanocatella*, *Methanoflexus*, *Methanorudis*, and *Methanovirga*, all gen. nova), *Methanomicrobiales* (*Methanofilum* and *Methanorbis*, both gen. nova), *Methanosarcinales* (*Methanofrustulum* and *Methanolapillus*, both gen. nova), *Methanomassiliicoccales* (*Methanomethylophilaceae* fam. nov., *Methanarcanum*, *Methanogranum*, *Methanomethylophilus*, *Methanomicula*, *Methanoplasma*, *Methanoprismaticola*, all gen. nova), and the new family *Bathycorpusculaceae* (*Bathycorpusculum* gen. nov.). Reclassification of amplicon libraries from this and previous studies using this new taxonomic framework revealed that arthropods harbor only CO₂ and methyl-reducing hydrogenotrophic methanogens. Numerous genus-level lineages appear to be present exclusively in arthropods,

suggesting long evolutionary trajectories with their termite, cockroach, and millipede hosts, and a radiation into various microhabitats and ecological niches provided by their digestive tracts (e.g., hindgut compartments, gut wall, or anaerobic protists). The distribution patterns among the different host groups are often complex, indicating a mixed mode of transmission and a parallel evolution of invertebrate and vertebrate-associated lineages.

KEYWORDS

archaea, methanogens, gut microbiota, termites, cockroaches, millipedes, Bathyarchaeia, Nitrososphaerales

Introduction

Methanogenic archaea play an important role in the fermentative breakdown of organic matter (Müller et al., 2018). They are common constituents of the intestinal microbiota of both invertebrate and vertebrate animals, where they thrive on the products of bacterial fermentations, namely molecular hydrogen, formate, methanol, and methylamines (Brune, 2019; Chibani et al., 2022).

Methane emission by termites was documented half a century ago by the seminal work of Breznak and coworkers (Brune, 2019). Although the phenomenon attracted attention because of its implications for the global methane budget, methane emissions from termites are dwarfed by those from ruminants and wetlands. Subsequent surveys of other invertebrates revealed that methanogenesis is restricted to only a few distinct groups of terrestrial arthropods, namely millipedes, termites and cockroaches (Blattodea), and scarab beetles (Hackstein and Stumm, 1994; Hackstein and van Alen, 2018; Brune, 2019).

Methanogens in arthropod guts are typically restricted to specific hindgut compartments, where they are localized on the cuticular lining, attached to filamentous bacteria on the hindgut wall, and associated with anaerobic protists (ciliates in cockroaches and millipedes; flagellates in all termite families except Termitidae or higher termites) (e.g., Leadbetter and Breznak, 1996; Sprenger et al., 2000). They consist almost exclusively of uncultured representatives, which have been identified in 16S rRNA-based surveys as members of the orders *Methanobacteriales* (phylum *Methanobacteriota*), *Methanomicrobiales* and *Methanosarcinales* (both phylum “*Halobacteriota*”), and *Methanomassiliicoccales* (phylum “*Thermoplasmatota*”); only a few species of the genera *Methanobrevibacter* and *Methanimicrococcus* have been isolated in pure culture (see reviews by Brune, 2018, 2019; Hackstein and van Alen, 2018). Some studies also identified non-methanogenic *Bathyarchaeales* and *Nitrososphaerales* (both phylum *Thermoproteota*) (e.g., Friedrich et al., 2001; Loh et al., 2021).

Despite these efforts, all major host groups are severely undersampled, and the diversity of methanogens in arthropods remains poorly resolved. The reference databases lack resolution because both full-length 16S rRNA gene sequences and genomes of arthropod-associated archaea are scarce. Also, the taxonomy of existing lineages is not well developed, and the names of many

taxa are provisional and not validly published, while other taxa are under-classified and require revision (Rinke et al., 2021).

To address these issues, we conducted a phylogenomic analysis of all archaeal genomes from arthropods using the taxonomy framework of the Genome Taxonomy Database (Parks et al., 2021), including a large number of metagenome-assembled genomes (MAGs) from termite guts (85 MAGs from 34 termite species) and diverse, so far undescribed, isolates obtained in our laboratory from cockroaches and millipedes. In parallel, we prepared full-length 16S rRNA gene libraries from more than 70 species of methane-emitting arthropods and incorporated them into the alignment of the SILVA database (version 138), together with full-length sequences from the termite gut metagenomes and unpublished clone libraries from our laboratory. Based on this comprehensive reference database, we reconstructed phylogenetic trees for all archaeal lineages in arthropod guts. In order to revise the taxonomy of the respective lineages, including a number of provisional *Candidatus* taxa from the literature, we then linked the lineages in the respective phylogenies via the 16S rRNA genes in the genomes, which allowed us to describe new species and higher taxa under the *Code of Nomenclature of Prokaryotes Described from Sequence Data* (SeqCode) (Hedlund et al., 2022; Whitman et al., 2022). Finally, we reclassified the archaeal 16S rRNA-gene libraries from arthropods guts from this study and selected datasets from the literature to provide an overview of the distribution of archaeal lineages across all host groups at the genus level.

Results

High-throughput sequencing of long-read amplicon libraries of archaeal 16S rRNA genes from the intestinal tract of cockroaches, termites, and millipedes (47 species, **Supplementary Tables 1, 4**) and hitherto unpublished clone libraries of archaeal 16S rRNA genes from termites and millipedes (15 species, **Supplementary Tables 1, 5**) substantially expanded literature information on archaeal diversity in methane-emitting arthropods (18 species, **Supplementary Table 5**). Phylogenetic analysis revealed that the archaeal communities consist mostly of methanogenic archaea of the orders *Methanobacteriales*, *Methanomicrobiales*, *Methanosarcinales*, and *Methanomassiliicoccales*. In a few species of soil-feeding termites and litter-feeding millipedes, the archaeal

communities comprised also non-methanogenic *Bathyarchaeales* and *Nitrososphaerales*.

Phylogenomic analysis of 85 archaeal MAGs from gut metagenomes of 34 termite species and 9 genomes of methanogens isolated from 5 millipede and 3 cockroach species revealed that almost all genus-level lineages occurring in arthropods were represented by one or more high-quality genomes (Figure 1 and Supplementary Table 2). Since the relative evolutionary divergence (RED) of several genera considerably exceeded the average values of other genus-level lineages in the respective phyla (Parks et al., 2018; Rinke et al., 2021), we harmonized the taxonomic ranks of the respective lineages by introducing additional genus-level taxa (Figure 2).

Using the 16S rRNA genes from the genomic datasets, it was possible to link most clades in the 16S rRNA-based trees to this new taxonomic framework (Figures 3–7). In each order except Nitrososphaerales, the sequences from arthropod guts typically formed one or more lineages that comprised only members of a particular host group, often without cultured representatives. Most lineages had representatives with sequenced genomes of sufficient quality to serve as nomenclatural type for the description of new species under SeqCode (see section “Taxonomy”).

Methanobacteriales

Members of the order *Methanobacteriales* are the most common archaeal lineage in the intestinal tract of arthropods. All phylotypes fall within the radiation of the genus *Methanobrevibacter* sensu lato (Figure 3). Based on the RED values among members of this genus (Rinke et al., 2021), the current Genome Taxonomy Database (GTDB) distinguishes between *Methanobrevibacter* sensu stricto (which contains the type species, *Methanobrevibacter ruminantium*) and four additional genus-level lineages, *Methanobrevibacter_A* to *Methanobrevibacter_D* (hereafter Mbb_A–D). All sequences from arthropod guts fall into the radiation of Mbb_A, Mbb_C, and Mbb_D; the lineages *Methanobrevibacter* sensu stricto and Mbb_B are not represented in arthropods.

In the phylogenomic analysis, the new MAGs from termite guts further expanded the evolutionary divergence within the radiation of Mbb_C and Mbb_D, resulting in RED values for the internal nodes that require the introduction of additional genus-level taxa (Figure 2A). In accordance with the taxonomic ranks suggested by Rinke et al. (2021), we propose that the *Methanobrevibacter* species that do not fall within the radiation of *Methanobrevibacter* sensu stricto be placed in the new genera *Methanocatella* (Mbb_A), *Methanarmilla* (Mbb_B), *Methanobaculum*, *Methanobinarius*, and *Methanorudis* (Mbb_C), and *Methanacia*, *Methanoflexus*, and *Methanovirga* (Mbb_D), using the genomes of previously described species and uncultured archaea as nomenclatural type (see section “Taxonomy”).

The genera *Methanocatella* and *Methanarmilla* consist exclusively of isolates or uncultured archaea from the intestinal tract of mammals (Figure 3). Mbb_A comprises a large clade of 16S rRNA sequences from cockroaches and a few millipedes (Mbb_A2) and a smaller clade (Mbb_A1) of sequences from lower termites (*Reticulitermes flavipes* and *Hodotermopsis sjoestedti*),

which are well separated from the genus *Methanocatella* but lack representatives with high- or medium-quality genomes. In the phylogenomic analysis, we identified a single low-quality genome (Hm464_bin.79) from the lower termite *Hodotermes mossambicus* that occupies a sister position to the genus *Methanocatella*, suggesting that the Mbb_A clade comprises additional genus-level taxa from arthropod guts (Figure 2B).

The remaining genera consist almost exclusively of representatives from the guts of termites, cockroaches, and millipedes (Figure 3). While the genera *Methanacia*, *Methanobaculum*, *Methanobinarius*, and *Methanoflexus* have cultured representatives, the genera *Methanorudis* and *Methanovirga* consist exclusively of uncultured archaea. Two clades in the radiation of Mbb_D that consist exclusively of clones from cockroaches (Mbb_D2) and higher termites (Mbb_D1) lack representatives with sequenced genomes. A few clones from the genus *Methanobinarius* were not obtained from arthropods guts but were recovered from a sapropelic ciliate or anaerobic bioreactors.

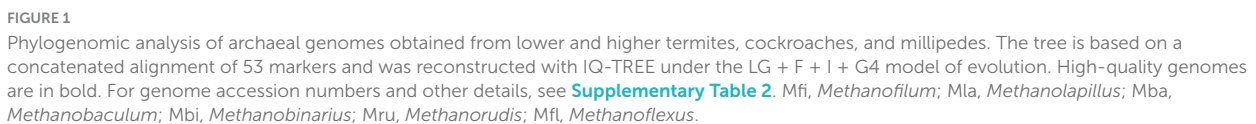
Methanomicrobiales

Representatives of the order *Methanomicrobiales* form several arthropod-specific clusters in the families *Methanospirillaceae* and *Methanocorpusculaceae* (Figure 4). The clones that fall into the radiation of *Methanospirillaceae* form a genus-level lineage that is sister to the genus *Methanospirillum* (Figure 4). The clade consists exclusively of uncultured methanogens from the intestinal tract of higher termites and several cockroaches. Since the MAGs from termite guts form a well-separated genus-level clade (WRER01 in GTDB) also in the phylogenomic tree (Figure 2B), we propose to classify them in the new genus *Methanofilum* (see section “Taxonomy”).

The clones that fall into the radiation of *Methanocorpusculaceae* form several lineages that occupy basal positions to the genus *Methanocorpusculum*. One lineage consists exclusively of sequences from millipedes, including three isolates from our laboratory (Protasov and Brune, unpublished results). It is loosely affiliated with additional lineages of uncultured representatives from cockroaches, millipedes, and termites. Members of the genus *Methanocorpusculum* form a well-supported cluster with a lineage of uncultured archaea from mammalian feces. In the phylogenomic analysis, however, only the genomes from mammalian feces fall into the genus *Methanocorpusculum*, whereas the MAGs from termites and the genomes of millipede isolates form a separate genus-level clade that also includes MAGs from wombat and chicken feces (Figure 2B). We propose to classify the members of this clade in the new genus *Methanorbis* (see section “Taxonomy”).

Methanosarcinales

In the order *Methanosarcinales*, most sequences from arthropods guts fall into two genus-level clusters in the family *Methanosarcinaceae* (Figure 5). One of the clusters contains all representatives from termites and cockroaches, including *Methanimicrococcus blatticola* isolated from the cockroach



The two clusters are sister to a clade of uncultured archaea from the rumen or feces of mammals, including endosymbionts of rumen ciliates. In the phylogenomic analysis (**Figure 2C**), the clade is represented by several genomes from ruminants and anaerobic

With a few exceptions, the *Methanomassiliicoccales* from arthropod guts are representatives of the so-called intestinal clade, a family-level cluster that comprises several highly enriched cultures but no isolates. Phylogenetic analysis revealed the presence of several arthropod-specific lineages that are well separated from lineages found in the mammalian guts or anaerobic digesters (Figure 6). One of these lineages comprises numerous representatives from the guts of termites, cockroaches, and millipedes, including the previously characterized *Candidatus* Methanoplasma termitum (Lang et al., 2015). Based on several genomes from lower and higher termites that form a well-separated

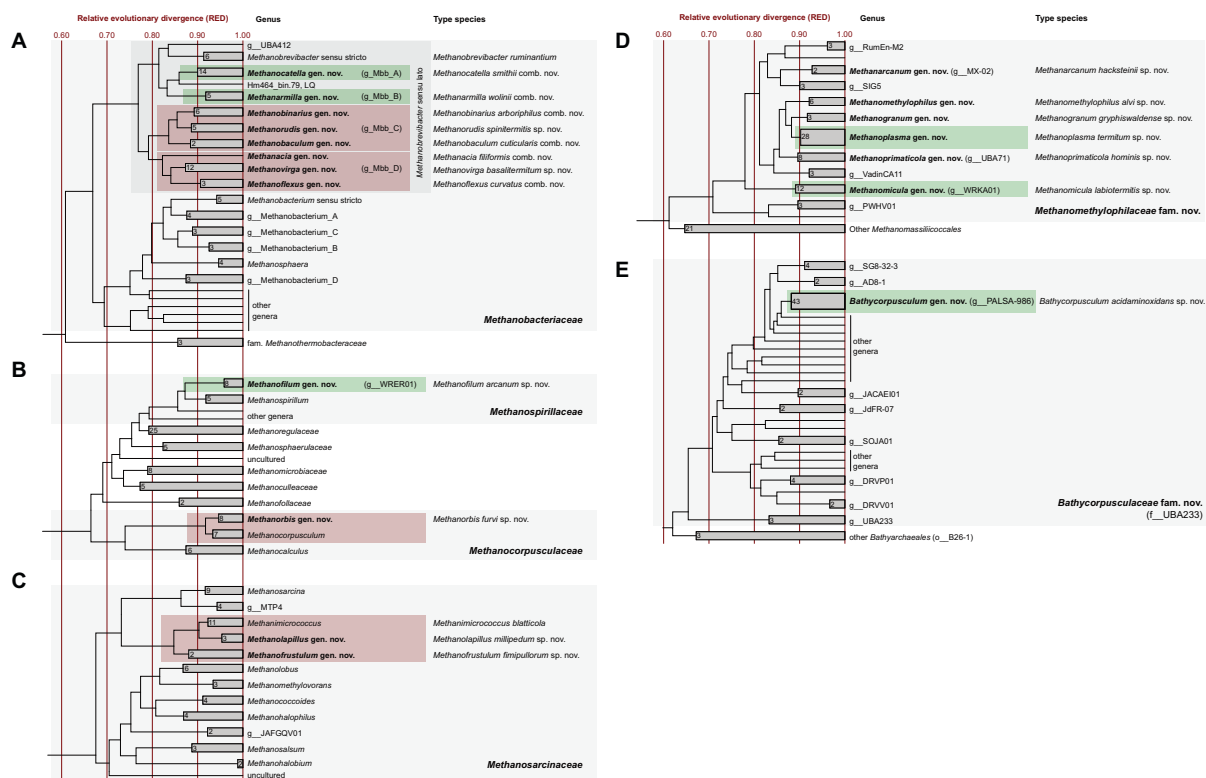


FIGURE 2

Rank-normalized phylogenies of the archaeal families that harbor isolates or MAGs from arthropod guts (A–E). Taxa with representatives from arthropod guts are highlighted in color; green shading indicates genera that are recognized also in GTDB, red shading indicates those that are expanded in the present study. Newly proposed genera are in bold; the corresponding type species are given. Provisional names in the GTDB taxonomy are in parentheses. The maximum-likelihood tree is based on an alignment of 53 marker genes and was normalized using relative evolutionary divergence (RED) values determined with *PhyloRank*. The number of genomes in the collapsed clades is indicated. For an expanded version of the original tree, including genome accession numbers, see [Supplementary Figure 1](#).

clade in the phylogenomic tree ([Figure 2D](#)), we propose to place the members of this lineage in the new genus *Methanoplasma* and the new family *Methanomethylophilaceae* (see section “Taxonomy”).

Three other genus-level clusters that split off at basal nodes in the *Methanomethylophilaceae* tree consist exclusively of uncultured methanogens from arthropods ([Figure 6](#)). One cluster consists exclusively of representatives from cockroaches (cluster C), and another is a mixed cluster comprising sequences from millipedes and higher termites (cluster M). The third cluster consists exclusively of representatives from higher termites, including the 16S rRNA genes of several MAGs. Members of this cluster form a genus-level clade also in the phylogenomic analysis ([Figure 2D](#)) and are assigned to the new genus *Methanomicula* (see section “Taxonomy”).

Bathyarchaeales

Members of the class *Bathyarchaeia* were represented exclusively in higher termites. In the 16S rRNA-based analysis, the clones fall within the radiation of two termite-specific clades previously described as *Ca. Termiticorpusculum* (TB1) and *Ca. Termitimicrobium* (TB2) in the recently described *Bathyarchaeales* ([Loh et al., 2021; Khomyakova et al., 2023](#)). Based on the 16S rRNA gene phylogeny, the phylotypes from termite guts represent a

monophyletic group among various lineages of uncultured archaea from marine sediments, salt marshes, and anaerobic digesters ([Figure 7](#)). Phylogenomic analysis revealed that TB1 and TB2 are polyphyletic and separated by MAGs from hot spring sediments, anaerobic digesters, and permafrost soil (g_PALSA_986 in GTDB; [Figure 2E](#)). We propose to place members of the genus PALSA_986 in the new genus *Bathycorpusculum*, with *Bathycorpusculum acidaminoxidans* as type species (see section “Taxonomy”).

Nitrososphaerales

A small number of sequences from arthropod guts fall within the radiation of *Nitrososphaerales*, where they cluster with uncultured archaea in the genera *Candidatus Nitrosocosmicus* and g_UBA10452 (*Nitrososphaeraeaceae*) ([Supplementary Figure 8](#)). They were absent in most gut samples but were present in low abundance in several humivorous termites, millipedes, and the larva of the scarab beetle *Pachnoda ephippiata* ([Figure 8](#) and [Supplementary Tables 4, 5](#)). In cases where individual compartments were sampled (soil-feeding termites of the genera *Amitermes*, *Isognathotermes*, *Polyspathotermes*, and *Ophiotermes*, and the humivorous larva of *P. ephippiata*), the same phylotypes dominated the clone libraries of food soil, nest material, and often also the anterior gut regions ([Supplementary Table 3](#)),

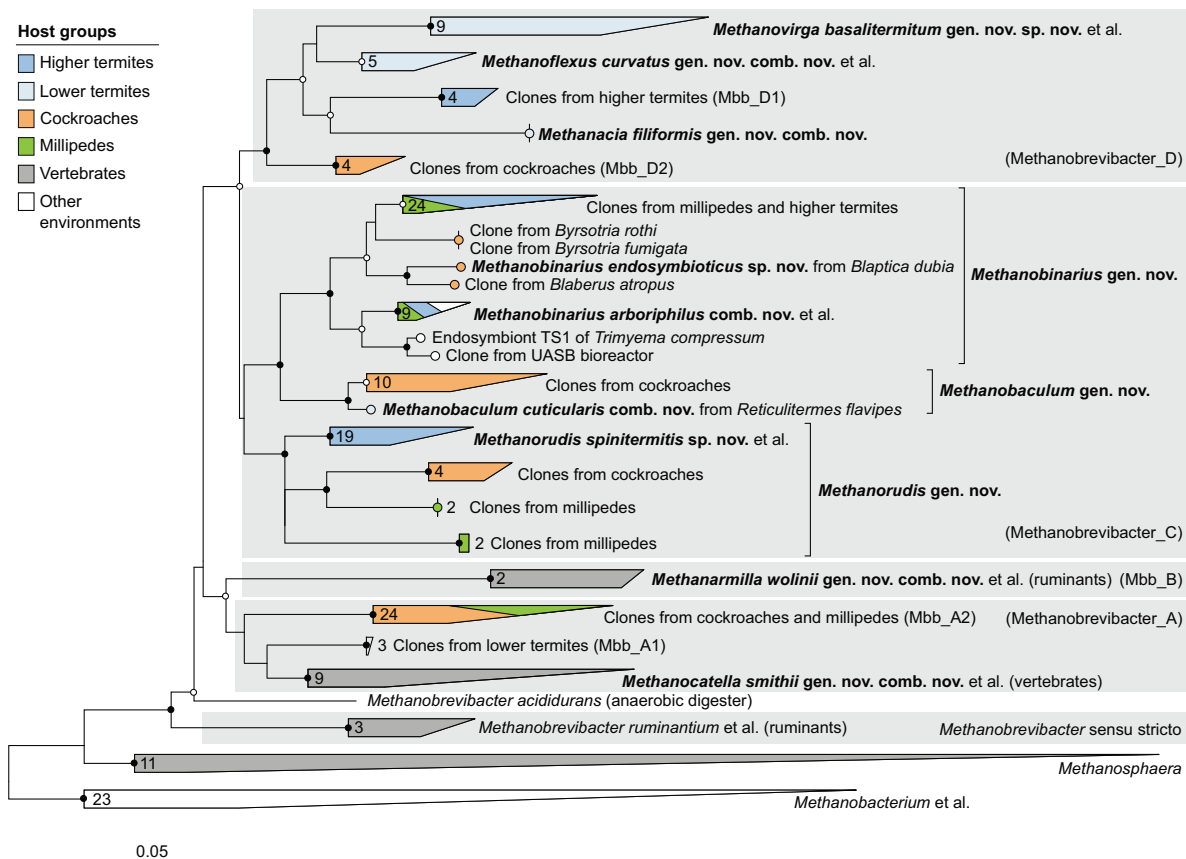


FIGURE 3

Phylogenetic tree of *Methanobacteriaceae*, illustrating the position of the sequences from arthropods obtained in this and previous studies. Other *Methanobacteriales* were used as outgroup. The maximum-likelihood tree is based on a curated alignment of near-full-length 16S rRNA genes (>1,400 sites) and was generated using IQ-TREE under the GTR + I + G4 model of evolution. Bullets on internal nodes indicate SH-aLRT/ultrafast bootstrap support (●, both $\geq 95/99\%$; ○, both $\geq 80/95\%$; 1,000 replicates each). The scale bar indicates the number of substitutions per site. Color coding indicates host groups. Newly described taxa and their type species are in bold; provisional names in the GTDB taxonomy in parentheses. The number of sequences in the collapsed clades is indicated. A fully expanded version of the tree, including accession numbers, is included in [Supplementary Figure 3](#).

suggesting that they are transient microbiota and originated from the environment.

Distribution of archaeal groups across host lineages

We assessed archaeal community structure in methane-emitting arthropods by classifying the 16S rRNA gene libraries obtained in this and previous studies using the phylogenetic framework of our curated reference database ([Supplementary Tables 4, 5](#)). A comparison of representative samples from all host groups revealed that the distribution of methanogenic taxa among arthropods is complex ([Figure 8](#)).

Members of *Methanobacteriales* are present in almost all arthropod species investigated but are unevenly distributed among host groups ([Figure 8](#)). While the genera *Methanobaculum*, *Methanobinarius*, and *Methanorudis* are present in all host groups, *Methanovirga* and *Methanoflexus* are present only in lower termites, and *Methanacia* only in *Reticulitermes* spp. The cockroach cluster (Mbb_A2), which is related to the

genus *Methanocatella*, also contains representatives from several termites and millipedes. Although *Methanobacteriales* dominate the archaeal community in numerous representatives of each host group, they are frequently outnumbered by members of other orders even in closely related hosts. Most striking are the large differences in the occurrence of certain genera between independent samples of the same host species (e.g., *Isognathotermes fungifaber*, *Embiratermes neotenicus*, *Panesthia angustipennis* and *Anadenobolus monilicornis*), which corroborates that the specificity of both hosts and symbionts for their respective partners is not always strict. The complete absence of *Methanobacteriales* from the amplicon libraries of the cockroach *Gyna cafferorum* and two *Glyptotermes* species is noteworthy ([Supplementary Table 4](#)), whereas their absence from several clone libraries should be interpreted with caution because of insufficient sampling depth ([Supplementary Table 5](#)).

Representatives of *Methanomicrobiales* are common in higher termites and millipedes and of lower abundance in cockroaches ([Figure 8](#)). Members of the genus *Methanorubis* occur in millipedes and cockroaches, where they often dominate the archaeal community, but are absent in termites, with the notable exception

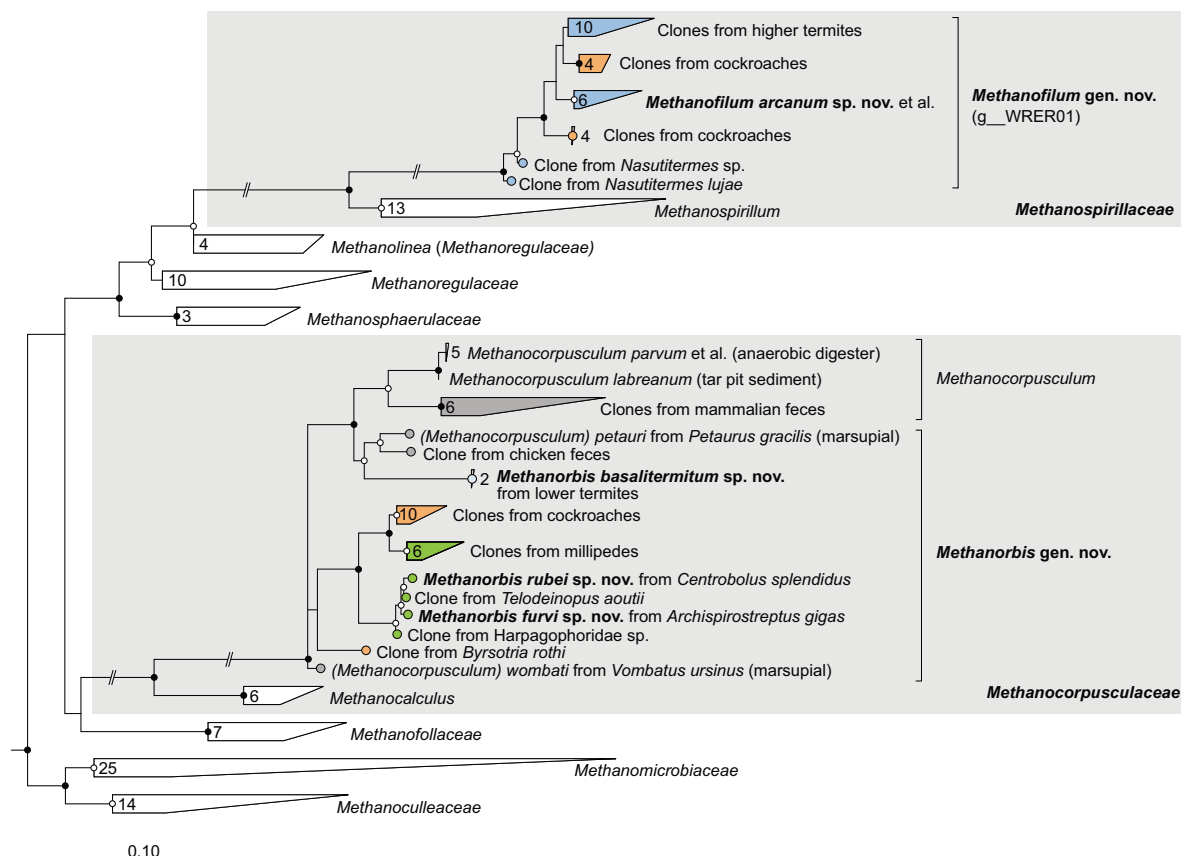


FIGURE 4

Phylogenetic tree of *Methanomicrobiales*, illustrating the position of the sequences from arthropods obtained in this and previous studies. *Methanocellales* were used as outgroup. Color coding and other details are the same as in [Figure 1](#). A fully expanded version of the tree, including accession numbers, is included in [Supplementary Figure 4](#).

of several *Reticulitermes* species. The genus *Methanofilum*, which occurs in all higher termites and, although in lower abundance, also in many cockroaches, is not encountered in lower termites and millipedes.

