



## OPEN ACCESS

## EDITED BY

Giovanna Suzzi,  
University of Teramo,  
Italy

## REVIEWED BY

Andres Fernando Gonzalez,  
University of Los Andes, Colombia  
Daniel Agyirifo,  
University of Cape Coast, Ghana

## \*CORRESPONDENCE

Aristóteles Góes-Neto  
arigoesneto@icb.ufmg.br

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

## SPECIALTY SECTION

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

RECEIVED 14 July 2022

ACCEPTED 04 October 2022

PUBLISHED 03 November 2022

## CITATION

Lima COC, De Castro GM, Solar R, Vaz ABM, Lobo F, Pereira G, Rodrigues C, Vandenberghe L, Martins Pinto LR, da Costa AM, Koblitz MGB, Benevides RG, Azevedo V, Uetanabaro APT, Soccol CR and Góes-Neto A (2022) Unraveling potential enzymes and their functional role in fine cocoa beans fermentation using temporal shotgun metagenomics. *Front. Microbiol.* 13:994524. doi: 10.3389/fmicb.2022.994524

## COPYRIGHT

© 2022 Lima, De Castro, Solar, Vaz, Lobo, Pereira, Rodrigues, Vandenberghe, Martins Pinto, da Costa, Koblitz, Benevides, Azevedo, Uetanabaro, Soccol and Góes-Neto. 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.

# Unraveling potential enzymes and their functional role in fine cocoa beans fermentation using temporal shotgun metagenomics

Carolina O. de C. Lima<sup>1†</sup>, Giovanni M. De Castro<sup>2</sup>, Ricardo Solar<sup>2</sup>, Aline B. M. Vaz<sup>2</sup>, Francisco Lobo<sup>2</sup>, Gilberto Pereira<sup>3</sup>, Cristine Rodrigues<sup>3</sup>, Luciana Vandenberghe<sup>3</sup>, Luiz Roberto Martins Pinto<sup>4</sup>, Andréa Miura da Costa<sup>4†</sup>, Maria Gabriela Bello Koblitz<sup>5†</sup>, Raquel Guimarães Benevides<sup>1</sup>, Vasco Azevedo<sup>2</sup>, Ana Paula Trovatti Uetanabaro<sup>2,4</sup>, Carlos Ricardo Soccol<sup>3</sup> and Aristóteles Góes-Neto<sup>1,2\*</sup>

<sup>1</sup>Department of Biological Sciences, State University of Feira de Santana (UEFS), Feira de Santana, Bahia, Brazil, <sup>2</sup>Institute of Biological Sciences, Federal University of the Minas Gerais (UFMG), Belo Horizonte, Minas Gerais, Brazil, <sup>3</sup>Bioprocess Engineering and Biotechnology Department, Federal University of the Paraná (UFPR), Curitiba, Paraná, Brazil, <sup>4</sup>Department of Biological Sciences, State University of Santa Cruz (UESC), Ilhéus, Bahia, Brazil, <sup>5</sup>Food and Nutrition Graduate Program (PPGAN), Federal University of the State of Rio de Janeiro (UNIRIO), Rio de Janeiro, Rio de Janeiro, Brazil

Cocoa beans fermentation is a spontaneous process, essential for the generation of quality starting material for fine chocolate production. The understanding of this process has been studied by the application of high-throughput sequencing technologies, which grants a better assessment of the different microbial taxa and their genes involved in this microbial succession. The present study used shotgun metagenomics to determine the enzyme-coding genes of the microbiota found in two different groups of cocoa beans varieties during the fermentation process. The statistical evaluation of the most abundant genes in each group and time studied allowed us to identify the potential metabolic pathways involved in the success of the different microorganisms. The results showed that, albeit the distinction between the initial (0h) microbiota of each varietal group was clear, throughout fermentation (24–144h) this difference disappeared, indicating the existence of selection pressures. Changes in the microbiota enzyme-coding genes over time pointed to the distinct ordering of fermentation at 24–48h (T1), 72–96h (T2), and 120–144h (T3). At T1, the significantly more abundant enzyme-coding genes were related to threonine metabolism and those genes related to the glycolytic pathway, explained by the abundance of sugars in the medium. At T2, the genes linked to the metabolism of ceramides and hopanoids lipids were clearly dominant, which are associated with the resistance of microbial species to extreme temperatures and pH values. In T3, genes linked to trehalose metabolism, related to the response to heat stress, dominated. The results obtained in this study provided insights into the potential functionality

of microbial community succession correlated to gene function, which could improve cocoa processing practices to ensure the production of more stable quality end products.

#### KEYWORDS

*Theobroma cacao*, cocoa beans fermentation, ecological succession, microbiome, functional analysis

## Introduction

Chocolate is one of the most popular food products in the world. Its production depends on several stages, from the planting of cocoa trees to the molding of bars and chocolates (De Vuyst and Leroy, 2020). An important part of this process takes place on the farm and involves the spontaneous fermentation