The order *Methanosarcinales* is represented in all host groups except lower termites (Figure 8). Members of the genus *Methanimicrococcus* are restricted to cockroaches and higher termites, where they often dominate the archaeal community, and the humivorous larva of the scarab beetle *P. ephippiata*. The genus *Methanolapillus* occurs exclusively in millipedes, where it frequently represents the predominant lineage of methanogens.

Members of the *Methanomassiliicoccales* occur in all host groups and can dominate the archaeal community in certain host species (Figure 8). Members of the genus *Methanoplasma* are found in termites, cockroaches, and millipedes, whereas the genus *Methanomicula* occurs exclusively in higher termites, typically in high relative abundance. Members of cluster C, which occur at low abundance in most cockroaches, are also found in some Macrotermitinae (a subfamily of fungus-cultivating higher termites), whereas members of cluster M occur in millipedes, cockroaches, and soil-feeding higher termites. Members of both clusters are present in the humivorous larva of the scarab beetle *P. ephippiata*.

Discussion

Our comprehensive analysis of the archaeal diversity in the intestinal tract of terrestrial arthropods known to emit methane reveals distinct clades of methanogens from the orders *Methanobacteriales*, *Methanomassiliicoccales*, *Methanomicrobiales*, and *Methanosarcinales*. Almost all lineages exhibit a high specificity for a particular host group (i.e., termites, cockroaches, or millipedes) and occupy sister positions to lineages from vertebrates, indicating a common evolutionary origin of host-associated methanogens. Linking the 16S rRNA-based diversity data to a phylogenomic analysis of more than 80 archaeal MAGs from termite guts and the genomes of 9 isolates from cockroaches and millipedes allowed the description of novel genera for each order and a taxonomic revision of methanogens and other archaea in arthropod guts.

Taxonomic revision of the genus
Methanobrevibacter

The first methanogens isolated from arthropod guts were members of the genus *Methanobrevibacter* (Leadbetter and Breznak, 1996; Leadbetter et al., 1998). Together

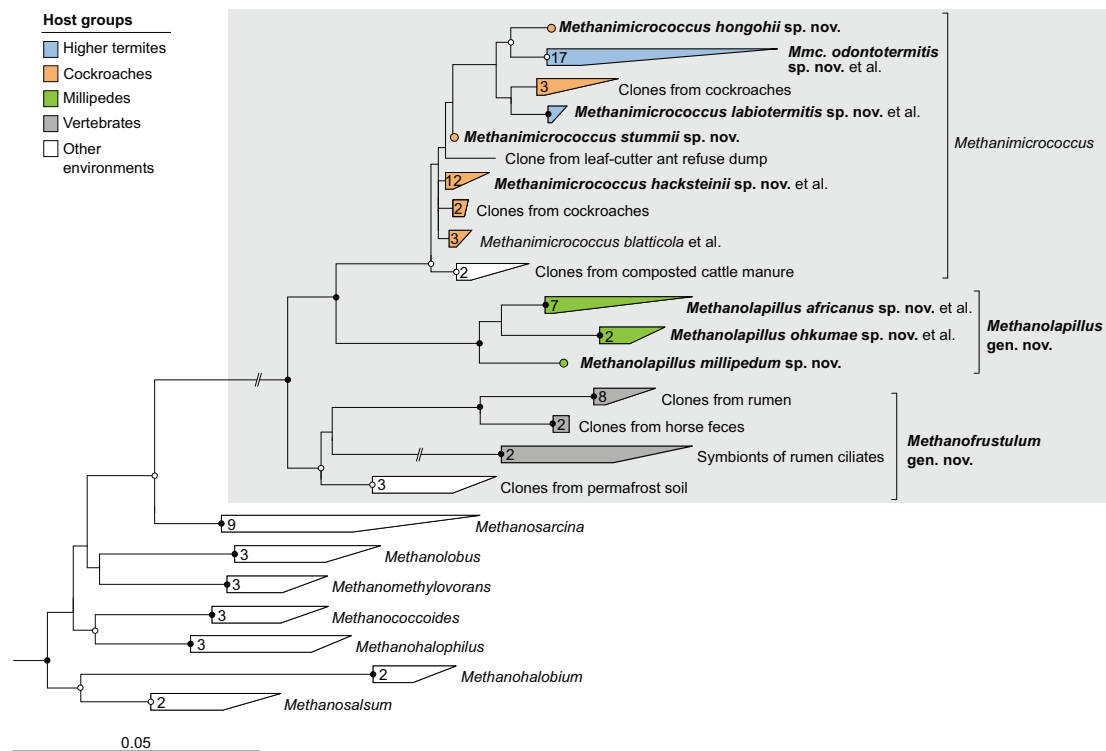


FIGURE 5

Phylogenetic tree of *Methanosarcinaceae* illustrating the position of the sequences from arthropods obtained in this and previous studies. Other *Methanosarcinales* were used as outgroup. Color coding and other details are the same as in Figure 1. A fully expanded version of the tree, including accession numbers, is included in Supplementary Figure 5.

with the 16S rRNA gene sequences of numerous uncultured representatives recovered from the intestinal tract of arthropods (reviewed in Brune, 2018), they represent lineages distinct from those that colonize the intestinal tract of vertebrates (Figure 3). Earlier studies using 16S rRNA and multi-locus gene sequence analyses had already suggested that the genus *Methanobrevibacter* is severely underclassified, comprising multiple genus-level clades that apparently coevolved with different host groups (i.e., ruminants, humans, and termites) (Dighe et al., 2004; Poehlein et al., 2018). This notion was then corroborated by a phylogenomic analysis that expanded the taxonomy of archaea to include genomes from uncultured lineages in the phylogenetic framework of GTDB and suggested that the high levels of RED within the genus *Methanobrevibacter* require the introduction of additional genus-level taxa (Rinke et al., 2021).

Our significantly expanded datasets of 16S rRNA gene sequences and MAGs from arthropod guts underscore the need for taxonomic revision. Based on the RED values of the internal nodes in the radiation of the genus *Methanobrevibacter* sensu lato, we propose to reclassify all species that do not fall into the *M. ruminantium* clade (comprising the type species of the genus *Methanobrevibacter*) into eight new genera: *Methanocatella* (Mbb_A), *Methanarmilla* (Mbb_B), *Methanobaculum*, *Methanobinarius*, and *Methanorudis* (Mbb_C), and *Methanacia*, *Methanoflexus*, and *Methanovirga* (Mbb_D) (Figure 2A). The presence of additional arthropod-specific clusters in the radiation of Mbb_A and Mbb_D (Figure 3), which lack representatives with sequenced genomes, suggests the

presence of additional genus-level lineages that are candidates for future taxonomic revision. The same is true for the species *Methanobrevibacter acididurans*, which has no close relatives in public databases and whose genome remains to be sequenced.

Notably, each of the new genera is specific for a particular host group. While members of *Methanobrevibacter* sensu stricto, *Methanocatella*, and *Methanarmilla* are associated with the intestinal tract of ruminants and other vertebrates, all other genera are associated with arthropods and often include subclades restricted to either termites, cockroaches, or millipedes. The genera *Methanacia*, *Methanoflexus*, and *Methanovirga* currently consist exclusively of representatives from termites (Figure 3). The presence of a clade of unclassified phylotypes from higher termites (Mbb_D1) within the radiation of Mbb_D and another clade from cockroaches (Mbb_D2) in sister position to all clades from termites is in agreement with the evolutionary origin of termites among cockroaches (Inward et al., 2007) and a coevolutionary history of the members of the Mbb_D clade with their dictyopteran hosts.

The genera *Methanobaculum*, *Methanobinarius*, and *Methanorudis* also consist of lineages that are specific to particular arthropod host groups. In the genus *Methanobinarius*, the close relatedness among representatives from distantly related host lineages, i.e., termites (class Insecta) and millipedes (class Myriapoda), suggests an environmental transfer of methanogens between these soil-dwelling arthropods. This is underscored by the presence of *Methanobinarius* clones in anaerobic bioreactors (Carrillo-Reyes et al., 2014) and in the free-living, anaerobic ciliate *Trimyema compressum* (Shinzato et al., 2007).

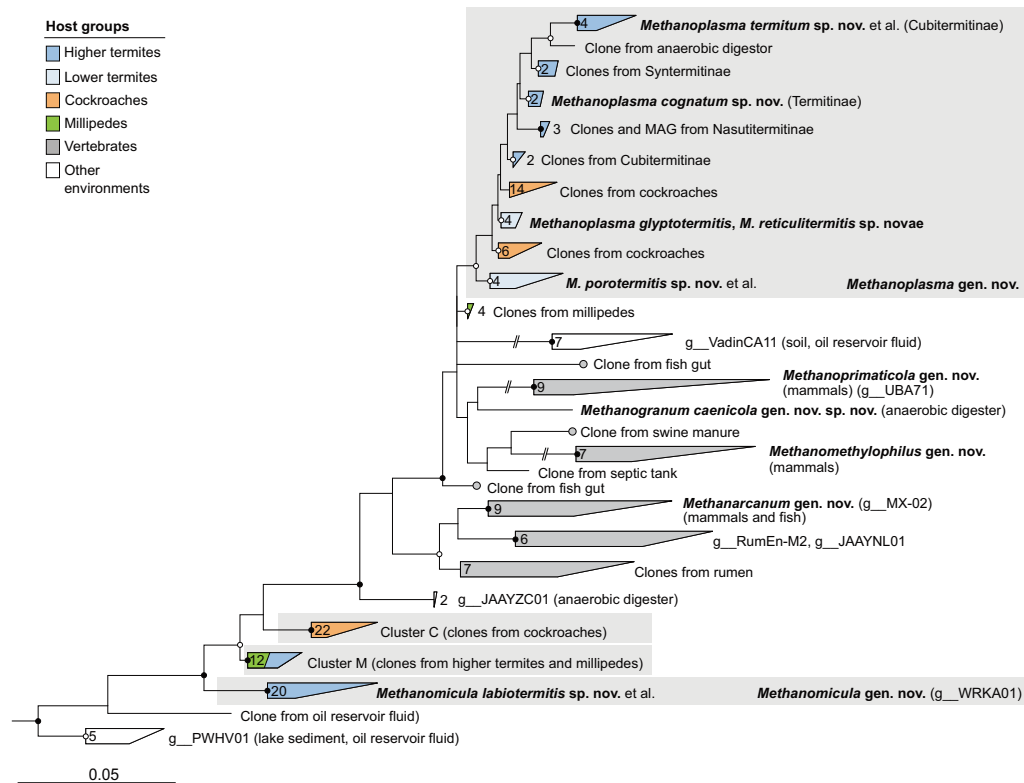


FIGURE 6

Phylogenetic tree of *Methanomethylophilaceae*, illustrating the position of the sequences from arthropods obtained in this and previous studies. Other *Methanomassiliicoccales* were used as outgroup. Color coding and other details are the same as in [Figure 1](#). A fully expanded version of the tree, including accession numbers, is included in [Supplementary Figure 6](#).

Representatives of the genus *Methanosphaera*, which are typical for the intestinal tract of mammals ([Hoedt et al., 2018](#); [Thomas et al., 2022](#)), were not detected in any arthropod host. Members of the genus *Methanobacterium* were only rarely encountered, indicating that the strains of *Methanobacterium bryantii* isolated from several higher termites ([Deevong et al., 2004](#)) are not common members of the archaeal microbiota of arthropods.

New arthropod-specific genera in *Methanomicrobiales*

The order *Methanomicrobiales* comprises two so far unclassified arthropod-specific clades in the families *Methanocorpusculaceae* and *Methanospirillaceae*. In the phylogenomic analysis, the genomes from termites and millipedes are sister to the genus *Methanocorpusculum* ([Figure 2B](#)). A recent phylogenomic analysis revealed that the host-associated members of the genus *Methanocorpusculum* form two major clades ([Volmer et al., 2023](#)). Host clade 2, which is sister to the environmental clade comprising all described species of the genus *Methanocorpusculum*, consists entirely of uncultured archaea from mammalian feces. Host clade 1 consists of genomes from the feces of chickens and marsupials, including *Ca. Methanocorpusculum faecipullorum* and two isolates, *Methanocorpusculum petauri* and *Methanocorpusculum wombati* ([Gilroy et al., 2021](#); [Volmer et al., 2023](#)). Since all members of host clade 1 fall within the radiation

of genomes from millipedes and termites and are well separated from the remaining members of the genus *Methanocorpusculum* ([Supplementary Figure 2](#)), we propose to place them in the new genus *Methanorbis* (see section “Taxonomy”).

Amplicon sequencing of archaea in the intestinal tract of vertebrates has suggested that ancestral members of the genus *Methanocorpusculum* were present in the last common ancestor of ungulates ([Thomas et al., 2022](#)). Our results support a common ancestry of both arthropod-associated and vertebrate-associated *Methanocorpusculaceae*. Although the genera *Methanorbis* and *Methanocorpusculum* are not fully resolved in 16S rRNA-based analyses ([Figure 4](#)), phylogenomic analysis places all genomes from arthropods, including all MAGs from termite guts, into the genus *Methanorbis* ([Supplementary Figure 2](#)). The two sister clades of uncultured *Methanocorpusculaceae* from cockroaches and millipedes that lack representatives with sequenced genomes most likely belong also to the genus *Methanorbis*. The widespread distribution of *Methanorbis* in cockroaches and millipedes suggests that the representatives associated with chicken and marsupial feces were acquired by consuming arthropods. So far, it remains open whether the genus *Methanocorpusculum*, including the environmental clade, originated from a free-living or an arthropod-associated ancestor.

Members of the new genus *Methanofilum* occur exclusively in cockroaches and higher termites ([Figures 4, 8](#)). The absence of *Methanospirillaceae* in the intestinal tract of all other animals suggests that the genus arose within the order

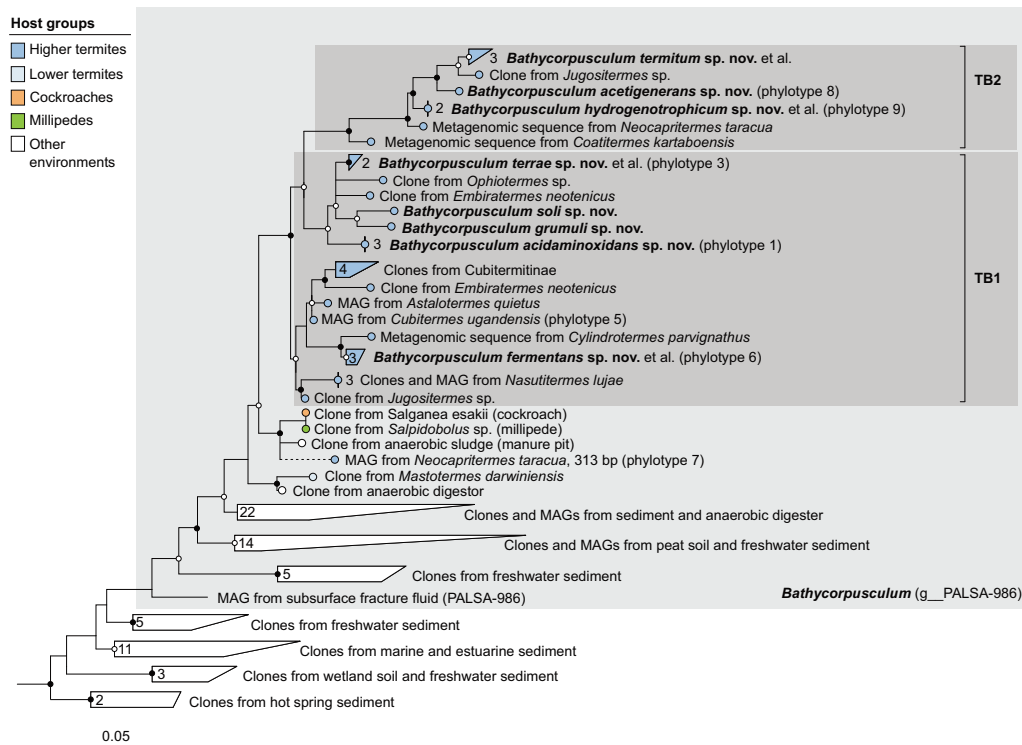


FIGURE 7

Phylogenetic tree of *Bathycorpusculaceae*, illustrating the position of the sequences from arthropods obtained in this and previous studies. Other *Bathyarchaeales* were used as an outgroup. Color coding and other details are the same as in Figure 1. A fully expanded version of the tree, including accession numbers, is included in Supplementary Figure 7.

Blattodea, descending from free-living ancestors that occurred in their soil environment. This would mirror the situation with the genus *Methanosphaera*, which occurs exclusively in vertebrates and presumably evolved from an ancestral lineage of *Methanobacteriaceae* (Hoedt et al., 2018). The genus *Methanomicrobium*, which is common in the gut of ruminants (Janssen and Kirs, 2008; Henderson et al., 2015), is not represented in arthropods.

Novel lineages of host-associated *Methanosarcinales*

Arthropod-associated members of *Methanosarcinales*, which were first detected in archaeal clone libraries of higher termites (Ohkuma et al., 1999; Friedrich et al., 2001), cockroaches (Sprenger et al., 2000), and scarab beetle larvae (Egert et al., 2003), are abundant in all host groups except lower termites (Figure 8). The genus *Methanimicrococcus* comprises all representatives from cockroaches and higher termites, whereas the new genus *Methanolapillus* harbors those associated with millipedes (Figure 4), suggesting that the lineages in Insecta and Diplopoda evolved independently from each other.

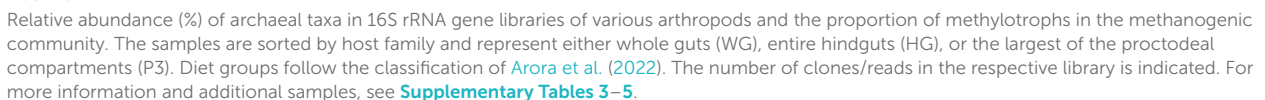
Unlike the sister genus *Methanosarcina*, whose members have the widest substrate range among methanogens, all isolates and genomes in the host-associated genera *Methanimicrococcus* and *Methanolapillus* examined to date are obligately hydrogen-dependent methylotrophs. Genomic analysis of the

uncultivated representatives of the mammal-associated genus *Methanofrustulum* is pending, but all arthropod-associated lineages have lost the methyl branch of the Wood-Ljungdahl pathway and use methanol and methylamines as substrates only in the presence of hydrogen (Sprenger et al., 2007; Thomas et al., 2021; Protasov and Brune, unpublished results).

The arthropod-specific genera occupy a sister position to the new genus *Methanofrustulum*, whose representatives were first detected in horses but are found also in ruminants and other ungulates (Lwin and Matsui, 2014; Huang et al., 2016), where they have been identified as endosymbionts of rumen ciliates (Regensbogenova et al., 2004). The wide distribution of *Methanimicrococcus*-related archaea (most likely representing *Methanofrustulum*) in short-read amplicon libraries of vertebrates (Thomas et al., 2022) suggests parallel evolution of gut-associated *Methanosarcinaceae* in vertebrates and arthropods.

Methanomassiliicoccales – Taxonomic update of the intestinal clade

The order *Methanomassiliicoccales* consists exclusively of obligately hydrogen-dependent methylotrophs (Zinke et al., 2021). It comprises two major clades that occur in contrasting environments. While members of the so-called environmental clade, represented by the family *Methanomassiliicoccaceae*, occur predominantly in anoxic soils, sediments, wetlands, and subsurface habitats, members of the so-called intestinal clade are found mostly



The genus *Methanoplasma*, which occurs exclusively in arthropod guts, belongs to an apical clade of the family *Methanomethylophilaceae*, with the genera *Methanogranum*, *Methanomethylophilus*, and *Methanoprismaticola* as its closest relatives (Figure 2). The same clade is also represented in the 16S rRNA-based analysis, although the branching order of its members is not fully resolved (Figure 6). Within the genus *Methanoplasma*, representatives from cockroaches and the phylogenetically

older termite families branch more deeply than those from the phylogenetically younger higher termites, suggesting co-evolution between *Methanoplasma* and its blattodean hosts. An exception is the presence of an apical lineage in millipedes, which was most likely acquired by an environmental transfer from soil-feeding Cubitermitinae.

By contrast, the genus *Methanomicula*, which occurs exclusively in higher termites, occupies a basal position in the phylogeny of *Methanomethylphilaceae* (Figure 2). It contains no genomes from cockroaches and millipedes, but both host groups are represented in the 16S-based analyses (Figure 6). As in the case of the genus *Methanimicrococcus*, *Methanomicula* is also consistently absent in lower termites (Figure 8). While some older clone libraries described in the literature have been undersampled and suffer from primer bias against *Methanomassiliicoccales* (see discussion in Paul et al., 2012), the high abundance of *Methanoplasma* in amplicon libraries of lower termites suggests that the absence of *Methanomicula* is not an artifact.

The genus *Bathycorpusculum* and description of the family *Bathycorpusculaceae*

Members of the “*Bathyarchaeota*,” a name coined by Meng et al. (2014) for the Miscellaneous Crenarchaeotal Group (MCG), are presently considered a class-level lineage in the phylum *Thermoproteota*. The class *Bathyarchaeia* was formally described only recently, based on its first cultured representative, *Bathyarchaeum tardum* isolated from the anaerobic sediment of a coastal lake (Khomyakova et al., 2023). The type genus, *Bathyarchaeum*, and its family, *Bathyarchaeaceae* (formerly BA1), belong to the order *Bathyarchaeales* (formerly B26-1).

Members of the order *Bathyarchaeales*, at the time referred to as the “freshwater cluster” of the “*Crenarchaeota*,” were first detected in arthropod guts in archaeal clone libraries of soil-feeding termites (Friedrich et al., 2001). A genome-centric analysis of these termite-specific lineages identified them as members of the family UBA233 (also referred to as subgroup MCG-6 or Bathy-6) and tentatively assigned them to the candidate taxa “*Termiticorpusculum*” (TB1) and “*Termitimicrobium*” (TB2) (Loh et al., 2021). In the taxonomic framework of GTDB, they belong to a single genus-level lineage, described here as the new genus *Bathycorpusculum* (formerly PALSA-986), which also includes MAGs from peat soils, sediments, and anaerobic digesters, in the new family *Bathycorpusculaceae* (Figure 2E and Supplementary Figure 7).

Comparative genome analysis of *Bathycorpusculum* MAGs revealed a purely fermentative metabolism based on amino acids with the potential for reductive acetogenesis from H_2 and CO_2 (TB1) or possibly methylated compounds (TB2) (Loh et al., 2021) but ruled out the capacity for methanogenesis or alkane oxidation, which had been reported for other lineages of *Bathyarchaeia* (Evans et al., 2019). Until recently, host-associated *Bathyarchaeia* had been detected only in higher termites (Friedrich et al., 2001; Grieco et al., 2013; Shi et al., 2015). However, short-read amplicon libraries of archaea in the intestinal tract of animals documented the presence of *Bathyarchaeia* also in various species of vertebrates (Youngblut et al., 2021; Thomas et al., 2022). Thomas et al. (2022) demonstrated

that the short reads from diverse vertebrates cluster in a sister position to representatives of the genus *Bathycorpusculum*. In contrast to the situation in higher termites, where members of the genus *Bathycorpusculum* appear to be part of the autochthonous microbiota (Loh et al., 2021), the vertebrate hosts are only distantly related, which supports the hypothesis that they were acquired independently from either host environment or diet (Youngblut et al., 2021; Thomas et al., 2022). Remarkably, amplicon libraries of cockroaches and millipedes also contain rare phylotypes that fall outside the radiation of TB1 and TB2 from termites in a lineage that also includes a clone from a manure pit (Figure 7 and Supplementary Figure 7).

Nitrososphaerales and other transient microbiota

Members of the order *Nitrososphaerales* were detected in archaeal clone libraries of soil-feeding termites more than 20 years ago (Friedrich et al., 2001). At the time referred to as the “Terrestrial Cluster” of “*Crenarchaeota*,” they were subsequently placed in the candidate phylum “*Thaumarchaeota*” (Brochier-Armanet et al., 2008), which was recently described as *Nitrososphaerota* (Oren and Garrity, 2021). The GTDB taxonomy classifies the class *Nitrososphaeria* in the phylum *Thermoproteota*.

Members of the order *Nitrososphaerales* are aerobic, ammonia-oxidizing archaea and occur in a wide range of marine and terrestrial ecosystems (Pester et al., 2011). In arthropod guts, they were detected in only a few soil or litter-feeding species of all host groups (Figure 8). The phylotypes do not form arthropod-specific lineages but fall into the radiation of uncultured representatives from soils and sediments (Supplementary Figure 8), suggesting that they are part of a transient microbiota taken up from the environment. This is supported by their prevalence in the food soil and midgut of the humivorous larva of *P. ephippiata* or in the food soil and/or nest material of soil-feeding termites (e.g., *Isognathotermes* spp., *Ophiotermes* sp., and *Nitiditermes orthognatus*), where they are abundant also in the anterior gut compartments but not in the hindgut (P1–P4; Supplementary Table 3). This is consistent with the consumption of nest material by soil-feeding termites (Nalepa et al., 2001).

Using the *amoA* gene as a functional marker, ammonia-oxidizing archaea have been detected particularly in the guts of soil-feeding termites and humivorous scarab beetle larvae (Majeed et al., 2014; Miambi et al., 2022). Considering the high ammonia concentrations in the guts of these host groups (Ji and Brune, 2006; Ngugi et al., 2011) and the considerable influx of oxygen into the peripheral regions of all gut compartments (Brune et al., 1995), *Nitrososphaerales* may be metabolically active during gut passage and contribute to the emissions of the greenhouse gas N_2O by termites and scarab beetle larvae (Ngugi and Brune, 2012; Brauman et al., 2015).

Methanogens of the genera *Methanocella* and *Methanosarcina* were not detected in gut samples of arthropods but in the food soil of *Cubitermes fungifaber* and *P. ephippiata*. The latter were also absent in most gut samples of millipedes but highly abundant in their excreta (Šustr et al., 2014). A representative of the genus *Methanothrix* was detected in the gut of a single sample of

TABLE 1 New taxa and new combinations of archaea proposed under SeqCode and the designated nomenclatural type.

Taxon and descriptor of new taxa (type genus or species)	New species or combination (type genome)
Methanobacteriales	
<i>Methanocatella</i> gen. nov. Protasov and Brune (<i>Methanocatella smithii</i> comb. nov.)	<i>Methanocatella smithii</i> comb. nov. (GCF_000016525)
	<i>Methanocatella gottschalkii</i> comb. nov. (GCF_003814835)
	<i>Methanocatella millerae</i> comb. nov. (GCF_900103415)
	<i>Methanocatella oralis</i> comb. nov. (GCF_001639275)
	<i>Methanocatella thaueri</i> comb. nov. (GCF_003111625)
	<i>Methanocatella woesei</i> comb. nov. (GCF_003111605)
<i>Methanarmilla</i> gen. nov. Protasov and Brune (<i>Methanarmilla wolinii</i> comb. nov.)	<i>Methanarmilla wolinii</i> comb. nov. (GCF_000621965)
	<i>Methanarmilla boviskoreani</i> comb. nov. (GCF_000320505)
<i>Methanobinarius</i> gen. nov. Protasov and Brune (<i>Methanobinarius arboriphilus</i> comb. nov.)	<i>Methanobinarius arboriphilus</i> comb. nov. (GCF_002072215)
	<i>Methanobinarius endosymbioticus</i> comb. nov. (GCA_003315655)
<i>Methanobaculum</i> gen. nov. Protasov and Brune (<i>Methanobaculum cuticularis</i> comb. nov.)	<i>Methanobaculum cuticularis</i> comb. nov. (GCA_001639285)
<i>Methanoflexus</i> gen. nov. Protasov and Brune (<i>Methanoflexus curvatus</i> comb. nov.)	<i>Methanoflexus curvatus</i> comb. nov. (GCF_001639295)
	<i>Methanoflexus mossambicus</i> sp. nov. (GCA_031261915)
<i>Methanorudis</i> gen. nov. Protasov and Brune (<i>Methanorudis spinitermitis</i> sp. nov.)	<i>Methanorudis spinitermitis</i> sp. nov. (GCA_031286225)
<i>Methanovirga</i> gen. nov. Protasov and Brune (<i>Methanovirga basalitermitum</i> sp. nov.)	<i>Methanovirga aequatorialis</i> sp. nov. (GCA_031282205)
	<i>Methanovirga australis</i> sp. nov. (GCA_031272765)
	<i>Methanovirga basalitermitum</i> sp. nov. (GCA_031284445)
	<i>Methanovirga meridionalis</i> sp. nov. (GCA_031289325)
	<i>Methanovirga procula</i> sp. nov. (GCA_031280375)
<i>Methanacia</i> gen. nov. Protasov and Brune (<i>Methanacia filiformis</i> comb. nov.)	<i>Methanacia filiformis</i> comb. nov. (GCF_001639265)
Methanomicrobiales	
<i>Methanorbis</i> gen. nov. Protasov and Brune (<i>Methanorbis furvi</i> sp. nov.)	<i>Methanorbis basalitermitum</i> sp. nov. (GCA_031287415)
	<i>Methanorbis furvi</i> sp. nov. (GCA_032714615)
	<i>Methanorbis rubei</i> sp. nov. (GCA_032714495)
<i>Methanofilum</i> gen. nov. Protasov and Brune (<i>Methanofilum arcanum</i> sp. nov.)	<i>Methanofilum arcanum</i> sp. nov. (GCA_031285085)
Methanosarcinales	
<i>Methanimicrococcus</i> (<i>Methanimicrococcus blatticola</i>)	<i>Methanimicrococcus hacksteinii</i> sp. nov. (GCA_032714515)
	<i>Methanimicrococcus hongohii</i> sp. nov. (GCA_032594095)
	<i>Methanimicrococcus labiatermitis</i> sp. nov. (GCA_009784005)
	<i>Methanimicrococcus odontotermitis</i> sp. nov. (GCA_031286065)
	<i>Methanimicrococcus stummii</i> sp. nov. (GCA_032594435)
<i>Methanolapillus</i> gen. nov. Protasov and Brune (<i>Methanolapillus millepedarum</i> sp. nov.)	<i>Methanolapillus africanus</i> sp. nov. (GCA_032714475)
	<i>Methanolapillus ohkumae</i> sp. nov. (GCA_032594355)
	<i>Methanolapillus millepedarum</i> sp. nov. (GCA_032594115)
<i>Methanofrustulum</i> gen. nov. Protasov and Brune (<i>Methanofrustulum fimipullorum</i> sp. nov.)	<i>Methanofrustulum fimipullorum</i> sp. nov. (GCA_012518265)
Methanomassiliicoccales	
<i>Methanomethylophilaceae</i> fam. nov. Gaci et al. (<i>Methanomethylophilus</i> gen. nov.)	
<i>Methanomethylophilus</i> gen. nov. Borrel et al. (<i>Methanomethylophilus alvi</i> sp. nov.)	<i>Methanomethylophilus alvi</i> sp. nov. (GCA_000300255)
<i>Methanarcanum</i> gen. nov. Chibani et al. (<i>Methanarcanum hacksteinii</i> sp. nov.)	<i>Methanarcanum hacksteinii</i> sp. nov. (GCA_006954405)
<i>Methanoprismaticola</i> gen. nov. Chibani et al. (<i>Methanoprismaticola hominis</i> sp. nov.)	<i>Methanoprismaticola hominis</i> sp. nov. (GCA_006954465)

(Continued)

TABLE 1 (Continued)

Taxon and descriptor of new taxa (type genus or species)	New species or combination (type genome)
<i>Methanogranum</i> gen. nov. Iino et al. (<i>Methanogranum gryphiswaldense</i> sp. nov.)	<i>Methanogranum gryphiswaldense</i> sp. nov. (GCA_019262145)
<i>Methanoplasma</i> gen. nov. Lang and Brune (<i>Methanoplasma termitum</i> sp. nov.)	<i>Methanoplasma termitum</i> sp. nov. (GCF_000800805)
	<i>Methanoplasma cognatum</i> sp. nov. (GCA_009777615)
	<i>Methanoplasma glyptotermis</i> (sp. nov. GCA_031267895)
	<i>Methanoplasma porotermis</i> sp. nov. (GCA_031290095)
	<i>Methanoplasma reticulitermis</i> sp. nov. (GCA_031287135)
<i>Methanomicula</i> gen. nov. Protasov and Brune (<i>Methanomicula labiotermis</i> sp. nov.)	<i>Methanomicula labiotermis</i> sp. nov. (GCA_009780575)
Bathyarchaeales	
<i>Bathycorpusculaceae</i> fam. nov. Loh and Brune (<i>Bathycorpusculum</i> gen. nov.)	
<i>Bathycorpusculum</i> gen. nov. Loh and Brune (<i>Bathycorpusculum acidaminoxidans</i> sp. nov.)	<i>Bathycorpusculum acidaminoxidans</i> sp. nov. (GCA_009786255)
	<i>Bathycorpusculum acetigenerans</i> sp. nov. (GCA_009781675)
	<i>Bathycorpusculum fermentans</i> sp. nov. (GCA_009787175)
	<i>Bathycorpusculum hydrogenotrophicum</i> sp. nov. (GCA_009783705)
	<i>Bathycorpusculum grumuli</i> sp. nov. (GCA_009776805)
	<i>Bathycorpusculum soli</i> sp. nov. (GCA_031277345)
	<i>Bathycorpusculum terrae</i> sp. nov. (GCA_009784175)
	<i>Bathycorpusculum termitum</i> sp. nov. (GCA_031254875)

The protologues including the descriptors of each new taxon, the etymologies of the new taxon names, and the full descriptions of all taxa are given in the [Supplementary Data File 1](#).

I. fungifaber (Cu547). Members of these genera are typical for soil and sediment habitats but do not belong to the autochthonous archaeal microbiota in arthropod guts.

Hydrogenotrophic vs. methylotrophic lineages

Methanogens colonizing the gut of arthropods reduce either CO₂ or methyl groups to methane using hydrogen as electron donor. The former are hydrogenotrophic methanogens from the orders *Methanobacteriales* and *Methanomicrobiales*; the latter are obligately hydrogen-dependent methyl reducers from the orders *Methanosarcinales* and *Methanomassiliicoccales*. Obligately methyl-reducing *Methanosarcinales*, represented exclusively by the genera *Methanimicrococcus* and *Methanolapillus* (Thomas et al., 2021; Protasov and Brune, unpublished results), and all members of the order *Methanomassiliicoccales* lack the methyl branch of the Wood–Ljungdahl pathway and are therefore restricted to methylated substrates such as methanol or methylamines (e.g., Borrel et al., 2014; Lang et al., 2015; Söllinger et al., 2016). Members of the genus *Methanospaera*, which are obligately methyl-reducing methanogens in the intestinal tract of mammals (Miller and Wolin, 1985; Fricke et al., 2006), are absent in arthropods.

The proportion of methylotrophic lineages in the methanogenic communities of arthropod guts differs substantially among host species (Figure 8). While each host family (and subfamily of higher termites) comprises species that harbor only hydrogenotrophic methanogens, all major host groups comprise representatives with a high abundance of methylotrophs. Even among lower termites, which were previously thought to harbor

mainly hydrogenotrophic *Methanobacteriales* (reviewed by Brune, 2019), several species are heavily colonized by methylotrophs. Since methyl-reducing methanogens will always outcompete CO₂-reducing hydrogenotrophs for H₂, the abundance of methylotrophs in intestinal tracts is most likely regulated by the availability of methyl groups (Feldewert et al., 2020).

Methyl-disproportioning and acetoclastic methanogens are absent in the intestinal tract of arthropods. Although many members of the order *Methanosarcinales* can dismutate methyl groups to methane and CO₂, their independence from external hydrogen is of little advantage in the intestinal tract of animals, where they are outcompeted by methyl-reducing methanogens owing to their low affinity for methanol and other methylated compounds (Sprenger et al., 2007; Feldewert et al., 2020). The absence of acetoclastic methanogens, however, is more puzzling, since acetate concentrations are much higher in intestinal tracts than in sediments. Although it is generally assumed that the rather slow-growing members of this guild cannot cope with the short residence times of the intestinal contents, it remains enigmatic why they do not avoid washout by attaching to intestinal surfaces or protists (see Brune, 2019).

Microhabitats in arthropod guts

The intestinal tracts of arthropods are characterized by steep radial gradients of oxygen and hydrogen between gut wall and lumen and strong axial dynamics of these and other physicochemical parameters (Brune, 2014, 2019). Hence, it is not surprising that their archaeal communities are diverse and differ between gut compartments (Friedrich et al., 2001).

Hydrogenotrophic *Methanobacteriaceae* of the genera *Methanacia*, *Methanobaculum*, and *Methanoflexus* have been localized on the chitinous lining of the hindgut wall of termites (Leadbetter and Breznak, 1996; Leadbetter et al., 1998). The methyl-reducing *M. blatticola* colonizes the same microhabitat in the hindgut of cockroaches (Sprenger et al., 2000). Although attachment to the hindgut wall is thought to prevent washout, it comes at a cost. Hydrogenotrophs attached to the gut wall are not only severely hydrogen-limited but also exposed to the constant influx of oxygen into this microhabitat (Tholen and Brune, 2000). Both *Methanobaculum cuticularis* and *M. blatticola* actively remove oxygen from their environment (Leadbetter and Breznak, 1996; Sprenger et al., 2007). While *Methanobinarius arboriphilus* and *M. cuticularis* reduce oxygen using H₂ via a F₄₂₀-dependent oxidase (Seedorf et al., 2004; Tholen et al., 2007), the mechanism employed by *M. blatticola* is unclear.

Associations with protists

Association with protists prevents washout and allows methanogens to position themselves in the anoxic lumen of the hindgut, where hydrogen supply is also better than at the gut wall (Brune, 2019). Moreover, most anaerobic protists in the hindgut of arthropods possess hydrogenosomes and provide a stable substrate source for their hydrogenotrophic symbionts. Associations between methanogens and protists are also common in sediments and are regarded as mutualistic because of the cross-feeding of H₂ (Fenchel and Finlay, 2018; Treitli et al., 2023).

Ciliates of the genus *Nyctotherus*, which are found in the gut of cockroaches and millipedes, are commonly colonized by methanogens of the genera *Methanobaculum* and *Methanobinarius*, including *Methanobinarius endosymbioticus* from *Nyctotherus ovalis* (Gijzen et al., 1991; van Hoek et al., 2000; Lind et al., 2018; Supplementary Figure 3). Free-living relatives of these ciliates in sediments, however, are associated with hydrogenotrophic methanogens of the genera *Methanoregula*, *Methanocorpusculum*, *Methanoplanus*, or *Methanobacterium* (Fenchel and Finlay, 2018), suggesting that the endosymbionts are not host-specific but were acquired independently from the pool of methanogens present in the respective environment. This is also the case for rumen ciliates, which are associated with a close relative of *M. ruminantium* (Tokura et al., 1999).

Parabasalid flagellates of lower termites are also frequently associated with methanogens (Odelson and Breznak, 1985; Lee et al., 1987). They were identified as members of the genera *Methanovirga* and *Methanoflexus* (Ohkuma et al., 1995; Tokura et al., 2000; Hara et al., 2004; Inoue et al., 2008; Supplementary Figure 3). Both genera occur exclusively in the gut of lower termites, corroborating that flagellate-associated methanogens have also evolved multiple times from free-living lineages in their respective environment.

Host specificity and mode of transmission

Although methanogens from arthropods typically form clusters specific for a particular host group, evidence of a co-cladogenesis is

not always conclusive. In addition, there are numerous examples of host switching within a given clade (Figure 8). This suggests that the archaeal microbiota in arthropods exhibits a mixed mode of transmission, including both vertical transfer from parents to offspring and environmental exchange (e.g., through predation or co-habitation), as suggested already for the bacterial microbiota of termites (Bourguignon et al., 2018). Environmental transfer would also explain why millipede-associated archaea frequently cluster with those of termites, as millipedes are frequently found in termite nests (Mwabvu, 2005; Choosai et al., 2009).

Many lineages of methanogens present in millipedes, cockroaches, and higher termites are also found in the larva of *P. ephippiata* (Egert et al., 2003). With the exception of the genus *Methanobinarius*, all lineages found in the humivorous larva of *P. ephippiata* are absent in the root-feeding larva of the closely related *Melolontha melolontha* (Egert et al., 2005). An exchange of methanogens between arthropods that occur in the same habitat (e.g., via the feces) seems likely, especially given the close relatedness of the respective phylotypes (Supplementary Figures 3–7).

Conclusion

Arthropods harbor unique lineages of methanogens from several orders. They comprise both CO₂-reducing and methyl-reducing hydrogenotrophs. Some lineages (*Methanimicrococcus*, *Methanolapillus*, *Methanorbis*, *Methanomicula*, and *Methanoplasma*) are sister groups of lineages from the intestinal tract of vertebrates, indicating a common evolutionary origin from non-intestinal ancestors, whereas other lineages must have arisen only in arthropods (*Methanofilum*). The deep-branching phylogenies of each host-associated clade (at least at the genus level) indicate that they have coevolved with their intestinal niches over a long period of time since acquisition from the environment. The occurrence of the same lineages in unrelated host groups suggests the presence of similar ecological niches in the gut of methane-emitting arthropods. However, the reason for the absence of methanogens in all other arthropod groups remains unclear.

Taxonomy

Most archaea from the arthropod guts belong to genus-level lineages that are either unclassified or require reclassification. The presence of both high-quality genomes and 16S rRNA gene sequences for most lineages allow the proposal of new taxa under the *Code of Nomenclature of Prokaryotes Described from Sequence Data* (SeqCode) (Hedlund et al., 2022; Whitman et al., 2022). The new names and new combinations are listed in Table 1, along with the designated nomenclatural type. The authors of previously proposed *Candidatus* names are assigned as descriptors for the corresponding new taxa. The protologues including etymologies and the full descriptions are given in Supplementary Data File 1). The new isolates will be described also under ICNP in separate publications once genome analysis and phenotypic characterization are completed.

Materials and methods

Samples and DNA extraction

Termite colonies that were collected in the field were sampled within a week of collection. Samples from termite colonies maintained in other laboratories were processed within a few days after arrival. Species identity was confirmed by comparing their mitochondrial cytochrome oxidase II (COII) gene sequences (Pester and Brune, 2006) with those in public databases. COII gene sequences that were not represented in public databases have been submitted to NCBI GenBank. Cockroaches, beetles, and millipedes were obtained from commercial breeders and dissected within a few days after arrival. Detailed information on all samples is given in [Supplementary Table 1](#).

Specimens were immobilized on ice, decapitated, and dissected with sterile forceps. Whole guts or individual gut sections were pooled and homogenized in phosphate buffer (Köhler et al., 2012; Schauer et al., 2012). The number of specimens included in each sample was adjusted to account for size differences between species (one gut or gut section for millipedes, scarab beetles and cockroaches, 3–10 for termites). DNA for clone libraries was extracted using a bead-beating protocol with subsequent phenol–chloroform purification (Paul et al., 2012). DNA for amplicon libraries was purified using the DNeasy soil kit (Qiagen), which also includes mechanical disruption with zirconium beads, according to the manufacturer's instructions.

Clone libraries of 16S rRNA genes

Archaeal 16S rRNA genes were amplified as previously described (Paul et al., 2012), using the archaea-specific forward primer Ar109f (5'-AMDGCTCAGTAACACGT-3') with either the archaea-specific reverse primer Ar912r (5'-CTCCCCGCCAATTCTTTA-3') or the prokaryote-specific reverse primer 1490R (5'-GGHTACCTTGTTACGACTT-3'). The PCR products were purified using the MinElute PCR Purification Kit (Qiagen) and cloned using the pGEM-T vector kit (Promega). For each library, between 20 and 50 clones with correctly sized inserts were bidirectionally sequenced with M13 primers on a capillary sequencer at GATC-Biotech (Konstanz, Germany).

Amplicon libraries of 16S rRNA genes for next generation sequencing

Barcoded 16S rRNA amplicons were generated in two rounds of PCR. In the first round, 16S rRNA genes were amplified using primers Ar109F and 1490R (see above) tagged with M13 sequences at the 5' end (M13F 5'-TGTAACGACGCGCCAGT-3'; M13R 5'-GGAAACAGCTATGACCATG-3'). A 5' block (5'-NH₄-C₆) was added to each primer to ensure that no untagged amplicons were carried over into the second PCR. In the second round, samples were multiplexed by attaching unique barcodes (16-mers) to each end of the

amplicons using bar-coded M13 forward and reverse primers (Pacific Biosciences).

In both rounds, the PCR conditions followed standard PacBio amplicon generation protocols, except that the HiFi Hot Start DNA Polymerase (Roche Life Science) was replaced with Herculanase (Agilent). Round 1: initial denaturation step (92°C for 2 min), followed by 35 cycles of denaturation (94°C for 20 s), annealing (52°C for 30 s), and extension (68°C for 45 s), and a final extension step (68°C for 7 min). Round 2: initial denaturation step (95°C for 3 min), followed by 12 cycles of denaturation (95°C for 30 s), annealing (57°C for 30 s), and extension (72°C for 1 min), and a final extension step (72°C for 7 min).

The barcoded amplicons were purified using AMPure PB beads (Beckman Coulter) following the manufacturer's protocol, pooled at equimolar concentrations, and ligated with SMRTbell adapters following standard PacBio library preparation protocols. The library was sequenced on a Pacific Biosciences Sequel II platform at the Dresden Genome Center (DGC), Dresden, Germany, using one SMRT 8 M cell with the Sequel II Binding Kit 2.1 containing the Sequel Polymerase 2.0 and with a movie length of 600 min. Circular consensus (CCS) reads were generated using the CCS v. 6.4.0 Bioconda package (pbbioconda, Pacific Biosciences) (Grüning et al., 2018).

Read curation and taxonomic classification PacBio amplicons

Read curation followed the pipeline of Martijn et al. (2017) with adaptations for single 16S rRNA gene sequencing. Briefly, the base-calling confidence of the raw CCS reads was assessed, and sequences with associated read quality scores below 0.99 were removed from the dataset. Sequences were curated using *mothur* software (Schloss, 2020), first by demultiplexing barcoded amplicons with the *trim.seqs* command and sorting the samples by host species, followed by removal of primer sequences. Chimeras were removed using the *UCHIME* package (Edgar et al., 2011) integrated in *mothur* with our curated reference database (see below). Quality-trimmed reads were clustered into operational taxonomic units (OTUs) at 97% sequence identity level using the *VSEARCH* tool (Rognes et al., 2016), again using our reference database as a template.

Phylogenetic analysis of 16S rRNA genes

Sequences were imported into the ARB-SILVA database (v. 138¹) using the *ARB* software package (v. 7.0²), aligned with the SINA Aligner (v1.2.12) (Ludwig et al., 2004; Pruesse et al., 2012; Yilmaz et al., 2014), and placed into the phylogenetic framework of the guide tree using the *ARB* parsimony tool. The overall alignment was manually improved using the alignment editor integrated in *ARB*, taking into account secondary structures.

¹ <http://www.arb-silva.de>

² <http://www.arb-home.de>

Multiple sequence alignments comprising representative sequences of different archaeal orders or classes were exported with appropriate outgroup sequences. Maximum-likelihood trees were inferred using IQ-TREE 2 with the GTR + I + G4 substitution model suggested by the ModelFinder tool (Kalyaanamoorthy et al., 2017; Minh et al., 2020). Node support was assessed using the Shimodaira–Hasegawa approximate-likelihood ratio test (SH-aLRT) and ultrafast bootstrap analysis (Guindon et al., 2010; Hoang et al., 2018).

The Dictyopteran Gut Microbiota Reference Database (DictDb)

In this study, our in-house 16S rRNA reference database was expanded to include archaeal sequences from both host-associated and environmental samples. The current iteration of the Dictyopteran Gut Microbiota Reference Database (DictDb v. 5.1 Archaea) was built upon the framework of the latest release (v. 138.1) of the Silva 16S rRNA database (Quast et al., 2013) and includes only archaea. An extension of the previously published DictDb v. 3.0 (Mikaelyan et al., 2015), which covers only bacterial sequences, will be introduced in an upcoming publication. The curated database was further enriched with near-full-length 16S rRNA sequences from studies targeting archaeal diversity in arthropod guts, both from our research group and from the literature. These include the 16S rRNA gene sequences of our MAGs and metagenomes from termite guts (Hervé et al., 2020), and representative sequences obtained in the present study. We also included the curated 16S rRNA sequences provided by SBDI Sativa (Lundin and Andersson, 2021) to establish robust links to the GTDB taxonomy and the genome-based phylogenies. The curated taxon-specific trees of DictDb were used as sources for the reference alignment and taxonomy files for the analysis of next generation sequencing data with mothur (see above). The database was further enriched by adding sequences that are not included in the reference trees (mostly shorter sequences from this and previous studies) using the parsimony tool implemented in Arb.

Genome sequencing

High-molecular-weight DNA of pure cultures was isolated with the DNAEasy Blood & Tissue Kit (Qiagen) following the manufacturer's protocol. The quality of isolated DNA was first checked by agarose gel electrophoresis and validated using an Agilent Bioanalyzer 2100 and the Agilent DNA 12000 kit as recommended by the manufacturer (Agilent Technologies, Waldbronn, Germany). The concentration and purity of the isolated DNA was first estimated with a Nanodrop ND-1000 instrument (PeqLab Erlangen, Germany), and the exact concentration was determined using the Qubit® dsDNA HS Assay kit as recommended by the manufacturer (Life Technologies GmbH, Darmstadt, Germany). Illumina sequencing libraries were prepared using the Nextera XT DNA Sample Preparation kit. To assess the quality and size of the libraries, samples were run on an Agilent Bioanalyzer 2100 using the Agilent High Sensitivity

DNA kit according to the manufacturer's instructions. DNA concentration of the libraries was determined using the Qubit® dsDNA HS Assay kit (Life Technologies GmbH). The libraries were sequenced using a MiSeq system and the reagent kit v3 with 600 cycles as recommended by the manufacturer (Illumina, San Diego, CA, USA). Quality control and quality-filtering of the generated Illumina reads were performed with FastQC v0.11.5 (Andrews, 2010) and Trimmomatic v0.39 (Bolger et al., 2014) using default parameters, respectively. Genomes were assembled with the SPAdes genome assembler software v3.15.2 with default parameters (Bankevich et al., 2012). The quality of the *de novo* assembly was validated using Qualimap v2.2.1 (García-Alcalde et al., 2012).

The genomes of strains Hf6, Ac7, Am2, and Es2 were additionally sequenced using Nanopore technology. Libraries were prepared with 1.5 µg high-molecular-weight DNA using the Ligation Sequencing lit 1D (SQK-LSK109) and the Native Barcode Expansion kit (EXP-NBD104 and EXP-NBD114) as recommended by the manufacturer (Oxford Nanopore Technologies). Libraries were sequenced for 72 h using a MinION device Mk1B and a SpotON Flow Cell R9.4.1 (Oxford Nanopore Technologies). Basecalling and demultiplexing was done with the MinKNOW software and Guppy in high accuracy mode. The generated reads were quality filtered using fastp v0.23.2 (Chen et al., 2018), and the remaining adapters were removed using porechop v0.2.4.³ Hybrids were assembled using Unicycler v0.5.0 with default settings.

Genes were predicted and the assembled genomes were annotated using Prokka v1.14.5 (Seemann, 2014) with default settings.

Phylogenomic analysis

Genomes were classified using the GTDB toolkit (GTDB-Tk v2.3.0) with GTDB release 214 as reference (Chaumeil et al., 2022). The alignment of 53 archaeal marker genes generated by the GTDB toolkit was used to infer a phylogenomic tree with IQ-TREE 2 under the LG + F + I + G4 model suggested by modelfinder. Branch support was assessed by ultrafast bootstrap approximation (1,000 replicates). For rank normalization, RED values were calculated from the annotated tree according to Parks et al. (2018) using PhyloRank (v. 1.12⁴). The archaeal phylogenomics tree (Figure 1) was rendered using iTOL v. 6.8 and edited in Inkscape v. 1.0.1 (Letunic and Bork, 2021).

Data availability statement

Newly obtained representative OTU sequences were submitted to NCBI GenBank under the accession numbers OP851801–OP852117; OQ724653–OQ724818; OR354372–OR354382,

³ <https://github.com/rrwick/Porechop>

⁴ <https://github.com/dparks1134/PhyloRank>

and OR451225–OR451228. Clone library sequences from this study were submitted under the accession numbers OP713915–OP714075 and OR449907–OR449908. Binned small subunit (SSU) sequences extracted from MAGs were submitted under the accession numbers OQ730111–OQ730154; OR140526–OR140534, and OR359878–OR359882. The accession numbers of the new isolates and MAGs are listed in **Supplementary Table 2**. The Dictyopteran gut reference database (DictDb v. 5.1 Archaea) as Arb file and the accompanying mothur reference files are available on GitHub: <https://github.com/brunelab/databases/>.

Ethics statement

The manuscript presents research on animals that do not require ethical approval for their study.

Author contributions

EP: Conceptualization, Resources, Investigation, Data curation, Formal analysis, Validation, Visualization, Writing – original draft, Writing – review and editing. JON: Conceptualization, Resources, Investigation, Data curation, Formal analysis, Validation, Visualization, Writing – original draft. JMKS: Conceptualization, Methodology, Investigation, Data curation, Formal analysis, Validation, Visualization, Writing – review and editing. USM: Data curation, Formal analysis, Visualization, Writing – review and editing. VH: Data curation, Formal analysis, Visualization, Writing – review and editing. CD: Data curation, Formal analysis, Visualization, Writing – review and editing. KL: Investigation, Formal analysis, Writing – review and editing. LM: Investigation, Formal analysis, Writing – review and editing. KP: Methodology, Investigation, Formal analysis, Writing – review and editing. AP: Methodology, Data curation, Writing – review and editing. TK-R: Data curation, Formal analysis, Writing – review and editing. EM: Resources, Investigation, Formal analysis, Writing – review and editing. HIB: Resources, Investigation, Formal analysis, Writing – review and editing. CF: Resources, Writing – review and editing. DKN: Resources, Writing – review and editing. RP: Resources, Writing – review and editing. DS-D: Resources, Writing – review and editing. JŠ: Resources, Writing – review and editing. RD: Resources, Methodology, Writing – review and editing. AB: Conceptualization, Funding acquisition, Project administration, Supervision, Resources, Data curation, Formal analysis, Validation, Visualization, Writing – original draft, Writing – review and editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This study

was funded by the Max Planck Society and a grant of the Deutsche Forschungsgemeinschaft (DFG) in the Collaborative Research Center SFB 987. EP, JMKS, and USM received scholarships from the International Max Planck Research School Principles of Microbial Life: From molecules to cells, from cells to interactions (IMPRS- μ Life). JON received a scholarship from the Deutscher Akademischer Austauschdienst (DAAD). CD received a scholarship from the International Max Planck Research School for Environmental, Cellular and Molecular Microbiology (IMPRS-Mic). None of the funding bodies was involved in the design of the study, the collection, analysis, or interpretation of data, or in writing the manuscript.

Acknowledgments

We thank Kiyoto Maekawa (University of Toyama), Christine Nalepa (North Carolina State University), Rudolf H. Scheffrahn (University of Florida), and Gaku Tokuda (University of the Ryukyus) for insect samples, and Karen Brune for correcting the manuscript. We are grateful to the Kenya Wildlife Service (KWS) for the permission to collect termites from Kenya and to Rudolf H. Scheffrahn for identifying *Alyscotermes trestus*. We also thank Mechthild Bömeke and Melanie Heinemann for technical assistance.

Conflict of interest

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.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2023.1281628/full#supplementary-material>

References

- Andrews, S. (2010). *FastQC: a quality control tool for high throughput sequence data*. Cambridge: Babraham Bioinformatics.
- Arora, J., Kinjo, Y., Šobotnik, J., Bušek, A., Clitheroe, C., Stiblik, P., et al. (2022). The functional evolution of termite gut microbiota. *Microbiome* 10:78. doi: 10.1186/s40168-022-01258-3
- 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
- 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
- Borrel, G., Harris, H. M. B., Tottey, W., Mihajlovski, A., Parisot, N., Peyretailade, E., et al. (2012). Genome sequence of “*Candidatus Methanomethylophilus alvus*” Mx1201, a methanogenic archaeon from the human gut belonging to a seventh order of methanogens. *J. Bacteriol.* 194, 6940–6941. doi: 10.1128/JB.01867-12
- Borrel, G., Parisot, N., Harris, H. M. B., Peyretailade, E., Gaci, N., Tottey, W., et al. (2014). Comparative genomics highlights the unique biology of *Methanomassiliicoccales*, a Thermoplasmatales-related seventh order of methanogenic archaea that encodes pyrrolysine. *BMC Genomics* 15:679. doi: 10.1186/1471-2164-15-679
- Bourguignon, T., Lo, N., Dietrich, C., Šobotnik, J., Sidek, S., Roisin, Y., et al. (2018). Rampant host switching shaped the termite gut microbiome. *Curr. Biol.* 28, 649–654. doi: 10.1016/j.cub.2018.01.035
- Brauman, A., Majeed, M. Z., Buatois, B., Robert, A., Pablo, A. L., and Miambi, E. (2015). Nitrous oxide (N₂O) emissions by termites: Does the feeding guild matter? *PLoS One* 10:0144340. doi: 10.1371/journal.pone.0144340
- Brochier-Armanet, C., Boussau, B., Gribaldo, S., and Forterre, P. (2008). Mesophilic Crenarchaeota: Proposal for a third archaeal phylum, the Thaumarchaeota. *Nat. Rev. Microbiol.* 6, 245–252.
- Brune, A. (2014). Symbiotic digestion of lignocellulose in termite guts. *Nat. Rev. Microbiol.* 12, 168–180. doi: 10.1038/nrmicro3182
- Brune, A. (2018). “Methanogens in the digestive tract of termites,” in *(Endo)symbiotic Methanogenic Archaea*, ed. J. Hackstein (Berlin: Springer), doi: 10.1007/978-3-642-13615-3_6
- Brune, A. (2019). “Methanogenesis in the digestive tracts of insects and other arthropods,” in *Handbook of Hydrocarbon and Lipid Microbiology*, eds A. Stams and D. Sousa (Cham: Springer), doi: 10.1007/978-3-540-77587-4_56
- Brune, A., Emerson, D., and Breznak, J. A. (1995). The termite gut microflora as an oxygen sink: Microelectrode determination of oxygen and pH gradients in guts of lower and higher termites. *Appl. Environ. Microbiol.* 61, 2681–2687.
- Campanaro, S., Treu, L., Rodriguez-R, L. M., Kovalovszki, A., Ziels, R. M., Maus, I., et al. (2020). New insights from the biogas microbiome by comprehensive genome-resolved metagenomics of nearly 1600 species originating from multiple anaerobic digesters. *Biotechnol. Biofuels* 13, 1–18. doi: 10.1186/s13068-020-01679-y
- Carrillo-Reyes, J., Celis, L. B., Alatraste-Mondragón, F., Montoya, L., and Razo-Flores, E. (2014). Strategies to cope with methanogens in hydrogen producing UASB reactors: Community dynamics. *Int. J. Hydrogen Energy* 39, 11423–11432. doi: 10.1016/j.ijhydene.2014.05.099
- Chaumeil, P.-A., Mussig, A. J., Hugenholtz, P., and Parks, D. H. (2022). GTDB-Tk v2: memory friendly classification with the genome taxonomy database. *Bioinformatics* 38, 5315–5316. doi: 10.1093/bioinformatics/btac672
- Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). Fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34, i884–i890. doi: 10.1093/bioinformatics/bty560
- Chibani, C. M., Mahnert, A., Borrel, G., Almeida, A., Werner, A., Brugère, J.-F., et al. (2022). A catalogue of 1,167 genomes from the human gut archaeome. *Nat. Microbiol.* 7, 48–61. doi: 10.1038/s41564-021-01020-9
- Choosai, C., Mathieu, J., Hanboonsong, Y., and Jouquet, P. (2009). Termite mounds and dykes are biodiversity refuges in paddy fields in north-eastern Thailand. *Environ. Conserv.* 36, 71–79. doi: 10.1017/S0376892909005475
- Deevong, P., Hattori, S., Yamada, A., Trakulnaleamsai, S., Ohkuma, M., Noparatnaropon, N., et al. (2004). Isolation and detection of methanogens from the gut of higher termites. *Microbes Environ.* 19, 221–226. doi: 10.1264/jsm.2.19.221
- Dighe, A. S., Jangid, K., González, J. M., Pidiyar, V. J., Patole, M. S., Ranade, D. R., et al. (2004). Comparison of 16S rRNA gene sequences of genus *Methanobrevibacter*. *BMC Microbiol.* 4:20. doi: 10.1186/1471-2180-4-20
- Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C., and Knight, R. (2011). UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* 27, 2194–2200. doi: 10.1093/bioinformatics/btr381
- Egert, M., Stingl, U., Bruun, L. D., Pommerenke, B., Brune, A., and Friedrich, M. W. (2005). Structure and topology of microbial communities in the major gut compartments of *Melolontha melolontha* larvae (Coleoptera: Scarabaeidae). *Appl. Environ. Microbiol.* 71, 4556–4566. doi: 10.1128/AEM.71.8.4556-4566.2005
- Egert, M., Wagner, B., Lemke, T., Brune, A., and Friedrich, M. W. (2003). Microbial community structure in midgut and hindgut of the humus-feeding larva of *Pachnoda ephippiata* (Coleoptera: Scarabaeidae). *Appl. Environ. Microbiol.* 69, 6659–6668. doi: 10.1128/AEM.69.11.6659-6668.2003
- Evans, P. N., Boyd, J. A., Leu, A. O., Woodcroft, B. J., Parks, D. H., Hugenholtz, P., et al. (2019). An evolving view of methane metabolism in the Archaea. *Nat. Rev. Microbiol.* 17, 219–232. doi: 10.1038/s41579-018-0136-7
- Feldewert, C., Lang, K., and Brune, A. (2020). The hydrogen threshold of obligately methyl-reducing methanogens. *FEMS Microbiol. Lett.* 367:fnaa137. doi: 10.1093/femsle/fnaa137
- Fenchel, T., and Finlay, B. J. (2018). “Free-living protozoa with endosymbiotic methanogens,” in *(Endo) symbiotic methanogenic archaea*, ed. J. Hackstein (Berlin: Springer), 1–11.
- Fricke, W. F., Seedorf, H., Henne, A., Krüer, M., Liesegang, H., Hedderich, R., et al. (2006). The genome sequence of *Methanosphaera stadtmanae* reveals why this human intestinal archaeon is restricted to methanol and H₂ for methane formation and ATP synthesis. *J. Bacteriol.* 188, 642–658. doi: 10.1128/JB.188.2.642-658.2006
- Friedrich, M. W., Schmitt-Wagner, D., Lueders, T., and Brune, A. (2001). Axial differences in community structure of *Crenarchaeota* and *Euryarchaeota* in the highly compartmentalized gut of the soil-feeding termite *Cubitermes orthognathus*. *Appl. Environ. Microbiol.* 67, 4880–4890. doi: 10.1128/AEM.67.10.4880-4890.2001
- Gaci, N., Borrel, G., Tottey, W., O’Toole, P. W., and Brugère, J. F. (2014). Archaea and the human gut: New beginning of an old story. *World J. Gastroenterol.* 20, 16062–16078. doi: 10.3748/wjg.v20.i43.16062
- 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
- Gijzen, H. J., Broers, C. A. M., Barughare, M., and Stumm, C. K. (1991). Methanogenic bacteria as endosymbionts of the ciliate *Nyctotherus ovalis* in the cockroach hindgut. *Appl. Environ. Microbiol.* 57, 1630–1634. doi: 10.1128/aem.57.6.1630-1634.1991
- Gilroy, R., Ravi, A., Getino, M., Pursley, I., Horton, D. L., Alikhan, N. F., et al. (2021). Extensive microbial diversity within the chicken gut microbiome revealed by metagenomics and culture. *PeerJ* 9, 1–142. doi: 10.7717/peerj.10941
- Grieco, M. A. B., Cavalcante, J. J., Cardoso, A. M., Vieira, R. P., Machado, E. A., Clementino, M. M., et al. (2013). Microbial community diversity in the gut of the South American termite *Cornitermes cumulans* (Isoptera: Termitidae). *Microb. Ecol.* 65, 197–204.
- Grünig, B., Dale, R., Sjödin, A., Rowe, J., Chapman, B. A., Tomkins-Tinch, C. H., et al. (2018). Bioconda: Sustainable and comprehensive software distribution for the life sciences. *Nat. Methods* 15, 475–476. doi: 10.1038/s41592-018-0046-7
- 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.
- Hackstein, J. H. P., and Stumm, C. K. (1994). Methane production in terrestrial arthropods. *Proc. Natl. Acad. Sci. U. S. A.* 91, 5441–5445. doi: 10.1073/pnas.91.12.5441
- Hackstein, J. H. P., and van Alen, T. A. (2018). “Methanogens in the gastro-intestinal tract of animals,” in *(Endo)symbiotic Methanogenic Archaea*, ed. J. H. P. Hackstein (Berlin: Springer), doi: 10.1007/978-3-642-13615-3_8
- Hara, K., Shinzato, N., Oshima, T., and Yamagishi, A. (2004). Endosymbiotic *Methanobrevibacter* species living in symbiotic protists of the termite *Reticulitermes speratus* detected by fluorescent in situ hybridization. *Microbes Environ.* 19, 120–127. doi: 10.1264/jsm.2.19.120
- Hedlund, B. P., Chuvochina, M., Hugenholtz, P., Konstantinidis, K. T., Murray, A. E., Palmer, M., et al. (2022). SeqCode: a nomenclatural code for prokaryotes described from sequence data. *Nat. Microbiol.* 7, 1702–1708. doi: 10.1038/s41564-022-01214-9
- Henderson, G., Cox, F., Ganesh, S., Jonker, A., Young, W., Janssen, P. H., et al. (2015). Rumen microbial community composition varies with diet and host, but a core microbiome is found across a wide geographical range. *Sci. Rep.* 5:14567. doi: 10.1038/srep14567
- Hervé, V., Liu, P., Dietrich, C., Sillam-Dussès, D., Stiblik, P., Šobotnik, J., et al. (2020). Phylogenomic analysis of 589 metagenome-assembled genomes encompassing all major prokaryotic lineages from the gut of higher termites. *PeerJ* 8:e8614.
- Hoang, D. T., Chernomor, O., Von Haeseler, A., Minh, B. Q., and Vinh, L. S. (2018). UFBoot2: Improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.* 35, 518–522. doi: 10.1093/molbev/msx281
- Hoedt, E. C., Parks, D. H., Volmer, J. G., Rosewarne, C. P., Denman, S. E., McSweeney, C. S., et al. (2018). Culture- and metagenomics-enabled analyses of the *Methanosphaera* genus reveals their monophyletic origin and differentiation according to genome size. *ISME J.* 12, 2942–2953. doi: 10.1038/s41396-018-0225-7

- Huang, X. D., Martinez-Fernandez, G., Padmanabha, J., Long, R., Denman, S. E., and McSweeney, C. S. (2016). Methanogen diversity in indigenous and introduced ruminant species on the Tibetan plateau. *Archaea* 2016:5916067. doi: 10.1155/2016/5916067
- Iino, T., Tamaki, H., Tamazawa, S., Ueno, Y., Ohkuma, M., Suzuki, K. I., et al. (2013). Candidatus *Methanogranum caenicola*: A novel methanogen from the anaerobic digested sludge, and proposal of *Methanomassiliicoccaceae* fam. nov. and *Methanomassiliicoccales* ord. nov., for a methanogenic lineage of the class *Thermoplasmata*. *Microbes Environ.* 28, 244–250. doi: 10.1264/jsme2.ME12189
- Inoue, J. I., Noda, S., Hongoh, Y., Ui, S., and Ohkuma, M. (2008). Identification of endosymbiotic methanogen and ectosymbiotic spirochetes of gut protists of the termite *Coptotermes formosanus*. *Microbes Environ.* 23, 94–97. doi: 10.1264/jsme2.23.94
- Inward, D., Beccaloni, G., and Eggleston, P. (2007). Death of an order: A comprehensive molecular phylogenetic study confirms that termites are eusocial cockroaches. *Biol. Lett.* 3, 331–335. doi: 10.1098/rsbl.2007.0102
- Janssen, P. H., and Kirs, M. (2008). Structure of the archaeal community of the rumen. *Appl. Environ. Microbiol.* 74, 3619–3625. doi: 10.1128/AEM.02812-07
- Ji, R., and Brune, A. (2006). Nitrogen mineralization, ammonia accumulation, and emission of gaseous NH_3 by soil-feeding termites. *Biogeochemistry* 78, 267–283. doi: 10.1007/s10533-005-4279-z
- Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., Von Haeseler, A., and Jermini, L. S. (2017). ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nat. Methods* 14, 587–589. doi: 10.1038/nmeth.4285
- Khomyakova, M. A., Merkel, A. Y., Mamiy, D. D., Klyukina, A. A., and Slobodkin, A. I. (2023). Phenotypic and genomic characterization of *Bathyrarchaeum tardum* gen. nov., sp. nov., a cultivated representative of the archaeal class *Bathyrarchaeia*. *Front. Microbiol.* 14:1214631. doi: 10.3389/fmicb.2023.1214631
- Köhler, T., Dietrich, C., Scheffrahn, R. H., and Brune, A. (2012). High-resolution analysis of gut environment and bacterial microbiota reveals functional compartmentation of the gut in wood-feeding higher termites (*Nasutitermes* spp.). *Appl. Environ. Microbiol.* 78, 4691–4701. doi: 10.1128/aem.00683-12
- Lang, K., Schuldes, J., Klingl, A., Poehlein, A., Daniel, R., and Brune, A. (2015). New mode of energy metabolism in the seventh order of methanogens as revealed by comparative genome analysis of “*Candidatus* *Methanoplasma termitum*.” *Appl. Environ. Microbiol.* 81, 1338–1352. doi: 10.1128/AEM.03389-14
- Leadbetter, J. R., and Breznak, J. A. (1996). Physiological ecology of *Methanobrevibacter cuticularis* sp. nov. and *Methanobrevibacter curvatus* sp. nov., isolated from the hindgut of the termite *Reticulitermes flavipes*. *Appl. Environ. Microbiol.* 62, 3620–3631. doi: 10.1128/aem.62.10.3620-3631.1996
- Leadbetter, J. R., Crosby, L. D., and Breznak, J. A. (1998). *Methanobrevibacter filiformis* sp. nov., a filamentous methanogen from termite hindguts. *Arch. Microbiol.* 169, 287–292.
- Lee, M. J., Schreurs, P. J., Messer, A. C., and Zinder, S. H. (1987). Association of methanogenic bacteria with flagellated protozoa from a termite hindgut. *Curr. Microbiol.* 15, 337–341. doi: 10.1007/BF01577591
- Letunic, I., and Bork, P. (2021). Interactive tree of life (iTOL) v5: An online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* 49, W293–W296. doi: 10.1093/nar/gkab301
- Lind, A. E., Lewis, W. H., Spang, A., Guy, L., Embley, T. M., and Ettema, T. J. G. (2018). Genomes of two archaeal endosymbionts show convergent adaptations to an intracellular lifestyle. *ISME J.* 12, 2655–2667. doi: 10.1038/s41396-018-0207-9
- Loh, H. Q., Hervé, V., and Brune, A. (2021). Metabolic potential for reductive acetogenesis and a novel energy-converting [NiFe] hydrogenase in *Bathyrarchaeia* from termite guts – a genome-centric analysis. *Front. Microbiol.* 11:635786. doi: 10.3389/fmicb.2020.635786
- Ludwig, W., Strunk, O., Westram, R., Richter, L., Meier, H., Yadukumar, A., et al. (2004). ARB: A software environment for sequence data. *Nucleic Acids Res.* 32, 1363–1371. doi: 10.1093/nar/gkh293
- Lundin, D., and Andersson, A. (2021). SBDI Sativa curated 16S GTDB database. *SciLifeLab*. Dataset. doi: 10.17044/scilifelab.14869077.v3
- Lwin, K. O., and Matsui, H. (2014). Comparative analysis of the methanogen diversity in horse and pony by using mcrA gene and archaeal 16S rRNA gene clone libraries. *Archaea* 2014, 1–11. doi: 10.1155/2014/483574
- Majeed, M. Z., Miambi, E., Barois, I., Randriamanantsoa, R., Blanchart, E., and Brauman, A. (2014). Contribution of white grubs (Scarabaeidae: Coleoptera) to N_2O emissions from tropical soils. *Soil Biol. Biochem.* 75, 37–44. doi: 10.1016/j.soilbio.2014.03.025
- Martijn, J., Lind, A. E., Spiers, I., Juzokaite, L., Bunikis, I., Vinnere Pettersson, O., et al. (2017). Amplicon sequencing of the 16S-ITS-23S rRNA operon with long-read technology for improved phylogenetic classification of uncultured prokaryotes. *BioRxiv* [Preprint]. doi: 10.1101/234690
- Meng, J., Xu, J., Qin, D., He, Y., Xiao, X., and Wang, F. (2014). Genetic and functional properties of uncultivated MCG archaea assessed by metagenome and gene expression analyses. *ISME J.* 8, 650–659. doi: 10.1038/ismej.2013.174
- Miambi, E., Jusselme, T. M. D., Châtelliers, C. C., des, Robert, A., Delort, A., et al. (2022). Potential gross and net N_2O production by the gut of different termite species are related to the abundance of nitrifier and denitrifier groups. *Front. Microbiomes* 1:1017006. doi: 10.3389/frmbi.2022.1017006
- Mikaelian, A., Köhler, T., Lampert, N., Rohland, J., Boga, H., Meuser, K., et al. (2015). Classifying the bacterial gut microbiota of termites and cockroaches: A curated phylogenetic reference database (DictDb). *Syst. Appl. Microbiol.* 38, 472–482. doi: 10.1016/j.syapm.2015.07.004
- Miller, T. L., and Wolin, M. J. (1985). *Methanospaera stadmaniae* gen. nov., sp. nov.: a species that forms methane by reducing methanol with hydrogen. *Arch. Microbiol.* 141, 116–122. doi: 10.1007/BF00423270
- Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., Von Haeseler, A., et al. (2020). IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* 37, 1530–1534. doi: 10.1093/molbev/msaa015
- Müller, N., Timmers, P., Plugge, C. M., Stams, A. J., and Schink, B. (2018). “Syntrophy in methanogenic degradation,” in (*Endo*)symbiotic methanogenic archaea, ed. J. H. Hackstein (Heidelberg: Springer), doi: 10.1007/978-3-642-13615-3
- Mwabvu, T. (2005). The density and distribution of millipedes on termite mounds in miombo woodland, Zimbabwe. *Afr. J. Ecol.* 6, 101–103. doi: 10.1016/s1468-1641(10)60322-2
- Nalepa, C. A., Bignell, D. E., and Bandi, C. (2001). Detritivory, coprophagy, and the evolution of digestive mutualisms in Dictyoptera. *Insectes Soc.* 48, 194–201. doi: 10.1007/PL00001767
- Ngugi, D. K., and Brune, A. (2012). Nitrate reduction, nitrous oxide formation, and anaerobic ammonia oxidation to nitrite in the gut of soil-feeding termites (*Cubitermes* and *Ophiotermes* spp.). *Environ. Microbiol.* 14, 860–871. doi: 10.1111/j.1462-2920.2011.02648.x
- Ngugi, D. K., Ji, R., and Brune, A. (2011). Nitrogen mineralization, denitrification, and nitrate ammonification by soil-feeding termites: A ^{15}N -based approach. *Biogeochemistry* 103, 355–369. doi: 10.1007/s10533-010-9478-6
- Odelson, D. A., and Breznak, J. A. (1985). Nutrition and growth characteristics of *Trichomitopsis termitidis*, a cellulolytic protozoan from termites. *Appl. Environ. Microbiol.* 49, 614–621. doi: 10.1128/aem.49.3.614-621.1985
- Ohkuma, M., Noda, S., Horikoshi, K., and Kudo, T. (1995). Phylogeny of symbiotic methanogens in the gut of the termite *Reticulitermes speratus*. *FEMS Microbiol. Lett.* 134, 45–50. doi: 10.1016/0378-1097(95)00379-J
- Ohkuma, M., Noda, S., and Kudo, T. (1999). Phylogenetic relationships of symbiotic methanogens in diverse termites. *FEMS Microbiol. Lett.* 171, 147–153. doi: 10.1016/S0378-1097(98)00593-X
- Oren, A., and Garrity, G. M. (2021). Valid publication of the names of forty-two phyla of prokaryotes. *Int. J. Syst. Evol. Microbiol.* 71:005056.
- Parks, D. H., Chuvochina, M., Rinke, C., Mussig, A. J., Chaumeil, P., and Hugenholtz, P. (2021). GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. *Nucleic Acids Res.* 2021, 1–10. doi: 10.1093/nar/gkab776
- Parks, D. H., Chuvochina, M., Waite, D. W., Rinke, C., Skarshewski, A., Chaumeil, P. A., et al. (2018). A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat. Biotechnol.* 36:996. doi: 10.1038/nbt.4229
- Paul, K., Nonoh, J. O., Mikulski, L., and Brune, A. (2012). “*Methanoplasmatales*,” *Thermoplasmatales*-related archaea in termite guts and other environments, are the seventh order of methanogens. *Appl. Environ. Microbiol.* 78, 8245–8253. doi: 10.1128/aem.02193-12
- Pester, M., and Brune, A. (2006). Expression profiles of fhs (FTHFS) genes support the hypothesis that spirochaetes dominate reductive acetogenesis in the hindgut of lower termites. *Environ. Microbiol.* 8, 1261–1270. doi: 10.1111/j.1462-2920.2006.01020.x
- Pester, M., Schleper, C., and Wagner, M. (2011). The Thaumarchaeota: An emerging view of their phylogeny and ecophysiology. *Curr. Opin. Microbiol.* 14, 300–306. doi: 10.1016/j.mib.2011.04.007
- Poehlein, A., Schneider, D., Soh, M., Daniel, R., and Seedorf, H. (2018). Comparative genomic analysis of members of the genera *Methanospaera* and *Methanobrevibacter* reveals distinct clades with specific potential metabolic functions. *Archaea* 2018:7609847. doi: 10.1155/2018/7609847
- Pruesse, E., Peplies, J., and Glöckner, F. O. (2012). SINA: Accurate high-throughput multiple sequence alignment of ribosomal RNA genes. *Bioinformatics* 28, 1823–1829. doi: 10.1093/bioinformatics/bts252
- Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., et al. (2013). The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Res.* 41, 590–596. doi: 10.1093/nar/gks1219
- Regensbogenova, M., McEwan, N. R., Javorsky, P., Kisidayova, S., Michalowski, T., Newbold, C. J., et al. (2004). A re-appraisal of the diversity of the methanogens associated with the rumen ciliates. *FEMS Microbiol. Lett.* 238, 307–313. doi: 10.1016/j.femsle.2004.07.049

- Rinke, C., Chuvochina, M., Mussig, A. J., Chaumeil, P.-A., Davin, A. A., Waite, D. W., et al. (2021). A standardized archaeal taxonomy for the Genome Taxonomy Database. *Nat. Microbiol.* 6, 946–959. doi: 10.1038/s41564-021-00918-8
- Rognes, T., Flouri, T., Nichols, B., Quince, C., and Mahé, F. (2016). VSEARCH: A versatile open source tool for metagenomics. *PeerJ* 2016, 1–22. doi: 10.7717/peerj.2584
- Schauer, C., Thompson, C. L., and Brune, A. (2012). The bacterial community in the gut of the cockroach *Shelfordella lateralis* reflects the close evolutionary relatedness of cockroaches and termites. *Appl. Environ. Microbiol.* 78, 2758–2767. doi: 10.1128/aem.07788-11
- Schloss, P. D. (2020). Reintroducing mothur: 10 Years Later. *Appl. Environ. Microbiol.* 86, 1–13.
- Seedorf, H., Dreisbach, A., Hedderich, R., Shima, S., and Thauer, R. K. (2004). F₄₂₀H₂ oxidase (FprA) from *Methanobrevibacter arboriphilus*, a coenzyme F₄₂₀-dependent enzyme involved in O₂ detoxification. *Arch. Microbiol.* 182, 126–137. doi: 10.1007/s00203-004-0675-3
- Seemann, T. (2014). Prokka: Rapid prokaryotic genome annotation. *Bioinformatics* 30, 2068–2069. doi: 10.1093/bioinformatics/btu153
- Shi, Y., Huang, Z., Han, S., Fan, S., and Yang, H. (2015). Phylogenetic diversity of Archaea in the intestinal tract of termites from different lineages. *J. Basic Microbiol.* 55, 1021–1028.
- Shinzato, N., Watanabe, I., Meng, X. Y., Sekiguchi, Y., Tamaki, H., Matsui, T., et al. (2007). Phylogenetic analysis and fluorescence in situ hybridization detection of archaeal and bacterial endosymbionts in the anaerobic ciliate *Trimyema compressum*. *Microb. Ecol.* 54, 627–636. doi: 10.1007/s00248-007-9218-1
- Söllinger, A., Schwab, C., Weinmaier, T., Loy, A., Tveit, A. T., Schleper, C., et al. (2016). Phylogenetic and genomic analysis of *Methanomassiliicoccales* in wetlands and animal intestinal tracts reveals clade-specific habitat. *FEMS Microbiol. Ecol.* 92, 1–12. doi: 10.1093/femsec/fiv149
- Sprenger, W. W., Hackstein, J. H. P., and Keltjens, J. T. (2007). The competitive success of *Methanomicrococcus blatticola*, a dominant methylotrophic methanogen in the cockroach hindgut, is supported by high substrate affinities and favorable thermodynamics. *FEMS Microbiol. Ecol.* 60, 266–275. doi: 10.1111/j.1574-6941.2007.00287.x
- Sprenger, W. W., Van Belzen, M. C., Rosenberg, J., Hackstein, J. H. P., and Keltjens, J. T. (2000). *Methanomicrococcus blatticola* gen. nov., sp. nov., a methanol- and methylamine-reducing methanogen from the hindgut of the cockroach *Periplaneta americana*. *Int. J. Syst. Evol. Microbiol.* 50, 1989–1999. doi: 10.1099/00207713-50-6-1989
- Šustr, V., Chroňáková, A., Semanová, S., Tajovský, K., and Šimek, M. (2014). Methane production and methanogenic archaea in the digestive tracts of millipedes (Diplopoda). *PLoS One* 9:e102659. doi: 10.1371/journal.pone.0102659
- Tholen, A., and Brune, A. (2000). Impact of oxygen on metabolic fluxes and in situ rates of reductive acetogenesis in the hindgut of the wood-feeding termite *Reticulitermes flavipes*. *Environ. Microbiol.* 2, 436–449. doi: 10.1046/j.1462-2920.2000.00127.x
- Tholen, A., Pester, M., and Brune, A. (2007). Simultaneous methanogenesis and oxygen reduction by *Methanobrevibacter cuticularis* at low oxygen fluxes. *FEMS Microbiol. Ecol.* 62, 303–312. doi: 10.1111/j.1574-6941.2007.00390.x
- Thomas, C. M., Quémener, E. D., Gribaldo, S., and Borrel, G. (2022). Factors shaping the abundance and diversity of the gut archaeome across the animal kingdom. *Nat. Commun.* 13:3358. doi: 10.1038/s41467-022-31038-4
- Thomas, C. M., Taib, N., Gribaldo, S., and Borrel, G. (2021). Comparative genomic analysis of *Methanimicrococcus blatticola* provides insights into host-adaptation in archaea and the evolution of methanogenesis. *ISME Commun.* 1:47. doi: 10.1038/s43705-021-00050-y
- Tokura, M., Ohkuma, M., and Kudo, T. (2000). Molecular phylogeny of methanogens associated with flagellated protists in the gut and with the gut epithelium of termites. *FEMS Microbiol. Ecol.* 33, 233–240. doi: 10.1016/S0168-6496(00)00065-9
- Tokura, M., Tajima, K., and Ushida, K. (1999). Isolation of *Methanobrevibacter* sp. as a ciliate-associated ruminal methanogen. *J. Gen. Appl. Microbiol.* 45, 43–47. doi: 10.2323/jgam.45.43
- Treitli, S., Hanousková, P., Benes, V., Brune, A., Čepička, I., and Hampl, V. (2023). Hydrogenotrophic methanogenesis is the key process in the obligately syntrophic consortium of the anaerobic amoeba *Pelomyxa schiedti*. *ISME J.* 17, 1884–1894. doi: 10.1038/s41396-023-01499-6
- van Hoek, A. H. A. M., van Alen, T. A., Sprakel, V. S. I., Leunissen, J. A. M., Brigge, T., Vogels, G. D., et al. (2000). Multiple acquisition of methanogenic archaeal symbionts by anaerobic ciliates. *Mol. Biol. Evol.* 17, 251–258. doi: 10.1093/oxfordjournals.molbev.a026304
- Volmer, J. G., Soo, R. M., Evans, P. N., Hoedt, E. C., Alsina, A. L. A., Woodcroft, B. J., et al. (2023). Isolation and characterisation of novel *Methanocorpusculum* species indicates the genus is ancestrally host - associated. *BMC Biol.* 21:59. doi: 10.1186/s12915-023-01524-2
- Weil, M., Hoff, K. J., Meißner, W., Schäfer, F., Söllinger, A., Wang, H., et al. (2021). Full genome sequence of a *Methanomassiliicoccales* representative enriched from peat soil. *Microbiol. Resour. Annu.* 10, e443–e421.
- Whitman, W. B., Chuvochina, M., Hedlund, B. P., Hugenholtz, P., Konstantinidis, K. T., Murray, A. E., et al. (2022). Development of the SeqCode: A proposed nomenclatural code for uncultivated prokaryotes with DNA sequences as type. *Syst. Appl. Microbiol.* 45:126305. doi: 10.1016/j.syapm.2022.126305
- Xie, F., Jin, W., Si, H., Yuan, Y., Tao, Y., Liu, J., et al. (2021). An integrated gene catalog and over from the gastrointestinal microbiome of ruminants. *Microbiome* 9, 1–20.
- Yilmaz, P., Parfrey, L. W., Yarza, P., Gerken, J., Pruesse, E., Quast, C., et al. (2014). The SILVA and “all-species Living Tree Project (LTP)” taxonomic frameworks. *Nucleic Acids Res.* 42, 643–648. doi: 10.1093/nar/gkt1209
- Youngblut, N. D., Reischer, G. H., Dauser, S., Maisch, S., Walzer, C., Stalder, G., et al. (2021). Vertebrate host phylogeny influences gut archaeal diversity. *Nature microbiology* 6, 1443–1454.
- Zinke, L. A., Evans, P. N., Santos-Medellin, C., Schroeder, A. L., Parks, D. H., Varner, R. K., et al. (2021). Evidence for non-methanogenic metabolisms in globally distributed archaeal clades basal to the *Methanomassiliicoccales*. *Environ. Microbiol.* 23, 340–357.



OPEN ACCESS

EDITED BY

Chuanlun Zhang,
Southern University of Science and
Technology, China

REVIEWED BY

Micaela Cerletti,
National University of Mar del Plata, Argentina
Dina Grohmann,
University of Regensburg, Germany
Xiaoyun Yang,
Southern University of Science and
Technology, China

*CORRESPONDENCE

Jörg Soppa

✉ soppa@bio.uni-frankfurt.de

[†]These authors have contributed equally to this work

RECEIVED 21 August 2023

ACCEPTED 09 November 2023

PUBLISHED 29 November 2023

CITATION

Üresin D, Pyper DJ, Borst A, Hadjeras L,
Gelhausen R, Backofen R, Sharma C,
Schwalbe H and Soppa J (2023)
Characterization of the zinc finger μ -protein
HVO_0758 from *Haloferax volcanii*: biological
roles, zinc binding, and NMR solution structure.
Front. Microbiol. 14:1280972.
doi: 10.3389/fmicb.2023.1280972

COPYRIGHT

© 2023 Üresin, Pyper, Borst, Hadjeras,
Gelhausen, Backofen, Sharma, Schwalbe and
Soppa. 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 the zinc finger μ -protein HVO_0758 from *Haloferax volcanii*: biological roles, zinc binding, and NMR solution structure

Deniz Üresin^{1†}, Dennis J. Pyper^{2†}, Andreas Borst¹, Lydia Hadjeras³,
Rick Gelhausen⁴, Rolf Backofen^{4,5}, Cynthia Sharma³,
Harald Schwalbe² and Jörg Soppa^{1*}

¹Institute for Molecular Biosciences, Goethe University, Frankfurt, Germany, ²Institute for Organic Chemistry and Chemical Biology, Center for Biomolecular Magnetic Resonance (BMRZ), Goethe University, Frankfurt, Germany, ³Institute of Molecular Infection Biology, University of Würzburg, Würzburg, Germany, ⁴Bioinformatics Group, Department of Computer Science, University of Freiburg, Freiburg, Germany, ⁵Signalling Research Centres BIOSS and CIBSS, University of Freiburg, Freiburg, Germany

It is increasingly recognized that very small proteins (μ -proteins) are ubiquitously found in all species of the three domains of life, and that they fulfill important functions. The halophilic archaeon *Haloferax volcanii* contains 282 μ -proteins of less than 70 amino acids. Notably, 43 of these contain two C(P)XCG motifs, suggesting their potential to complex a zinc ion. To explore the significance of these proteins, 16 genes encoding C(P)XCG proteins had been deleted, and the majority of mutants exhibited phenotypic differences to the wild-type. One such protein, HVO_2753, was thoroughly characterized in a previous study. In the present study an in-depth analysis of a second protein, HVO_0758, was performed. To achieve this goal, the HVO_0758 protein was produced heterologously in *Escherichia coli* and homologously in *H. volcanii*. The purified protein was characterized using various biochemical approaches and NMR spectroscopy. The findings demonstrated that HVO_0758 is indeed a *bona fide* zinc finger protein, and that all four cysteine residues are essential for folding. The NMR solution structure was solved, revealing that HVO_0758 is comprised of an N-terminal alpha helix containing several positively charged residues and a globular core with the zinc finger domain. The transcriptomes of the HVO_0758 deletion mutant and, for comparison, the HVO_2753 deletion mutant were analyzed with RNA-Seq and compared against that of the wild-type. In both mutants many motility and chemotaxis genes were down-regulated, in agreement to the phenotype of the deletion mutants, which had a swarming deficit. The two *H. volcanii* zinc-finger μ -proteins HVO_0758 and HVO_2753 showed many differences. Taken together, two zinc finger μ -proteins of *H. volcanii* have been characterized intensively, which emerged as pivotal contributors to swarming behavior and biofilm formation.

KEYWORDS

Archaea, *Haloferax volcanii*, small proteins, microproteins, zinc finger, TALON, NMR solution structure, RNA-Seq

Introduction

Until quite recently, small proteins have been overlooked for several reasons. In genome sequencing projects, open reading frames (ORFs) comprising fewer than 100 codons were by default not recognized as potential genes. This was a deliberate strategy to prevent the annotation of a large number of false positive gene predictions. Additionally, very small proteins disappeared in experimental approaches that were optimized for the characterization of normal-sized and large proteins. Furthermore, biochemical characterization of very small proteins faces specific challenges, for example because they often do not exhibit a specific fold in isolated form (Kubatova et al., 2020). Nevertheless, the past few years have unveiled the existence and vital roles of very small proteins across all three domains of life. Notable advancements encompass the establishment of ribosome profiling methods, advances in mass spectrometry (peptidomics), and enhancements in gene/ORF prediction through bioinformatics refinements. Several recent reviews summarize current knowledge about very small proteins in archaea, bacteria, and eukaryotes (Storz et al., 2014; Orr et al., 2020; Cassidy et al., 2021; Steinberg and Koch, 2021; Tharakan and Sawa, 2021; Chen et al., 2022; Gray et al., 2022; Leong et al., 2022; Schlesinger and Elsässer, 2022; Weidenbach et al., 2022; Zhang et al., 2022; Dong et al., 2023; Hassel et al., 2023).

In contrast to the many novel findings the field is so new that a common nomenclature does not exist. Very small proteins are denoted as μ -peptides, peptides, μ -proteins, small proteins, and sORF encoded proteins. In addition, the upper size limit varies from about 50 aa to 100 aa. In this contribution the name “ μ -proteins” will be used, to distinguish these proteins from “small proteins,” which are much smaller than an average sized protein of 300 aa (prokaryotes), but larger than the annotation limit of 100 aa.

Even before the general importance of μ -proteins in all domains of life was recognized, the group of the late Dieter Oesterhelt characterized the “low molecular weight proteome” of the halophilic archaeon *Halobacterium salinarum* (Klein et al., 2007). They identified 380 proteins smaller than 20 kDa, constituting around 9% of the proteome, mostly lacking known functions. Approximately 20 of these proteins contained dual C(P)XCG motifs, suggesting a potential role in binding zinc ions, making them putative single-domain zinc finger proteins.

Zinc finger proteins were first discovered in eukaryotes and were thought for long to be confined to this domain (Maret, 2013). Eukaryotic zinc finger proteins are typically larger proteins housing multiple zinc finger domains. Zinc finger domains are characterized by four amino acids that complex a central zinc ion, e.g., four cysteines or two cysteines and two histidines. The four amino acids are organized into two motifs, CXXC or HXXH (X representing any amino acid), which are separated by a linker of variable length. C4 zinc fingers often display the more elaborate C(P)XCG consensus sequence (P in only one motif).

Zinc fingers are very versatile interaction domains, enabling zinc finger proteins to interact with diverse biomolecules, encompassing DNA, RNA, soluble and membrane proteins (Matthews and Sunde, 2002; Eom et al., 2016). Moreover, engineered zinc finger nucleases have been applied for gene editing (Bak et al., 2018). Structurally, zinc finger proteins have been classified into eight groups (Krishna et al., 2003).

Zinc finger proteins are not confined to eukaryotes, but are also present in prokaryotes. And, as noted above, not all zinc fingers are part of large proteins, small single domain zinc finger proteins also exist. In archaea, zinc finger proteins constitute 8% of small proteins with fewer than 100 amino acids, compared to a mere 1.5% in bacterial small proteins (Tarasov et al., 2008). In a recent study we have generated *in-frame* deletion mutants of 16 genes encoding zinc finger μ -proteins in the haloarchaeon *Haloferax volcanii* (Nagel et al., 2019). 12 of these mutants exhibited distinct phenotypes from the wild-type, illustrating the diverse roles zinc finger μ -proteins play in haloarchaeal biological processes. These mutants did not only show loss-of-function characteristics, but also gain-of-function phenotypes, including improved biofilm formation (Nagel et al., 2019).

One protein, HVO_2753, was characterized in detail, including solving the NMR solution structure (Zahn et al., 2021). HVO_2753 was chosen because it is the only *H. volcanii* zinc finger protein with four C(P)XCG motifs, which are indicative for the formation of two zinc fingers. Surprisingly, a biochemical zinc assay and the NMR zinc titration and structure revealed consistently that HVO_2753 binds only one zinc ion. The second “zinc finger” has apparently lost the ability to bind a zinc ion, because the four cysteines do not point to the center of the “zinc finger.” This result shows that the presence of two C(P)XCG motifs in a protein sequence is highly indicative for the formation of a zinc finger and zinc binding, however, that exceptions exist and experimental verification is important.

Here, we report the detailed characterization of a second *H. volcanii* zinc finger μ -protein, HVO_0758. The protein was selected as its cognate deletion mutant showed interesting pleiotrophic phenotypes, i.e., a late onset of growth in synthetic glycerol medium, a lack of swarming, and enhanced biofilm formation. HVO_0758 was produced homologously in *H. volcanii* and heterologously in *Escherichia coli*, purified, and subjected to comprehensive characterization. Several conserved amino acids were selected and point mutants were generated. Notably, the NMR solution structure was solved with high resolution. Transcriptome comparisons were made between the deletion mutant, the wild-type, and the HVO_2753 deletion mutant. Lastly, characteristic features and the structures of the two in depth characterized proteins HVO_0758 and HVO_2753 were compared.

Results

Characteristic features of HVO_0758 and evolutionary conservation

HVO_0758 was selected because a deletion mutant of the cognate gene exhibited a swarming deficit and a gain of function in biofilm formation, indicating that the protein fulfills important functions (Nagel et al., 2019). HVO_0758 is a protein of 56 amino acids with two C(P)XCG-like motifs, none of which contains a proline at the second position (Figure 1A). It is comprised of an extremely high fraction of charged and hydrophilic amino acids, and thus can be predicted to be involved in many interactions with other biomolecules. Notably, this amino acid composition includes 13 positively charged amino acids (lysine and arginine), and consequently HVO_0758 has an isoelectric point of 7.6. Such alleviated pK_i is very untypical for haloarchaeal proteins, which typically have a high fraction of negatively charged

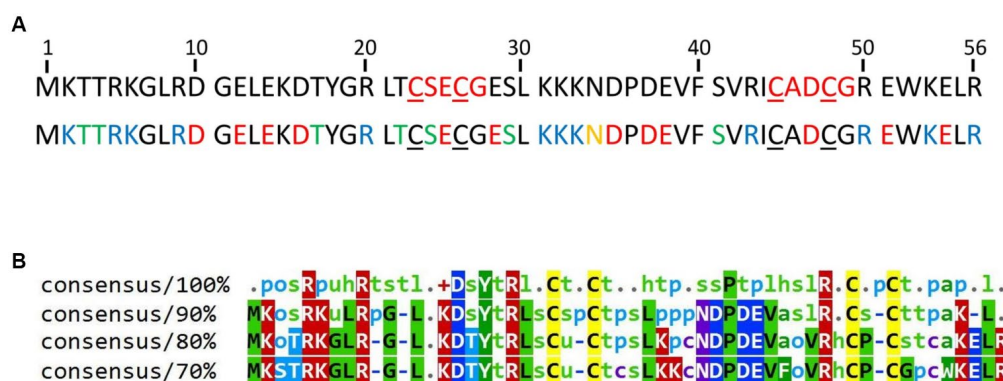


FIGURE 1

(A) Sequence of the protein HVO_0758. The four cysteines forming the zinc binding pocket are underlined. Top: The C(P)XCG motifs are highlighted in red. Bottom: Acidic amino acids are highlighted in red, basic residues in blue, residues with amide side chains in yellow and residues with hydroxyl groups in green. The accession number of the protein in the UniProtKB database is D4GTQ1. (B) Conservation of HVO_0758. A multiple sequence alignment was generated between HVO_0758 and the top 99 hits of a protein blast search. At each position the residues are shown that are, respectively, 100, 90, 80%, or 70% conserved. Lower-case letter legend: a → aromatic, h → hydrophobic, l → aliphatic, o → alcohol, p → polar, s → small (A, C, D, G, N, P, S, T, V), t → turnlike, u → tiny (A, G, S).

amino acids and a very low isoelectric point (Shukla, 2006; Oren, 2013). HVO_0758 is present in more than 200 other species of haloarchaea. However, there are no orthologs in other groups of archaea, in bacteria, or in eukaryotes. A multiple sequence alignment of 100 HVO_0758 orthologs revealed that many positions are highly conserved, in addition to the four cysteines (Figure 1B). Analysis of previous RNA-Seq and dRNA-Seq results (Babski et al., 2016; Laass et al., 2019) revealed that HVO_0758 is expressed during exponential growth under optimal conditions, that the transcript is leaderless and that it has a long 3'-UTR of about 200 nt (Supplementary Figure S1). A Northern blot analysis confirmed that the transcript is considerably longer than the ORF and has a long UTR (see below).

Heterologous production and characterization

In order to analyze the protein with NMR, it was heterologously produced in *E. coli* BL21(DE3) cells. The production was carried out in M9-medium with ^{15}N - NH_4Cl for single-labeled samples or ^{15}N - NH_4Cl and ^{13}C -glucose for double-labeled samples as sole nitrogen and carbon sources. The protein was produced with an N-terminal His₆-tag and a SUMO tag (Malakhov et al., 2004; Marblestone et al., 2006), which enabled the purification by twin-affinity chromatography. The tags were removed by the SUMO protease. Size exclusion chromatography was used as final purification step, resulting in a purity of >95%. The purification steps are visualized in Supplementary Figure S2.

We aimed to characterize the effect of salt on the structure of HVO_0758 from the halophilic archaeon *H. volcanii*, which has an optimal salt concentration of 2.1 M NaCl. We started out with a NaCl concentration of 300 mM. The HSQC spectrum of HVO_0758 has 68 visible signals, although from the sequence only 54 signals are expected (Figure 2A). With 2D ^{15}N -ZZ-exchange experiments (Montelione and Wagner, 1989; Farrow et al., 1994) we could show that these signals stem from two conformations that interconvert slowly, with one of them exhibiting a wide signal dispersion and the

other a narrow, indicating that the protein is present in a structured and an unstructured conformation. This is further supported by the temperature series (Supplementary Figures S3, S4). At high temperatures (328 K) the protein completely unfolds and only the signals of the second conformation remain.

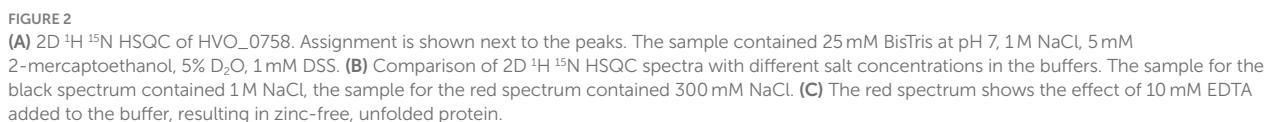
We increased the salt concentration of the protein buffer to 1 M NaCl, which led to the second (unstructured) conformation disappearing, showing that the folding of the protein is dependent on salt concentration (Figure 2B). Subsequent experiments for the investigation of structure and dynamics were thus carried out at 1 M NaCl.

Zinc binding

In order to analyze the effects of zinc binding, all zinc ions were removed from the isolated protein by addition of 10 eq. of EDTA (Figure 2C). Without zinc ions, only signals of the unfolded conformation remained, indicating that zinc ions are necessary for the native folded structure. Addition of zinc ions to the unfolded protein fully refolded the protein.

NMR assignment and TALOS prediction

NMR backbone and side-chain assignment were conducted manually using standard 2D and 3D experiments. The protein was measured with 1 M NaCl for correct folding conditions at 298 K. We were able to assign 99% of the backbone with the only missing signal being glutamate 28%. The assigned ^1H - ^{15}N HSQC is shown in Figure 2A. With the backbone chemical shift assignments, we used TALOS-N (Shen and Bax, 2013) to predict secondary structure elements of the protein. We identified one α -helix ranging from T4 to D10, as well as five β -strands (L13 to K15, L21 to C23, K31 to N34, S41 to C45 and E51 to L55; Figure 3, top panel). Furthermore, we obtained backbone torsion angles from TALOS-N, which were used as constraints in the structure calculation (see below).



We investigated the dynamic properties of HVO_0758 by analyzing T_1 and T_2 relaxation times as well as $\{^1\text{H}^{15}\text{N}\}$ -hetNOE. The relaxation data were used to calculate the Lipari-Szabo order parameter S^2 , providing information about the local flexibility of the protein. Furthermore, we determined the experimental rotational correlation time τ_c to be 4.01 ns (Figure 3).

frontiersin.org

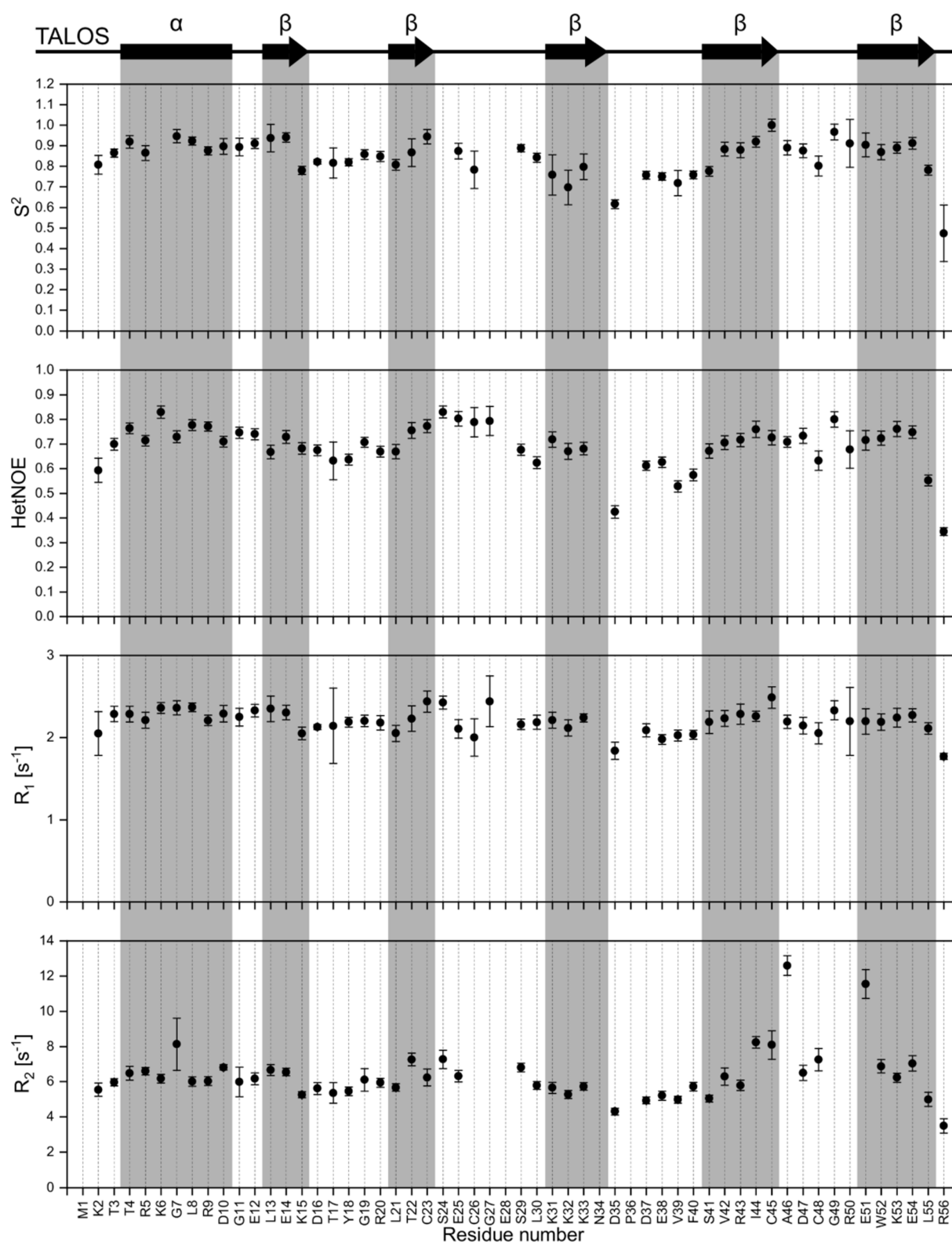


FIGURE 3

Dynamic studies of HVO_0758. On the top, the secondary structure prediction obtained from TALOS-N is shown. Below, the S^2 order parameter (calculated with TENSOR2) is shown, which is based on the experimental relaxation parameters ($^1\text{H}^{15}\text{N}$)-hetNOE, R_1 and R_2 .

Furthermore, we obtained $^3J_{\text{HNH}\alpha}$ coupling constants from a 3D HNHA spectrum which, based on the Karplus equation, can also be used to predict secondary structure formation

(Supplementary Figure S6; Vuister and Bax, 1993). Coupling constants in the range of 3 to 4.5 Hz and 8 to 9 Hz indicate α -helices and β -sheets, respectively.

3D structure calculation

For the structure calculation, a total of 88 backbone torsion angles (φ , ψ) were obtained from TALOS-N and included in the structure calculation as restraints. In addition, 41 $^3J_{\text{HNH}\alpha}$ scalar couplings from a 3D HNHA spectrum were used and 12 hydrogen bonds were defined by manually checking visible cross-peaks in the 3D NOESY-HSQC spectra for potential secondary structure elements (Table 1).

The 20 lowest energy calculated structures are shown as a bundle in Figure 4A. The protein has an N-terminal α -helix with

TABLE 1 List of NOE values.

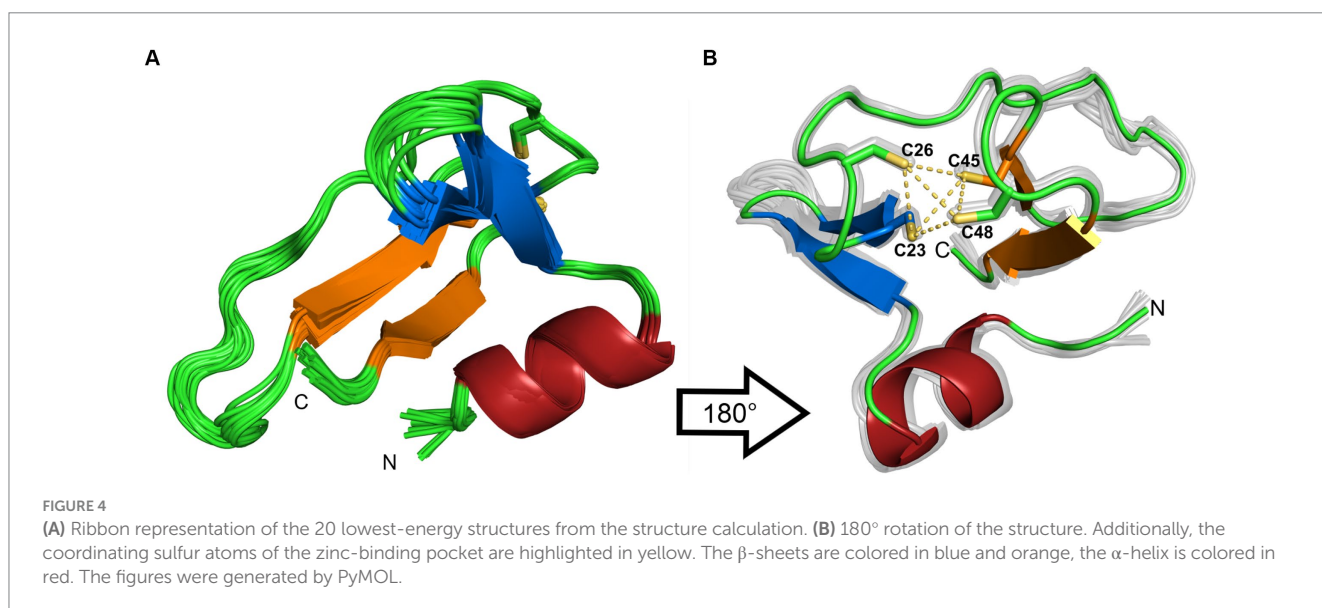
NOE distances	666
Short range ($ i-j \leq 1$)	369
Medium range ($1 < i-j < 5$)	73
Long range ($ i-j \geq 5$)	224
Hydrogen bonds	12
Backbone torsion angles (φ, ψ)	88
Scalar couplings ($^3J(\text{HN}, \text{H}\alpha)$)	41
Lower length of S-S bond in zinc binding pocket	3.65 Å
Upper length of S-S bond in zinc binding pocket	4 Å
Number of S-S restraints in zinc binding pocket	6
Ramachandran score (%)	
Most favored	92.3
Additionally allowed	7.7
Generously allowed	0
Disallowed	0
Average rmsd to mean (Å)	
Backbone	0.36
All atoms	1.06
Restraint violations	0

several positive charges, followed by a twisted antiparallel β -sheet. The middle part of the protein is comprised of a more flexible part that could be a potential third β -strand to the second anti-parallel β -sheet, which is located at the C-terminus. This is assumed due to some observed cross peaks in the NOESY spectra between K31 and I44, as well as K33 and V42. In addition, TALOS-N predicts a β -strand from K31 to N34. Two cysteines are found after the first β -sheet, which together with two more cysteines from the β -turn of the second β -sheet form a well-defined zinc binding pocket, that can be classified as a zinc-ribbon structure (Krishna et al., 2003).

To rule out dimerization, we compared the experimental rotational correlation time τ_c obtained via TENSOR2 (Dosset et al., 2000) with the calculated one from Hydromr (García et al., 2000). Hydromr calculates τ_c based on a provided structure, while TENSOR2 also takes measured relaxation parameters (R_1 , R_2 , hetNOE) into account. The obtained τ_c values are 4.301 ns for Hydromr and 4.074 ns for TENSOR2. Since the difference between them is small, we can conclude that the protein exists as a monomer.

Phenotypes of an in-frame deletion mutant

In a previous study 16 deletion mutants of C(P)XCG protein encoding genes were generated, and their phenotypes were characterized (Nagel et al., 2019). The HVO_0758 deletion mutant exhibited distinctive phenotypes, including a delayed onset of growth in glycerol medium, an inability to swarm, and enhanced biofilm formation in comparison to the wild-type. Prior to further characterizing HVO_0758, we sought to confirm these phenotypes. First, a multiple cycle PCR analysis was used to prove that the deletion was still homozygous (Supplementary Figure S7A). Because *H. volcanii* is highly polyploid, a deletion mutant might still retain a wild-type copy of the genome, potentially reverting to the wild-type over extended storage or cultivation. Furthermore, Northern blot analysis demonstrated the absence of HVO_0758 transcript (Supplementary Figure S7B).



Normally the swarm plate assay is performed in six well plates for 2 days (Nagel et al., 2019). To analyze the behavior of the HVO_0758 deletion mutant upon prolonged incubation, 8 cm Petri dishes were used and the incubation time was considerably extended from 42 h to 192 h (Supplementary Figure 8A). The deletion mutant did not swarm at all for the first 3 days, confirming the lack of swarming during 42 h observed before. However, then the mutant started to swarm with the same velocity as the wild-type. It can be hypothesized that a regulatory step missing in the mutant can be circumvented after prolonged incubation and the chemotaxis pathway can be induced in spite of the missing step.

The enhanced biofilm formation of the HVO_0758 deletion mutant was verified (Supplementary Figure 8B). After 48 h, the mutant exhibited over twofold greater biofilm formation compared to the wild-type.

Initially, replicating the glycerol medium phenotype proved challenging. Eventually, it was uncovered that the phenotype was contingent on the pre-culture's growth medium. If the pre-culture was cultivated in glycerol medium, the mutant culture initiated growth upon inoculation without any lag phase, paralleling the wild-type (Supplementary Figure 8C, left panel). In contrast, if the pre-culture was cultivated in a complex medium, the mutant's lag phase extended by a day relative to the wild-type (Supplementary Figure 8C, right panel).

Next, it was attempted to complement the phenotypes. The phenotype in glycerol medium could be fully complemented (Supplementary Figure S9). In contrast, the biofilm phenotype and the swarming phenotype could not be complemented. We have experienced the partial complementation of multiple phenotypes also with various additional deletion mutants of genes for zinc finger μ -proteins. Possible explanations are discussed below (see Discussion).

Transcriptome comparison between wild-type and mutant Δ HVO_0758

To gain further insight into the potential functions of HVO_0758, an alternative approach involved comparing the transcriptome of the deletion mutant to that of the wild-type. To this end, two cultures of each strain were grown to mid-exponential growth phase ($4\text{--}5 \times 10^8$ cells per ml) in complex medium. Mutant and wild-type grew identically in glucose medium, therefore, indirect effects based on growth rate-differences were prevented. Subsequently, total RNA was extracted, rRNA was depleted and RNA-Seq was used to determine the sequences of the remaining RNAs. The results have been deposited at the Gene Expression Omnibus (GEO)¹ and obtained the accession No. GSE228855. First, it was analyzed that the HVO_0758 transcript was indeed missing in the deletion mutant (Supplementary Figure S10A), which was in accordance with the Northern blot analysis (Supplementary Figure S7B).

All genes displaying more than twofold average transcript level difference between the two strains are listed in Supplementary Table S1

together with, e.g., quantitative transcript level difference, protein name, functional classification, and normalized average counts. A total of 78 genes exhibited more than twofold downregulation in the Δ HVO_0758 mutant, while 29 genes experienced more than twofold upregulation. The high number of 108 differentially regulated genes strongly suggests that HVO_0758 has important biological roles in *H. volcanii*.

In many cases not solitary genes were differentially regulated, but rather clusters of contiguous genes displaying co-regulation. Specifically, 15 clusters of co-regulated genes are represented among the 111 differentially regulated genes. The most prominent cluster was cluster 2 with genes of the large motility/chemotaxis gene cluster (mot/che). Notably, 11 out of the 25 genes (HVO_1201- HVO_1225) were downregulated twofold to twentyfold in the deletion mutant, in agreement with the chemotaxis defect described above. Additionally, downregulation was observed for genes responsible for transducer proteins such as MpcT (HVO_0420), and two paralogs of Htr15 (HVO_0555, HVO_3005), even though they reside outside of the mot/che gene cluster but still contribute to chemotaxis. Notably, 29 out of the 78 down-regulated genes encode “conserved hypothetical proteins.” A severe downregulation exceeding fivefold was observed for 11 of these genes, with the most intensely downregulated gene coding for a hypothetical protein. These results indicate that conserved hypothetical proteins might play as yet unknown functions in chemotaxis, or that HVO_0758 regulates additional important processes that have not been detected in the phenotypic analyses of the deletion mutant. Notably, the transcript level of gene HVO_B0382 encoding a TATA box binding protein (TBP) was fivefold down-regulated. *H. volcanii* contains four TBP paralogs, three are encoded on the major chromosome, and HVO_B0382 is encoded on the minor chromosome pHV3. An analysis of the number of reads using the Integrated Genome Browser visualized the severe downregulation of HVO_B0382 (Supplementary Figure S10B), which was also verified by Northern Blot analysis (Supplementary Figure S7B). It is tempting to speculate that this TBP paralog might be involved in transcription initiation at several or many of the genes that are downregulated in the HVO_0758 deletion mutants.

Two genes encoding pilins (HVO_2450, HVO_2451) displayed more than twofold upregulation, in agreement with the increased biofilm formation of Δ HVO_0758. The largest up-regulated gene cluster, cluster 16, encompassed 12 genes (green in Supplementary Table S1). This set included genes encoding subunits of nitrite reductase, nitric oxide reductase, a copper-containing oxidoreductase, and halocyanine (HVO_2141, HVO_2147, HVO_2153, HVO_2150). Upregulation of genes for proteins linked to (anaerobic) redox processes could be interpreted as an adaptive response in preparation for biofilm formation, given the diminished oxygen availability within biofilms. Within the 29 up-regulated genes, seven encode “conserved hypothetical proteins.” The remaining up-regulated genes have various annotated functions, yet these functions did not provide insight into any additional biological role of HVO_0758 that could be tested.

Taken together, the phenotypic analysis together with the transcriptome analysis nicely complemented one another, showing that the downregulation of genes encoding the archaeella (archaeal flagella) and chemotaxis proteins in the deletion mutant lead to the observed defect in swarming.

¹ <http://www.ncbi.nlm.nih.gov/geo>

Transcriptome comparison between wild-type and mutant Δ HVO_2753

Recently we have reported the detailed analysis of another C(P) XCG zinc finger μ -protein of *H. volcanii*, HVO_2753 (Zahn et al., 2021). Similar analyses were performed as in the present study, including the generation of a deletion mutant, its phenotypic characterization, the heterologous and homologous production of the protein, its biochemical characterization, and the determination of the NMR solution structure. Notably, the deletion mutant also exhibited a swarming defect, similar to the deletion mutant discussed in this study. However, a transcriptome analysis had not been conducted in the former study. The interesting results of the transcriptome analysis of deletion mutant Δ HVO_0758 described above prompted us to address this gap, and a transcriptome analysis of Δ HVO_2753 was performed. Again, RNA was isolated from mid-exponential cultures (two biological replicates). Subsequently, RNA-Seq analysis was carried out as detailed above and in the Methods section. All RNA-Seq results have been deposited at the Gene Expression Omnibus (GEO, see Footnote 1) and obtained the accession No. GSE228855. [Supplementary Table S2](#) contains all genes displaying transcript levels that were more than twofold up- or downregulated in the deletion mutant and the same features as in [Supplementary Table S1](#). In total, 51 genes were down-regulated and 60 genes showed up-regulation. Once more, numerous differentially regulated genes clustered together in contiguous regions. In sum, 15 clusters of differentially co-regulated genes were identified.

Remarkably, all 17 genes that were most down-regulated in the Δ HVO_2753 mutant were localized in the mot/che gene cluster. For these genes the transcript levels were undetectable or extremely low. In fact, the transcript levels of all 25 genes of the HVO_1201-HVO_1225 gene cluster were down-regulated at least threefold. Once more, the transcript levels of additional chemotaxis genes that are situated outside of this gene cluster were also down-regulated, e.g., genes for the transducers Htr15 (HVO_0555), HemAT (HVO_1126), Htr15 (HVO_3005), Htr7 (HVO_1999), and BasB (HVO_0553). Again, the high number of down-regulated motility and chemotaxis genes is in excellent agreement with the observed swarming defect (Zahn et al., 2021). Further examination of the data revealed that among the remaining genes, 14 encode “conserved hypothetical proteins,” once more underscoring the lack of knowledge about the biological functions of many proteins.

Among the up-regulated genes, two gene clusters stand out prominently. The first, cluster 9, contains six genes from the *agl* (archaeal glycosylation) gene cluster, two of which are up-regulated

more than fivefold. These *agl* genes encode enzymes that are essential for biosynthesis of two distinct N-linked glycans, which are post-translationally coupled to the surface layer glycoprotein and additional proteins of *H. volcanii* (Pohlschroder and Esquivel, 2015; Eichler, 2020).

The second and notably larger cluster, Cluster 13, includes 20 up-regulated genes. It contains four genes encoding an ABC transporter with the annotated substrate “sugar” (HVO_2031 – HVO_2034), the gene for TrmB (HVO_2035), which has been shown to be a sugar sensor and a transcriptional regulator of sugar ABC transporters in *Thermococcus litoralis* and *Pyrococcus furiosus* (Lee et al., 2003, 2005), and a gene for a GalE sugar epimerase (HVO_2040). The remaining genes encode for seven conserved hypothetical proteins and proteins that are not indicative of sugar metabolism. Nevertheless, together the genes within the up-regulated clusters 9 and 13 suggest that in the Δ HVO_2753 mutant sugar import, sugar metabolism and glycoprotein production are induced.

[Table 2](#) gives an overview of the total numbers of differentially regulated genes, the numbers of up- and down-regulated genes, and several additional features of the transcriptome analyses of the two deletion mutants. [Supplementary Table S3](#) summarizes the genes of the mot/che gene cluster that are down-regulated in the two mutants and the annotated functions of the genes. Taken together, for both deletion mutants the transcriptome analyses and the phenotypic analyses were in excellent agreement.

Homologous production and purification of HVO_0758

For homologous overproduction in *H. volcanii*, the HVO_0758 gene was cloned into the shuttle vector pSD1/R1-6, which contains a strong synthetic promoter (Danner and Soppa, 1996). To generate a tagged fusion protein, the sequence for a C-terminal hexahistidine tag was added with one of the oligonucleotides used to amplify the gene. This strategy resulted in the homologous overproduction of a tagged fusion protein, which could be isolated via nickel chelating affinity chromatography. The elution fractions contained a high amount of HVO_0758, endogenous *H. volcanii* proteins that had a stretch of histidines (PitA and Cdc48; Allers et al., 2010), and some proteins that were co-isolated with HVO_0758 ([Supplementary Figure S11A](#)). Preparative size exclusion chromatography (SEC) was used as a second purification step to isolate pure monomeric HVO_0758 ([Supplementary Figure S11B](#)). The isolated protein was used for biochemical characterizations described below.

TABLE 2 Overview of the numbers of more than twofold differentially regulated genes between the wild-type and the in-frame deletion mutants Δ HVO_0758 and Δ HVO_2753.

Class of genes	Δ HVO_0758	Δ HVO_2753
Total No. of >2x regulated genes	108	111
No. of conserved hypothetical genes	40	29
No. of co-regulated gene clusters	19	15
No. of downregulated genes	78	51
No. in the mot/che gene cluster	11	25
No. of up-regulated genes	29	60

The transcript levels were determined via RNA-Seq in rRNA-depleted RNA samples from two biological replicates of cultures in the mid-exponential growth phase.

Biochemical characterization of HVO_0758

First, the zinc content of HVO_0758 was quantified. The isolated protein was dialyzed against a low salt buffer, because haloarchaeal proteins typically denature at low salt (Oren, 2008; Gunde-Cimerman et al., 2018). Nevertheless, only a small fraction of 0.2 zinc ions per molecule HVO_0758 could be determined. Therefore, the protein was hydrolyzed by proteinase K, and the zinc quantification was repeated. Notably, 1.0 zinc ions per HVO_0758 molecule were measured, revealing that (1) HVO_0758 is a *bona fide* zinc finger protein, and (2) that zinc binding is largely, but not fully preserved at low salt. This result is in agreement with the NMR results described above, which revealed that at 1 M NaCl the protein had a native structure, while at 0.3 M NaCl the native and a non-native structure co-existed.

Next, the stability of HVO_0758 was studied using tryptophan fluorescence. The maximum of the emission spectrum of tryptophan is around 330 nm to 350 nm and it depends on the hydrophobicity/hydrophilicity of its environment. In a hydrophobic environment the maximum is at a shorter wavelength, in a hydrophilic environment it is at a longer wavelength. Therefore, tryptophan fluorescence is well suited to study the unfolding of proteins (Varejão and Reverter, 2023). HVO_0758 contains one tryptophan near the C-terminus (W52, of 56 aa), which is in a rather hydrophobic environment and can be expected to change the fluorescence properties upon protein denaturation (compare Figure 4). The fluorescence emission spectrum was measured at the optimal NaCl concentration of 2,100 mM as well as reduced concentrations of 1,050 mM, 630 mM, and 0 mM (Figure 5; Supplementary Figure S12). The NMR analysis

had revealed that HVO_0758 was stable and had a native fold at 1 M NaCl (see above). Nevertheless, the reduction of the salt concentration from 2,100 mM to 1,050 mM led to a slight red shift of the maximum from 349 nm to 350 nm, underscoring that the tryptophan fluorescence is very sensitive to slight changes in the surrounding of tryptophans (Figure 5). At a NaCl concentration of 630 mM, the maximum was further shifted to 351 nm, and a second maximum around 346 nm appeared. This second maximum became more pronounced in the absence of salt. These results show that at low salt concentrations HVO_0758 has two different conformations, in excellent agreement with the NMR results. Unexpectedly, the maximum of the new conformation was blue-shifted, indicating that W52 moved to a more hydrophobic environment and the protein did not fully unfold to a random coil. The addition of EDTA to the protein at the optimal salt concentration led to a large red-shift to 356 nm, showing that EDTA was able to remove the zinc ion from the protein at room temperature and high salt, and that the protein unfolded in the absence of the zinc ion, again, in agreement with the NMR results.

In another approach the isolated protein was analyzed by mass spectrometry to determine whether or not it carries post-translational modifications. However, the major peak was at 7.301 kDa, exactly the mass of the unmodified protein (Supplementary Figure S13). Two minor peaks with higher masses were observed, which represent the protein with one and with two pairs of oxidized cysteines, respectively. Therefore, the MS analysis did not give any indication of a posttranslational modification, in spite of the high number of hydrophilic and charged amino acids that could potentially be modified.

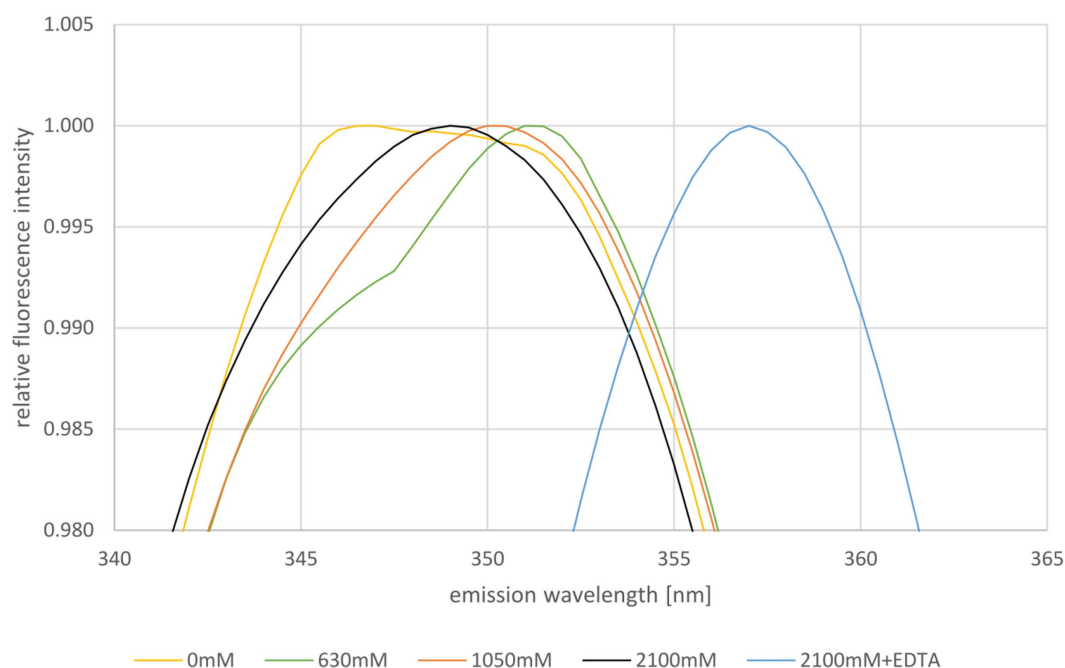


FIGURE 5

Normalized fluorescence emission spectra of HVO_0758 in buffers containing different NaCl concentrations, as indicated below. The relevant wavelength range from 340 nm to 365 nm is shown. Ten consecutive measurements per sample were performed and averaged. The spectra were normalized to 1 at the respective maximum fluorescence intensity.

Generation and characterization of point mutants

The next approach was the generation of point mutants, with the aim to unravel how the replacement of single amino acids affects zinc binding, folding, stability, and *in vivo* function. First, the four cysteines were replaced by alanines. However, the proteins became very unstable, and it was impossible to isolate even low amounts of protein via nickel chelating affinity chromatography from the mutants, in contrast to the wild-type HVO_0758 (Supplementary Figure S14A).

Next, the following three additional amino acids were selected and single alanine mutants were generated: arginine 9, glutamate 28, and lysine 31. Unfortunately, only very low protein amounts could be isolated from all three point mutants (Supplementary Figure S14B), indicating that all three mutations compromise the half-life of HVO_0758. The protein amounts were too low for the intended biochemical characterizations, and it also inhibited complementation studies. The large effect of all seven point mutations on the intracellular protein concentration led to the decision to terminate mutant studies.

Discussion

The project of characterizing C(P)XCG proteins from *H. volcanii* is part of the priority program “Small proteins of prokaryotes: an unexplored world” (SPP2002).² Notably, a characterization of the small proteome of *H. volcanii* combined ribosomal profiling with “small protein-optimized mass spectrometry” and led to the identification of 55 novel μ -proteins (Hadjeras et al., 2023). Only seven of these were found with both methods, while 40 were found solely by ribosomal profiling and 8 solely by peptidomics. The set of novel proteins overlapped, but was not identical to the set of novel proteins found in a second ribosomal profiling study with *H. volcanii* (Gelsinger et al., 2020). Together, these studies underscore that no experimental approach alone can unravel the full diversity of the μ -proteome of a species of interest.

The current *H. volcanii* project was chosen because 8% of all archaeal μ -proteins contain two C(P)XCG motifs and are putative zinc finger proteins, while that is only true for 1.5% of bacterial μ -proteins (Tarasov et al., 2008). A first attempt to clarify whether or not C(P)XCG proteins are important for archaeal biology and gain an insight into putative functions, 16 *in-frame* deletion mutants have been generated and phenotypically characterized, including HVO_0758 (Nagel et al., 2019). Currently, more than 30 C(P)XCG genes of *H. volcanii* have been deleted, and the majority exhibited phenotypic differences from the wild-type. Five genes could not be deleted and are thus essential, underscoring that C(P)XCG proteins fulfill important functions in *H. volcanii* and probably other archaea (data not shown). The aim is to select very few of these 43 C(P)XCG μ -proteins for a detailed analysis.

The first selected protein was HVO_2753, because it is the only *H. volcanii* protein with four C(P)XCG motifs, and its characterization has recently been published (Zahn et al., 2021). The next selected protein was HVO_0758, which was chosen based on the interesting

phenotype of the *in-frame* deletion mutant, i.e., a defect in swarming, enhanced biofilm formation, and a lag phase after inoculation in synthetic medium with glycerol (Nagel et al., 2019). The two proteins turned out to be very different, and, therefore, at the end of this discussion an overview of similarities and differences will be given.

HVO_0758 could be produced heterologously in *E. coli*, could be purified, and turned out to be in a folded conformation. This is not typical for μ -proteins from various prokaryotic species. In an overview study in the framework of SPP2002 it was attempted to produce and characterize 27 μ -proteins from nine bacterial and archaeal species (Kubatova et al., 2020). It turned out that only four of these proteins were folded and one protein was partially folded, whereas the majority of proteins were in the molten globule state, were totally unfolded, or could not be produced. Various reviews summarize the specific challenges in the heterologous or homologous production of μ -proteins, in contrast to average-sized proteins (Storz et al., 2014; Orr et al., 2020; Cassidy et al., 2021; Steinberg and Koch, 2021; Tharakan and Sawa, 2021; Chen et al., 2022; Gray et al., 2022; Leong et al., 2022; Schlesinger and Elsässer, 2022; Weidenbach et al., 2022; Zhang et al., 2022; Dong et al., 2023; Hassel et al., 2023).

The heterologously produced HVO_0758 turned out to have a fully folded conformation at 1 M NaCl, but to be a mixture of a native and a non-native conformation at 0.3 M NaCl. A mixture of two conformations was also observed with the homologously produced protein at low salt using tryptophan fluorescence. Together, these results indicate that HVO_0758 is not stable at low salt concentrations. However, the tryptophan fluorescence data as well as the zinc assay showed that it is not fully denatured at low salt. This is in contrast to other haloarchaeal proteins, which denature at low salt (Oren, 2008). Therefore, it seems that the complexation of four cysteines by the zinc ion protects HVO_0758 from denaturation, which typical haloarchaeal proteins experience at low salt. This might turn out to be a general advantage of haloarchaeal zinc finger proteins.

The zinc content of HVO_0758 was determined with two independent methods. On the one hand zinc was titrated to zinc-free HVO_0758, and the NMR analyses revealed that the addition of one zinc equivalent per HVO_0758 protein resulted in the generation of the native conformation. On the other hand, quantification of zinc ions after the proteolysis of isolated HVO_0758 revealed that one zinc ion had been bound to the protein prior to proteolysis. Together, these results showed that HVO_0758 is a *bona fide* zinc finger protein. This needs to be verified experimentally, because the C(P)XCG motifs are indicative for zinc binding, but do not prove it. For example, HVO_2753 has four C(P)XCG motifs, but binds only one zinc ion, while the second predicted “zinc finger” is zinc free (Table 3). Up to now 10 haloarchaeal μ -proteins with C(P)XCG motifs have been characterized, and four of them did not bind zinc, in contrast to the bioinformatics prediction (data not shown). The loss of metal binding in metal binding proteins by mutation, yielding folded and functional

TABLE 3 Predicted and experimentally quantified zinc content of the proteins HVO_0758 and HVO_2753.

Predicted	Experimentally quantified before proteolysis after proteolysis		
HVO_0758	1	0.17	1.0
HVO_2753	2	0.20	0.7

² <https://www.spp2002.uni-kiel.de/wordpress/>

metal-free descendants, is not uncommon in the evolution of metal binding proteins (Torrance et al., 2008). It seems that in these cases the protein structure is so stabilized by alternative interactions, that mutations can lead to the loss of zinc binding without the concomitant loss of the protein structure or function.

The NMR solution structure of HVO_0758 has been solved (Figure 4). In the Protein Database (PDB) no protein with a similar structure could be found, therefore, the structure will be helpful as template for future structure determinations of structure predictions (e.g., with AlphaFold) of homologous proteins and non-homologous proteins of similar structure.

HVO_0758 is characterized by an N-terminal alpha-helix of about 10 aa, which sticks out and looks like an interaction domain. It contains several positive amino acids, therefore, it is tempting to speculate that HCO_0758 might bind DNA or RNA. The unusual pI of 7.6 (for haloarchaeal proteins) and the differential regulation of many genes might point in the same direction. Unfortunately, attempts to co-isolate DNAs or RNAs or to measure DNA binding were unsuccessful until now.

The protein part C-terminal of the alpha-helix is folded around the zinc finger (Figure 4). The zinc binding seems to be of utmost importance for the protein structure, because in all four C->A mutants no protein could be detected (Supplementary Figure S14A), and because the native structure was lost upon addition of EDTA (NMR, tryptophan fluorescence). This is not always the case for zinc fingers in other proteins. For example, bacterial Ros proteins, which are transcriptional regulators, typically contain a C₂H₂ zinc finger (Malgieri et al., 2015). However, in a minority of homologs one of the two cysteines is replaced by another amino acid, nevertheless, these proteins with an incomplete zinc finger motif are able to bind zinc and DNA, and lose both features upon addition of EDTA (Baglivo et al., 2009). In addition, a C82D mutant of the well-studied Ros protein from *Agrobacterium tumefaciens* was generated, and also the mutant protein kept the ability to bind zinc and DNA (Baglivo et al., 2009). However, typically four cysteine/histidine ligands are essential for zinc binding, like in HVO_0758.

The phenotypic analyses revealed that the Δ HVO_0758 mutant exhibited an extended lag phase after cells pre-grown in complex medium were inoculated in synthetic medium with glucose. This seems to indicate that HVO_0758 might be important for the transition from one carbon source to another. Transitions between different conditions occur with extremely different velocities in *H. volcanii*. Whereas upregulation of the gene for the tryptophanase (HVO_0009) can be observed already 2 min after tryptophane addition to the medium (unpublished results), it takes *H. volcanii* about 20 h to adapt to a low salt medium containing 0.7 M NaCl (Jantzer et al., 2011). In an earlier study we have shown that *H. volcanii* exhibits a lag phase of about 1 h when cells pre-grown on casamino acids were switched to glucose as sole carbon and energy source (Zaigler et al., 2003). During this lag period more than 30 genes were transiently induced, which had identical transcript levels during steady state growth in both media. Some or all of the encoded proteins can be hypothesized to be important for the transition process, but not for steady states.

Deletion of both genes, HVO_0758 and HVO_2753, led to a complete loss of swarming during the first days, indicating that both zinc finger μ -proteins are essential for motility and chemotaxis. In agreement with the phenotype the analyses of the transcriptomes of

both mutants revealed that various genes of the large mot/che gene cluster are considerably down-regulated (Supplementary Table S3), nicely explaining the phenotype.

The Δ HVO_0758 mutant showed an enhanced biofilm formation, while the Δ HVO_2753 mutant had a defect in biofilm formation. In agreement with this difference, there was no overlap of up-regulated genes in the transcriptome analyses. In the transcriptome of Δ HVO_0758 genes for redox enzymes were induced, e.g., for nitrite reductase and nitrite oxide reductase. This can be speculated to be a pre-adaptation for biofilm formation, because oxygen availability is limited in biofilms. Also, two genes for pili were induced, which is in agreement with the essential roles of pili in biofilm formation of *H. volcanii* (Pohlschroder and Esquivel, 2015; Esquivel et al., 2016). In contrast, genes for sugar metabolism and glycan biosynthesis were up-regulated in the Δ HVO_2753 mutant. The *agl* (archaeal glycosylation) genes form a very large cluster in the genome of *H. volcanii* (HVO_2046 – HVO_2061). In *H. volcanii* the N-glycosylation pathways have been well-studied, and targets for N-glycosylation are, e.g., the major surface layer protein and pili (Esquivel et al., 2016; Kandiba et al., 2016; Tamir and Eichler, 2017; Pohlschroder et al., 2018; Shalev et al., 2018; Eichler, 2020; Schulze et al., 2021). Future studies are needed to unravel how an enhanced N-glycosylation of the S-layer and/or the pili might interfere with swarming and/or biofilm formation.

For the HVO_0758 deletion mutant, the growth deficit in glycerol medium could be complemented (Supplementary Figure S9), while the swarming phenotype and the biofilm phenotype could not be complemented (data not shown). Partial complementation was also observed for various additional mutants. The following reasons can possibly explain the failure to complement some of the phenotypes: (1) the His₆-tag can prevent interactions with other molecules that are essential for function, (2) the His₆-tag (more than 10% of the size μ -proteins) might influence the overall folding of a protein and thereby change the binding affinity to other molecules, (3) the His₆-tag might influence the half life of the protein, and (4) the function maybe sensitive to the intracellular concentration, which differs from the native concentration when the gene is expressed from a plasmid instead of the native chromosomal site.

Conclusion and outlook

The present study represents an in-depth analysis of the μ -protein HVO_0758, which contains two C(P)XCG motifs and thus had the potential to be a zinc finger protein. The protein was produced heterologously in *E. coli* and homologously in *H. volcanii* and the isolated protein was characterized with various approaches, including a quantitative zinc assay, tryptophan fluorescence, and different NMR measurements. An *in-frame* deletion mutant and seven point mutants were generated. HVO_0758 was shown to be a *bona fide* zinc finger protein and binds one zinc ion per protein molecule. The NMR solution structure revealed that it is comprised of an N-terminal alpha helix with several positively charged amino acids that is placed on top of a globular core, which is stabilized by the zinc finger. HVO_0758 is the second *H. volcanii* C(P)XCG μ -protein that has been thoroughly characterized. Table 4 summarizes the similarities and differences of the two proteins, HVO_0758 and HVO_2753.

TABLE 4 Comparison of characteristic features of HVO_0758 and HVO_2753.

Feature	HVO_0758	HVO_2753
Conserved in	Haloferax	Haloarchaea, methanogenic Archaea
No. aminoacids	56	59
pI value	7.6	6,7
Stability at low salt	partly	fully
No. C(P)XCG motifs	2	4
Predicted zinc fingers	1	2
Zinc ions bound per protein molecule	1	1
Zinc removal upon EDTA addition	yes	no
Cysteines essential for folding	yes	yes
Unordered N-Terminus	no	yes
Helix on top of globular core	yes	no
Phenotype of a deletion mutant:		
Growth in many media	normal	normal
Growth on glycerol	lag phase	normal
Swarming	loss	loss
Biofilm formation	increase	decrease
RNA-Seq: down-regulated	mot/che genes	mot/che genes
RNA-Seq: up-regulated	redox proteins	sugar metabolism
		Glycan synthesis

Of course it would be of interest to identify binding partners of HVO_0758, which might shed light on the molecular mechanism of regulating swarming and biofilm formation. However, initial attempts to co-isolate proteins, DNA, or RNA molecules that bind HVO_0758 specifically did not lead conclusive results, and thus they are omitted in this presentation. The experimental design of the co-isolation approaches will be optimized and the results will be published in the future.

Materials and methods

Databases and bioinformatics analyses

All gene and protein sequences derived from the *H. volcanii* genome were obtained from the HaloLex database (Pfeiffer et al., 2008).³ To visualize the transcript levels of HVO_0758 and further selected transcripts, the respective results of a RNA-Seq and a dRNA-Seq study (Babski et al., 2016; Laass et al., 2019) were visualized using the Integrated Genome Browser (Freese et al., 2016).

Homologous sequences of the HVO_0758 gene were searched for using NCBI BlastP (Altschul et al., 1990). The top 100 hits were retrieved and a multiple sequence alignment was generated using the ClustalOmega program at the EMBL-EBI site (Sievers et al., 2011).

Strains, media, and culture conditions

All strains generated in this study are derived from the *H. volcanii* H26 strain which also served as the wild-type control. The strains were grown in complex medium or synthetic media with different carbon and energy sources as described previously (Allers et al., 2004; Dambeck and Soppa, 2008; Jantzer et al., 2011). All strains with expression plasmids based on pSD1 were grown in the presence of Novobiocin (0.5 µg/mL; Danner and Soppa, 1996).

The *E. coli* strain XL1-Blue MRF' (Agilent Technologies, Waldbronn, Germany) was used for cloning and construction of all plasmids used in this study (Green and Sambrook, 2012).

Heterologous production in *Escherichia coli*, purification, and characterization

The ORF for HVO_0758 was introduced into the pE-SUMO vector via BsaI and XbaI restriction sites. This puts the protein directly behind the C-terminal double glycine of the SUMO protein, leading to a fusion protein of SUMO protein and the protein of interest with a His₆-tag on the N-terminus of the SUMO protein.

The plasmid carrying the ORF was used to transform BL21(DE3) *E. coli* cells, which were subsequently cultivated in M9 minimal medium enriched with ¹⁵N-NH₄Cl alone or together with ¹³C-Glucose for isotope-labeled samples. Ampicillin was added to the M9 medium (final concentration 1 mM) to ensure presence of the plasmid in the *E. coli* cells. The cultures were incubated at 37°C until turbidity reached an OD₆₀₀ of 0.6. Protein overproduction was then induced by addition of 1 mM IPTG. The cultures were further incubated overnight

³ www.halolex.mpg.de

and then harvested by centrifugation at 5000 rpm and 4°C for 15 min. Cell pellets were either flash-frozen and stored at −80°C or directly resuspended in buffer (25 mM BisTris pH 7, 300 mM NaCl, 5 mM 2-mercaptoethanol, 100 μM ZnCl₂) together with protease inhibitor (cOmplete™, Roche, Germany; one tablet per liter) for purification. The cells were lysed by french press and then centrifuged (16,000 x g, 45 min, 4°C) to remove cell debris. The supernatant was loaded onto a HisTrap for affinity purification and subsequently eluted at ~20% of an imidazole gradient up to 500 mM imidazole. Cleavage of the SUMO protein to obtain the pure desired protein was achieved by addition of Ulp1 (SUMO protease) during dialysis against buffer without imidazole. The protein was then applied to a HisTrap again and pooled in the flow-through, while the cleaved His-tagged SUMO protein bound to the column. In the final purification step, size-exclusion chromatography was carried out to obtain >95% pure protein. After every step, SDS-PAGE was carried out for relevant fractions of the columns to determine identity and purity of the protein.

NMR spectroscopic experimental data

NMR measurements were carried out in buffer containing 25 mM BisTris pH 7, 1 M NaCl, 5 mM 2-mercaptoethanol, 100 μM ZnCl₂ and 5% D₂O at 298 K. In order to reference spectra, 1 mM DSS was added and the proton methyl signal of the trimethylsilyl group set to 0.00 ppm. ¹³C and ¹⁵N referencing was done according to Wishart (2011). All used NMR spectrometers were manufactured by Bruker and equipped with cryogenic probeheads with z-axis gradient ¹H{¹³C,¹⁵N}. Spectrometer frequencies ranged from 600 to 900 MHz.

Backbone assignment was done in a 2D ¹H¹⁵N HSQC with 3D experiments HNCO, HN(CA)CO, HNCACB and HN(CO)CACB. Side-chain assignment was done in a 2D ¹H¹³C HSQC with 3D experiments H(CCO)NH, CC(CO)NH, HCCH-TOCSY, HCCH-COSY and ¹⁵N-TOCSY-HSQC. The aromatic assignment was done in a 2D ¹H¹³C TROSY together with a 3D ¹H¹³C NOESY-SOFAST-HMQC (mixing time 100 ms).

For the NOE-based distance restraints of the structure calculation 3D ¹H-¹H-¹⁵N-NOESY-HSQC and ¹H-¹H-¹³C-NOESY-HSQC spectra (mixing time for both 120 ms) were measured. A 3D HNHA spectrum was measured for the calculation of ³J_{HNHα} coupling constants (Vuister and Bax, 1993).

A temperature series was measured with a series of 1D ¹H and 2D ¹H¹⁵N HSQC spectra ranging from 278 K to 333 K in 5 K increments. Temperature coefficients were determined from this by plotting the amide proton chemical shift perturbations against the temperature with a linear fit (Baxter and Williamson, 1997).

2D ¹⁵N-ZZ-exchange experiments were carried out to confirm that two conformations of the protein were present at low NaCl concentrations. Mixing times were varied from 100 ms to 800 ms. Cross peaks could be observed best at 200 ms.

Relaxation experiments were performed with a ¹⁵N-labeled sample on a 600 MHz spectrometer. The ¹⁵N T₁ relaxation time was obtained from a pseudo-3D spectrum measuring a series of ¹H¹⁵N spectra with increasing relaxation delays (20, 60, 100, 200, 400, 600, 800, 1,200, 1,500 and 1800 ms). The ¹⁵N T₂ relaxation time was obtained from a pseudo-3D spectrum as well (delays: 16.96, 33.92, 67.84, 135.68, 169.60, 203.52, 237.44 and 271.36 ms).

{¹H}-¹⁵N-hetNOEs were measured as the ratio of signal intensities from two spectra recorded with and without amide proton saturation. Relaxation parameters were used to determine the S² order parameter and the experimental rotational correlation time both with TENSOR2.

All spectra were recorded and processed using TopSpin (different versions). Assignment was carried out using NMRFAM-SPARKY 1.470 (Lee et al., 2015).

Structure calculation

The structure calculation was performed with CYANA 3.98.13. Three 3D NOESY spectra were used as input for the fully automated NOE cross-peak assignment: ¹H¹⁵N-NOESY-HSQC, ¹H¹³C-NOESY-HSQC spectra (mixing time for both 120 ms) and ¹H¹³C-NOESY-SOFAST-HMQC (mixing time 100 ms). Peaks in the NOESY spectra were selected using the restricted peak picking of NMRFAM-SPARKY and manually checked and corrected. Peak shift tolerances were set to 0.02 ppm for protons and carbon atoms and 0.2 ppm for nitrogen atoms. Additional restraints included hydrogen bond distances derived from NOESY analysis, ³J_{HNHα} coupling constants from the 3D HNHA spectrum and dihedral angles from the TALOS-N prediction. Furthermore, the zinc-binding pocket was defined by setting lower and upper distance restraints between the four sulfur atoms of the coordinating cysteine residues to 3.65 Å and 4 Å, respectively. These distances were chosen by referencing the structures of other zinc-binding proteins. The restraints were only set after we saw that the binding pocket was also present in the calculated structure just through NOESY contacts. A standard protocol was used with 200 initial steps, 15,000 refinement steps and 20 final structures per iteration. The final bundle consisting of the lowest energy structures was deposited in the pdb (ID 8Q5B) and bmrB (ID 34844).

Generation of an *in-frame* deletion mutant

For the construction of all in-frame deletion mutants, the established Pop-In-Pop-Out method was applied (Allers et al., 2004; Hammelmann and Soppa, 2008). The oligonucleotides for the generation of the deletion mutant of HVO_0758 are listed in Supplementary Table S4. As *H. volcanii* is a polyploid organism with about 20–30 copies of the major chromosome (Breuert et al., 2006; Maurer et al., 2018), it happens sometimes that after the Pop_Out selection not all copies of the wild-type chromosome are replaced by the deletion variant of the respective gene, but that one or a few copies remain and the respective clones are heterozygous clones. Therefore, the homozygosity of the deletion mutant was verified by multicycle PCR using isolated genomic DNA as template, and the absence of the HVO_0758 transcript was verified via Northern blot analysis.

Phenotypic characterization: growth analysis

Growth analyses were performed in round-bottom microtiter plates as described previously (Jantzer et al., 2011). A pre-culture was grown in complex medium, harvested, washed once in basal salts and resuspended in basal salts to an OD₆₀₀ of 0.75. For each condition,

140 µL medium was mixed with 10 µL cell suspension, resulting in an OD₆₀₀ of 0.05 at the start of the growth experiment. The cultures were grown on a Heidolph Titramax 1,000 rotary shaker (Heidolph, Schwallbach, Germany) with 1,100 rpm at 42° C. The OD₆₀₀ was determined frequently using the microtiter plate photometer Spectramax 340 (Molecular Devices, Ismaning, Germany). Growth curves were obtained from average values of at least three biological replicates and their standard deviations.

Phenotypic characterization: swarming assay

The swarming assay was in general performed as described previously (Nagel et al., 2019). Twenty milliliter of synthetic glucose medium with 0.3% (w/v) agar was filled into petri dishes (Sarstedt, Nümbrecht, Germany) 1 day prior to start of the assay. The cells from pre-cultures in the respective medium were harvested by centrifugation and resuspended in basal salts to an OD₆₀₀ of 20. Four microliter of the cell suspension was injected deeply into the swarm agar in the center of the plates, because *H. volcanii* is swarming only at low oxygen concentrations. The plates were incubated for several days at 42°C in a Styrofoam box together with a cup of water to inhibit drying. The swarming diameter was measured and pictures were taken every 24 h.

Phenotypic characterization: biofilm formation

For the biofilm assay, cultures were grown in synthetic medium with glucose to the mid-exponential growth phase. The OD₆₀₀ was measured, cells were pelleted by centrifugation and resuspended in fresh glucose medium to an OD₆₀₀ of 0.5. For biofilm formation, 24 well microtiter plates were used (Sarstedt, Nümbrecht, Germany).

The biofilm assay consists of several steps, i.e., formation of a biofilm, removal of planktonic cells, fixation and staining of adherent cells, destaining and photometric quantification of the supernatant. The assay has been performed as described previously by Legerme et al. (2016), with some modifications. To this end, 2 mL of cell suspension was transferred into each well, and the plates were incubated without shaking at 42°C for 48 h.

After that, the supernatant was removed, and 1,000 µL of fixing solution [2% (w/v) acetic acid] was given into each well, and the plate was incubated for 5 min at room temperature. The supernatant was removed, and the plate was dried for 1–2 h at 37°C until all the liquid has evaporated. After that, 500 µL staining solution [0.1% (w/v) crystal violet] was given into each well, and it was incubated for 10 min at room temperature. After that, the supernatant was removed, and the wells were washed three times with 3 mL distilled water very carefully to not detach formed biofilm. Subsequently, 2 mL of destaining solution [10% (v/v) acetic acid, 30% (v/v) methanol] was given into each well, and the plate was incubated for 10 min at room temperature until all crystal violet bound in the biofilm was solved homogeneously in the supernatant. Then, 100 µL of the supernatant of each well was transferred into a new 96-well microtiter plate, and the OD₆₀₀ was recorded with a microtiter plate photometer (Spectramax 340, Molecular Devices, San Jose, CA, United States). A negative control

(medium without cells) was included in the assay, and its value of the destaining solution (about 0.05) was subtracted from the values of all tested strains.

Generation of the point mutations of HVO_0758

Point mutants of HVO_0758_NHis and HVO_0758_CHis were created via site-directed mutagenesis according to the protocol of the QuikChange II Site-Directed Mutagenesis Kit (Agilent Technologies, Santa Clara, United States). The oligonucleotides used for site-directed mutagenesis are listed in [Supplementary Table S4](#). Since pSD1-R1/6 is too large for this approach (>10 kbp), the native gene along with an either N- or C-terminal His-tag was first subcloned into the vector pSK(+). After site-directed mutagenesis, the mutated gene was excised with KpnI and either NdeI or NcoI and ligated with the vector pSD1-R1/6 linearized with the same enzymes. The sequences of the mutated gene versions were verified by sequencing (GATC/Eurofins; eurofinsgenomics.eu/en/custom-DNA-sequencing).

Northern blot analysis

The expression levels of selected genes were analyzed by Northern blot analysis.⁴ *H. volcanii* cultures were grown in synthetic glucose medium to mid-exponential growth phase (about 4–6 × 10⁸ cells/ml). Cells were harvested by centrifugation, and total RNA was isolated using phenol/chloroform extraction. Equal amounts of total RNA from each sample (usually 2–5 µg) were separated on a denaturing formaldehyde agarose gel. The RNA was transferred by capillary blotting onto a nylon membrane, and fixed by UV-cross-linking. Digoxigenin-labeled DNA probes were generated by PCR using DIG-dUTP and a dNTP mix with reduced dTTP concentration. The primers for probe generation are listed in [Supplementary Table S4](#). Hybridization was performed overnight at 50°C. The membrane was washed twice with 2 × SSC/0.1% SDS and twice with 1 × SSC/0.1% SDS. The probes were detected using an anti-DIG antibody coupled to alkaline phosphatase and the chemiluminescence substrate CDP star according to the instructions of the manufacturer (Roche, Mannheim, Germany). The signals were visualized with X-ray films (GE Healthcare), and the sizes were analyzed with the size marker ‘RiboRuler Low/High Range RNA Ladder’ (Thermo Fisher Scientific).

Transcriptome analysis using RNA-Seq

Cultures were grown to mid-exponential growth phase (4–6 × 10⁸ cells per ml) in glucose medium, cells were harvested, and total RNA was isolated as described above. Ribosomal RNAs were depleted from 1 µg of non-degraded DNase I-digested total RNA by subtractive hybridization using the Pan-Archaea_riboPOOL kit (siTOOLS, Germany) according to the manufacturer’s protocol using Dynabeads

⁴ <https://protocolexchange.researchsquare.com/article/pex-535/v1>

MyOne Streptavidin T1 beads (Invitrogen) and then subjected to cDNA libraries preparation.

cDNA libraries were prepared at Vertis Biotechnologies AG (Freising, Germany) using the Adapter ligation method. Briefly, the rRNA-depleted RNA samples were first fragmented using ultrasound (2 pulses of 30 s each at 4°C) and then directly used for cDNA synthesis. First, an oligonucleotide adapter was ligated to the 3' end of the RNA molecules. First-strand cDNA synthesis was performed using M-MLV reverse transcriptase and the 3' adapter as a primer. The first-strand cDNA was purified and the 5' Illumina TruSeq sequencing adapter was ligated to the 3' end of the antisense cDNA. The resulting cDNA was then PCR-amplified to about 10–20 ng/μl using a high fidelity DNA polymerase (12 PCR cycles). The cDNA was purified using the Agencourt AMPure XP kit (Beckman Coulter Genomics) and was analyzed by capillary electrophoresis. The primers used for PCR amplification were designed for TruSeq sequencing according to the instructions of Illumina. The following adapter sequences flank the cDNA inserts: TruSeq_Sense_primer: (NNNNNNNN=i5 Barcode for multiplexing) 5'-AATGATACGGCGACCACCGAGATCTACAC-NNNNNNNN-ACACTCTTTCCCTACACGACGCTCTTCCGATCT-3'.

TruSeq_Antisense_primer: (NNNNNNNN=i7 Barcode for multiplexing) 5'-CAAGCAGAAGACGGCATACGAGAT-NNNN-NNNN-GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT-3'.

cDNA libraries were pooled on an Illumina NextSeq 500 high-output flow cell and sequenced in single-end mode (75 cycles) with 10 million reads per RNA-seq library (at the Core Unit SysMed at the University of Würzburg). Raw sequencing reads in FASTQ format and normalized coverage files are available via Gene Expression Omnibus (GEO, see Footnote 1) under accession number GSE228855.

Bioinformatic analysis of the RNA-seq results

To assure high sequence quality, Illumina reads were quality and adapter trimmed via Cutadapt (Martin, 2011) version 1.16.1 using a cutoff Phred score of 20 (command line parameters: -quality-cutoff = 20 -m 20 -a AGATCGGAAGAGCACACGTCTGAACTCCA GTCAC). After trimming, the reads were mapped to the reference genome (*H. volcanii* DS2) using STAR (Dobin et al., 2013; Dobin and Gingeras, 2015) version 2.7.5b. Mapped reads were subsequently counted for all annotated genes using featureCounts (Liao et al., 2014) version 1.6.2. The quality of the raw data and each preprocessing step was assessed using multiQC (Ewels et al., 2016) version 1.6. MultiQC was used to aggregate statistics for featureCounts and fastQC version 0.11.6 into a clear html report.⁵

To facilitate visualization in a genome browser, HRIBO (Gelhausen et al., 2021) version 1.6.0 was used to generate perstrand genomic coverage plots for each library indicating the number of mapped reads per nucleotide. The used annotations for *H. volcanii*

DS2 was retrieved from NCBI. The coverage files were visualized using the Integrated Genome Browser IGB (Freese et al., 2016).

Differential expression analysis

Differential expression analysis for *Haloferax volcanii* DS2 H26 RNA-seq libraries was conducted using DESeq2 (Love et al., 2014) version 1.18.1. Significant results were retrieved by applying a cutoff to the log2 fold change [$\text{abs}(\log_2\text{FC}) > 2$] and the adjusted value of $p < 0.05$ for all output files. Additionally, plots for visualization and quality control were generated. These include heatmaps of the normalized read count correlation, principal component analysis (PCA) and MA-plots for each analyzed contrast.

Homologous production and native purification

For the homologous overproduction of HVO_0758 the respective gene was cloned into the expression vector pSD1-R1/6 containing a strong constitutive promoter (Danner and Soppa, 1996). The codons for a N-terminal His₆-tag were added with one of the primers (Supplementary Table S4), so that a fusion protein was produced, which could be isolated with nickel chelating chromatography. The sequence of the resulting plasmid was verified by sequencing, and it was used to transform the HVO_0758 deletion mutant. The plasmid complemented the phenotype in glycerol medium, and, thus, it was guaranteed that the produced protein was folded and functional *in vivo*. The production strain was grown in complex media with 0.5 μg/mL novobiocin overnight. Five hundred milliliter of complex medium with 0.5 μg/mL novobiocin was then inoculated with the pre-culture to a start OD₆₀₀ of 0.005. The production culture was grown for 24 h and the cells were harvested by centrifugation (6,500 g, 30 min, 4°C). IMAC was used as a first purification step. The pellet was suspended in 5 mL binding buffer (2.1 M NaCl, 20 mM HEPES pH 7.5, 20 mM imidazole, 1 mM PMSF) and the cells were lysed via sonication on ice. Cell fragments were pelleted by centrifugation (4,000 g, 30 min, 4°C) and the cleared lysate was mixed with Chelating Sepharose® Fast Flow (GE Healthcare) beads which had been charged with Ni²⁺ ions and suspended in binding buffer. The beads were washed four times with binding buffer to remove unspecifically bound proteins. The His-tagged HVO_0758 as well as other histidine-rich proteins were eluted with elution buffer (2.1 M NaCl, 20 mM HEPES pH 7.5, 300 mM imidazole).

Size exclusion chromatography was used as a second purification step in order to separate other specifically-bound proteins from the target protein. Elution fractions of the first purification step were loaded on a SuperDex 75 Increase (10/30) FPLC column (GE Healthcare) with 2.1 M NaCl, 20 mM HEPES pH 7.5 as mobile phase and a flow rate of 0.4 mL/min. A standard curve was generated with a chromatogram of a mixture of the proteins aprotinin (6.5 kDa), RNase A (13.7 kDa), Ovalbumin (44 kDa) and Conalbumin (57 kDa).

Samples of the different steps of the IMAC as well as the SEC were analyzed via Tricine-SDS-PAGE (Schägger and von Jagow, 1987).

⁵ <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>

Fluorimetric zinc quantification

To quantify the bound zinc of the protein, the highly zinc-specific fluorophore ZnAF-2F was used (Hirano et al., 2002). HVO_0758 was purified as described above. Protein concentration was determined via UV-absorption at 280 nm ($\epsilon_{280}=6,990 \text{ M}^{-1} \cdot \text{cm}^{-1}$) and the experiment was performed as described previously (Zahn et al., 2021). In short, 1 μM Protein was dialyzed against 25 mM NaCl and 20 mM HEPES, pH 7.5, and incubated with proteinase K (100 $\mu\text{g}/\text{mL}$) over night at 37°C to release the bound zinc. 3 μM of the fluorophore was added and the measurements were performed on a microtiter plate fluorimeter (ClarioStar, BMG LabTech, Ortenberg, Germany) with an excitation wavelength of 492 nm and detection at 517 nm. Four technical replicates were used for each biological replicate as well as the standard curve, which was generated with solutions containing 0, 0.5, 1 and 2 μM ZnSO_4 .

Stability analysis via tryptophan fluorescence

Fluorescence spectroscopy was performed in order to analyze the stability of HVO_0758_CHis in buffers containing various salt concentrations. 1 mg/mL of purified protein was either diluted in or dialyzed against a buffer containing 20 mM HEPES, pH 7.5, lowering the NaCl concentration to 1,050 mM, 630 mM or 0 mM, respectively. One measurement was recorded shortly after adding 20 mM EDTA to the sample in order to remove all zinc bound to the protein. Measurements were performed on a LS 55 Luminescence Spectrometer (PerkinElmer, Waltham, United States). The samples were excited at 280 nm and emission spectra were recorded from 300 nm to 440.5 nm with a step size of 0.5 nm. For each sample, 10 consecutive measurements were averaged.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found at: <https://www.ncbi.nlm.nih.gov/geo/>, GSE228855. <https://www.rcsb.org/>, 8Q5B, <https://bmrb.io/>, 34844.

Author contributions

DÜ: Conceptualization, Writing – original draft, Writing – review & editing, Formal analysis, Investigation, Visualization. DP: Formal analysis, Investigation, Methodology, Visualization, Writing – original draft, Writing – review & editing. AB: Conceptualization, Methodology, Supervision, Visualization, Writing – review & editing. LH: Formal analysis, Investigation, Methodology, Writing – review & editing.

References

Allers, T., Barak, S., Liddell, S., Wardell, K., and Mevarech, M. (2010). Improved strains and plasmid vectors for conditional overexpression of his-tagged proteins in *Haloflex volcanii*. *Appl. Environ. Microbiol.* 76, 1759–1769. doi: 10.1128/AEM.02670-09

editing. RG: Investigation, Methodology, Writing – review & editing. RB: Funding acquisition, Project administration, Supervision, Writing – review & editing. CS: Formal analysis, Writing – review & editing, Funding acquisition, Supervision. HS: Formal analysis, Funding acquisition, Supervision, Writing – original draft, Writing – review & editing. JS: Funding acquisition, Supervision, Writing – review & editing, Conceptualization, Project administration, Writing – original draft.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This collaborative work was supported by the German Research Foundation (DFG) in the framework of priority program SPP2002 “Small Proteins in Prokaryotes, an Unexplored World” (grant SO 264/26 to JS, grant SH580/7-1 and SH580/7-2 to CMS, grant BA 2168/21-2 to RB, grant BE 3869/5-1; www.spp2002.uni-kiel.de/wordpress). The center for biomolecular magnetic resonance (BMRZ) is supported by the state of Hesse, Germany.

Acknowledgments

We thank Uwe Linne and the staff of the “Service Centre for Mass Spectrometry and Elemental Analytik” of the University of Marburg for the mass spectrum of native HVO_0758.

Conflict of interest

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.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2023.1280972/full#supplementary-material>

Allers, T., Ngo, H.-P., Mevarech, M., and Lloyd, R. G. (2004). Development of additional selectable markers for the halophilic archaeon *Haloflex volcanii* based on the *leuB* and *trpA* genes. *Appl. Environ. Microbiol.* 70, 943–953. doi: 10.1128/aem.70.2.943-953.2004

- 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
- Babaki, J., Haas, K. A., Näther-Schindler, D., Pfeiffer, F., Förstner, K. U., Hammelmann, M., et al. (2016). Genome-wide identification of transcriptional start sites in the haloarchaeon *Haloferax volcanii* based on differential RNA-Seq (dRNA-Seq). *BMC Genomics* 17:629. doi: 10.1186/s12864-016-2920-y
- Baglivo, I., Russo, L., Esposito, S., Maligneri, G., Renda, M., Salluzzo, A., et al. (2009). The structural role of the zinc ion can be dispensable in prokaryotic zinc-finger domains. *Proc. Natl. Acad. Sci. U. S. A.* 106, 6933–6938. doi: 10.1073/pnas.0810003106
- Bak, R. O., Gomez-Ospina, N., and Porteus, M. H. (2018). Gene editing on center stage. *Trends Genet.* 34, 600–611. doi: 10.1016/j.tig.2018.05.004
- Baxter, N. J., and Williamson, M. P. (1997). Temperature dependence of ¹H chemical shifts in proteins. *J. Biomol. NMR* 9, 359–369. doi: 10.1023/a:1018334207887
- Breuer, S., Allers, T., Spohn, G., and Soppa, J. (2006). Regulated polyploidy in halophilic archaea. *PLoS One* 1:e92. doi: 10.1371/journal.pone.0000092
- Cassidy, L., Kaulich, P. T., Maaß, S., Bartel, J., Becher, D., and Tholey, A. (2021). Bottom-up and top-down proteomic approaches for the identification, characterization, and quantification of the low molecular weight proteome with focus on short open reading frame-encoded peptides. *Proteomics* 21:e2100008. doi: 10.1002/pmic.202100008
- Chen, L., Yang, Y., Zhang, Y., Li, K., Cai, H., Wang, H., et al. (2022). The small open Reading frame-encoded peptides: advances in methodologies and functional studies. *ChemBiochem* 23:e202100534. doi: 10.1002/cbic.202100534
- Dambeck, M., and Soppa, J. (2008). Characterization of a *Haloferax volcanii* member of the enolase superfamily: deletion mutant construction, expression analysis, and transcriptome comparison. *Arch. Microbiol.* 190, 341–353. doi: 10.1007/s00203-008-0379-1
- Danner, S., and Soppa, J. (1996). Characterization of the distal promoter element of halobacteria *in vivo* using saturation mutagenesis and selection. *Mol. Microbiol.* 19, 1265–1276. doi: 10.1111/j.1365-2958.1996.tb02471.x
- 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
- Dobin, A., and Gingeras, T. R. (2015). Mapping RNA-seq reads with STAR. *Curr. Protoc. Bioinformatics* 51, 11.14.1–11.14.19. doi: 10.1002/0471250953.bi1114s51
- Dong, X., Zhang, K., Xun, C., Chu, T., Liang, S., Zeng, Y., et al. (2023). Small open Reading frame-encoded Micro-peptides: an emerging protein world. *Int. J. Mol. Sci.* 24. doi: 10.3390/ijms241310562
- Dosset, P., Hus, J. C., Blackledge, M., and Marion, D. (2000). Efficient analysis of macromolecular rotational diffusion from heteronuclear relaxation data. *J. Biomol. NMR* 16, 23–28. doi: 10.1023/a:1008305808620
- Eichler, J. (2020). N-glycosylation in Archaea-new roles for an ancient posttranslational modification. *Mol. Microbiol.* 114, 735–741. doi: 10.1111/mmi.14569
- Eom, K. S., Cheong, J. S., and Lee, S. J. (2016). Structural analyses of zinc finger domains for specific interactions with DNA. *J. Microbiol. Biotechnol.* 26, 2019–2029. doi: 10.4014/jmb.1609.09021
- Esquivel, R. N., Schulze, S., Xu, R., Hippler, M., and Pohlschroder, M. (2016). Identification of *Haloferax volcanii* pilin N-Glycans with diverse roles in pilus biosynthesis, adhesion, and microcolony formation. *J. Biol. Chem.* 291, 10602–10614. doi: 10.1074/jbc.M115.693556
- Ewels, P., Magnusson, M., Lundin, S., and Käller, M. (2016). MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* 32, 3047–3048. doi: 10.1093/bioinformatics/btw354
- Farrow, N. A., Zhang, O., Forman-Kay, J. D., and Kay, L. E. (1994). A heteronuclear correlation experiment for simultaneous determination of ¹⁵N longitudinal decay and chemical exchange rates of systems in slow equilibrium. *J. Biomol. NMR* 4, 727–734. doi: 10.1007/BF00404280
- Freese, N. H., Norris, D. C., and Loraine, A. E. (2016). Integrated genome browser: visual analytics platform for genomics. *Bioinformatics* 32, 2089–2095. doi: 10.1093/bioinformatics/btw069
- García, L., de Torre, J., Huertas, M. L., and Carrasco, B. (2000). HYDRONMR: prediction of NMR relaxation of globular proteins from atomic-level structures and hydrodynamic calculations. *J. Magn. Reson.* 147, 138–146. doi: 10.1006/jmre.2000.2170
- Gelhausen, R., Svensson, S. L., Froschauer, K., Heyl, F., Hadjeras, L., Sharma, C. M., et al. (2021). HRIPO: high-throughput analysis of bacterial ribosome profiling data. *Bioinformatics* 37, 2061–2063. doi: 10.1093/bioinformatics/btaa959
- Gelsinger, D. R., Dallon, E., Reddy, R., Mohammad, F., Buskirk, A. R., and DiRuggiero, J. (2020). Ribosome profiling in archaea reveals leaderless translation, novel translational initiation sites, and ribosome pausing at single codon resolution. *Nucleic Acids Res.* 48, 5201–5216. doi: 10.1093/nar/gkaa304
- Gray, T., Storz, G., and Papenfort, K. (2022). Small Proteins. *Big Questions. J. Bacteriol* 204:e0034121. doi: 10.1128/JB.00341-21
- Green, M. R., and Sambrook, K. (2012). *Molecular cloning: A laboratory manual*. Cold Spring Harbor, NY, USA: Cold Spring Harbor Laboratory Press.
- Gunde-Cimerman, N., Plemenitaš, A., and Oren, A. (2018). Strategies of adaptation of microorganisms of the three domains of life to high salt concentrations. *FEMS Microbiol. Rev.* 42, 353–375. doi: 10.1093/femsre/fuy009
- Hadjeras, L., Bartel, J., Maier, L.-K., Maaß, S., Vogel, V., Svensson, S. L., et al. (2023). Revealing the small proteome of *Haloferax volcanii* by combining ribosome profiling and small-protein optimized mass spectrometry. *Microfluid. 4:uqad001*. doi: 10.1093/femsml/uqad001
- Hammelmann, M., and Soppa, J. (2008). Optimized generation of vectors for the construction of *Haloferax volcanii* deletion mutants. *J. Microbiol. Methods* 75, 201–204. doi: 10.1016/j.mimet.2008.05.029
- Hassel, K. R., Brito-Estrada, O., and Makarewich, C. A. (2023). Microproteins: overlooked regulators of physiology and disease. *iScience* 26:106781. doi: 10.1016/j.isci.2023.106781
- Hirano, T., Kikuchi, K., Urano, Y., and Nagano, T. (2002). Improvement and biological applications of fluorescent probes for zinc, ZnAFs. *J. Am. Chem. Soc.* 124, 6555–6562. doi: 10.1021/ja025567p
- Jantzer, K., Zerulla, K., and Soppa, J. (2011). Phenotyping in the archaea: optimization of growth parameters and analysis of mutants of *Haloferax volcanii*. *FEMS Microbiol. Lett.* 322, 123–130. doi: 10.1111/j.1574-6968.2011.02341.x
- Kandiba, L., Lin, C.-W., Aebi, M., Eichler, J., and Guerardel, Y. (2016). Structural characterization of the N-linked pentasaccharide decorating glycoproteins of the halophilic archaeon *Haloferax volcanii*. *Glycobiology* 26, 745–756. doi: 10.1093/glycob/cww014
- Klein, C., Aivaliotis, M., Olsen, J. V., Falb, M., Besir, H., Scheffer, B., et al. (2007). The low molecular weight proteome of *Halobacterium salinarum*. *J. Proteome Res.* 6, 1510–1518. doi: 10.1021/pr060634q
- Krishna, S. S., Majumdar, I., and Grishin, N. V. (2003). Structural classification of zinc fingers: survey and summary. *Nucleic Acids Res.* 31, 532–550. doi: 10.1093/nar/gkg161
- Kubatova, N., Pyper, D. J., Jonker, H. R. A., Saxena, K., Remmel, L., Richter, C., et al. (2020). Rapid biophysical characterization and NMR spectroscopy structural analysis of small proteins from Bacteria and Archaea. *ChemBiochem* 21, 1178–1187. doi: 10.1002/cbic.201900677
- Laass, S., Monzon, V. A., Kliemt, J., Hammelmann, M., Pfeiffer, F., Förstner, K. U., et al. (2019). Characterization of the transcriptome of *Haloferax volcanii*, grown under four different conditions, with mixed RNA-Seq. *PLoS One* 14:e0215986. doi: 10.1371/journal.pone.0215986
- Lee, S.-J., Engelmann, A., Horlacher, R., Qu, Q., Vierke, G., Hebbeln, C., et al. (2003). TrmB, a sugar-specific transcriptional regulator of the trehalose/maltose ABC transporter from the hyperthermophilic archaeon *Thermococcus litoralis*. *J. Biol. Chem.* 278, 983–990. doi: 10.1074/jbc.M210236200
- Lee, S.-J., Moulakakis, C., Koning, S. M., Hausner, W., Thomm, M., and Boos, W. (2005). TrmB, a sugar sensing regulator of ABC transporter genes in *Pyrococcus furiosus* exhibits dual promoter specificity and is controlled by different inducers. *Mol. Microbiol.* 57, 1797–1807. doi: 10.1111/j.1365-2958.2005.04804.x
- Lee, W., Tonelli, M., and Markley, J. L. (2015). NMR-FAM-SPARKY: enhanced software for biomolecular NMR spectroscopy. *Bioinformatics* 31, 1325–1327. doi: 10.1093/bioinformatics/btu830
- Legerme, G., Yang, E., Esquivel, R. N., Kiljunen, S., Savilahti, H., and Pohlschroder, M. (2016). Screening of a *Haloferax volcanii* transposon library reveals novel motility and adhesion mutants. *Life (Basel)* 6. doi: 10.3390/life6040041
- Leong, A. Z.-X., Lee, P. Y., Mohtar, M. A., Syafruddin, S. E., Pung, Y.-F., and Low, T. Y. (2022). Short open reading frames (sORFs) and microproteins: an update on their identification and validation measures. *J. Biomed. Sci.* 29:19. doi: 10.1186/s12929-022-00802-5
- Liao, Y., Smyth, G. K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30, 923–930. doi: 10.1093/bioinformatics/btt656
- Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15:550. doi: 10.1186/s13059-014-0550-8
- Malakhov, M. P., Mattern, M. R., Malakhova, O. A., Drinker, M., Weeks, S. D., and Butt, T. R. (2004). SUMO fusions and SUMO-specific protease for efficient expression and purification of proteins. *J. Struct. Funct. Genom.* 5, 75–86. doi: 10.1023/B:JSFG.0000029237.70316.52
- Maligneri, G., Palmieri, M., Russo, L., Fattorusso, R., Pedone, P. V., and Isernia, C. (2015). The prokaryotic zinc-finger: structure, function and comparison with the eukaryotic counterpart. *FEBS J.* 282, 4480–4496. doi: 10.1111/febs.13503
- Marblestone, J. G., Edavettal, S. C., Lim, Y., Lim, P., Zuo, X., and Butt, T. R. (2006). Comparison of SUMO fusion technology with traditional gene fusion systems: enhanced expression and solubility with SUMO. *Protein Sci.* 15, 182–189. doi: 10.1110/ps.051812706
- Maret, W. (2013). Zinc biochemistry: from a single zinc enzyme to a key element of life. *Adv. Nutr.* 4, 82–91. doi: 10.3945/an.112.003038
- Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* 17. doi: 10.14806/ej.17.1.200
- Matthews, J. M., and Sunde, M. (2002). Zinc fingers--folds for many occasions. *IUBMB Life* 54, 351–355. doi: 10.1080/15216540216035
- Maurer, S., Ludt, K., and Soppa, J. (2018). Characterization of copy number control of two *Haloferax volcanii* replication origins using deletion mutants and Haloarchaeal artificial chromosomes. *J. Bacteriol.* 200. doi: 10.1128/JB.00517-17

- Montelione, G. T., and Wagner, G. (1989). 2D chemical exchange NMR spectroscopy by proton-detected heteronuclear correlation. *J. Am. Chem. Soc.* 111, 3096–3098. doi: 10.1021/ja00190a072
- Nagel, C., Machulla, A., Zahn, S., and Soppa, J. (2019). Several one-domain zinc finger μ -proteins of *Haloferax Volcanii* are important for stress adaptation, biofilm formation, and swarming. *Genes (Basel)* 10. doi: 10.3390/genes10050361
- Oren, A. (2008). Microbial life at high salt concentrations: phylogenetic and metabolic diversity. *Saline Syst* 4:2. doi: 10.1186/1746-1448-4-2
- Oren, A. (2013). Life at high salt concentrations, intracellular KCl concentrations, and acidic proteomes. *Front. Microbiol.* 4:315. doi: 10.3389/fmicb.2013.00315
- Orr, M. W., Mao, Y., Storz, G., and Qian, S.-B. (2020). Alternative ORFs and small ORFs: shedding light on the dark proteome. *Nucleic Acids Res.* 48, 1029–1042. doi: 10.1093/nar/gkz734
- Pfeiffer, F., Broicher, A., Gillich, T., Klee, K., Mejía, J., Rampp, M., et al. (2008). Genome information management and integrated data analysis with HaloLex. *Arch. Microbiol.* 190, 281–299. doi: 10.1007/s00203-008-0389-z
- Pohlschroder, M., and Esquivel, R. N. (2015). Archaeal type IV pili and their involvement in biofilm formation. *Front. Microbiol.* 6:190. doi: 10.3389/fmicb.2015.00190
- Pohlschroder, M., Pfeiffer, F., Schulze, S., and Abdul Halim, M. F. (2018). Archaeal cell surface biogenesis. *FEMS Microbiol. Rev.* 42, 694–717. doi: 10.1093/femsre/fuy027
- Schägger, H., and von Jagow, G. (1987). Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. *Anal. Biochem.* 166, 368–379. doi: 10.1016/0003-2697(87)90587-2
- Schlesinger, D., and Elsässer, S. J. (2022). Revisiting sORFs: overcoming challenges to identify and characterize functional micropoteins. *FEBS J.* 289, 53–74. doi: 10.1111/febs.15769
- Schulze, S., Pfeiffer, F., Garcia, B. A., and Pohlschroder, M. (2021). Comprehensive glycoproteomics shines new light on the complexity and extent of glycosylation in archaea. *PLoS Biol.* 19:e3001277. doi: 10.1371/journal.pbio.3001277
- Shalev, Y., Soucy, S. M., Papke, R. T., Gogarten, J. P., Eichler, J., and Gophna, U. (2018). Comparative analysis of surface layer glycoproteins and genes involved in protein glycosylation in the genus *Haloferax*. *Genes (Basel)* 9. doi: 10.3390/genes9030172
- Shen, Y., and Bax, A. (2013). Protein backbone and sidechain torsion angles predicted from NMR chemical shifts using artificial neural networks. *J. Biomol. NMR* 56, 227–241. doi: 10.1007/s10858-013-9741-y
- Shukla, H. D. (2006). Proteomic analysis of acidic chaperones, and stress proteins in extreme halophile *Halobacterium* NRC-1: a comparative proteomic approach to study heat shock response. *Proteome Sci.* 4:6. doi: 10.1186/1477-5956-4-6
- 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
- Steinberg, R., and Koch, H.-G. (2021). The largely unexplored biology of small proteins in pro- and eukaryotes. *FEBS J.* 288, 7002–7024. doi: 10.1111/febs.15845
- Storz, G., Wolf, Y. I., and Ramamurthi, K. S. (2014). Small proteins can no longer be ignored. *Annu. Rev. Biochem.* 83, 753–777. doi: 10.1146/annurev-biochem-070611-102400
- Tamir, A., and Eichler, J. (2017). N-glycosylation is important for proper *Haloferax volcanii* S-layer stability and function. *Appl. Environ. Microbiol.* 83. doi: 10.1128/AEM.03152-16
- Tarasov, V. Y., Besir, H., Schwaiger, R., Klee, K., Furtwängler, K., Pfeiffer, F., et al. (2008). A small protein from the bop-brp intergenic region of *Halobacterium salinarum* contains a zinc finger motif and regulates bop and crtB1 transcription. *Mol. Microbiol.* 67, 772–780. doi: 10.1111/j.1365-2958.2007.06081.x
- Tharakan, R., and Sawa, A. (2021). Minireview: novel micropeptide discovery by proteomics and deep sequencing methods. *Front. Genet.* 12:651485. doi: 10.3389/fgene.2021.651485
- Torrance, J. W., Macarthur, M. W., and Thornton, J. M. (2008). Evolution of binding sites for zinc and calcium ions playing structural roles. *Proteins* 71, 813–830. doi: 10.1002/prot.21741
- Varejão, N., and Reverter, D. (2023). Using intrinsic fluorescence to measure protein stability upon thermal and chemical denaturation. *Methods Mol. Biol.* 2581, 229–241. doi: 10.1007/978-1-0716-2784-6_16
- Vuister, G. W., and Bax, A. (1993). Quantitative J correlation: a new approach for measuring Homonuclear three-bond J(HNH α) coupling constants in 15N-enriched proteins. *J. Am. Chem. Soc.* 115, 7772–7777. doi: 10.1021/ja00070a024
- Weidenbach, K., Gutt, M., Cassidy, L., Chibani, C., and Schmitz, R. A. (2022). Small proteins in Archaea, a mainly unexplored world. *J. Bacteriol.* 204:e0031321. doi: 10.1128/JB.00313-21
- Wishart, D. S. (2011). Interpreting protein chemical shift data. *Prog. Nucl. Magn. Reson. Spectrosc.* 58, 62–87. doi: 10.1016/j.pnmrs.2010.07.004
- Zahn, S., Kubatova, N., Pyper, D. J., Cassidy, L., Saxena, K., Tholey, A., et al. (2021). Biological functions, genetic and biochemical characterization, and NMR structure determination of the small zinc finger protein HVO_2753 from *Haloferax volcanii*. *FEBS J.* 288, 2042–2062. doi: 10.1111/febs.15559
- Zaigler, A., Schuster, S. C., and Soppa, J. (2003). Construction and usage of a one-fold-coverage shotgun DNA microarray to characterize the metabolism of the archaeon *Haloferax volcanii*. *Mol. Microbiol.* 48, 1089–1105. doi: 10.1046/j.1365-2958.2003.03497.x
- Zhang, Z., Li, Y., Yuan, W., Wang, Z., and Wan, C. (2022). Proteomics-driven identification of short open reading frame-encoded peptides. *Proteomics* 22:e2100312. doi: 10.1002/pmic.202100312

Frontiers in Microbiology

Explores the habitable world and the potential of microbial life

The largest and most cited microbiology journal which advances our understanding of the role microbes play in addressing global challenges such as healthcare, food security, and climate change.

Discover the latest Research Topics

[See more →](#)

Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact

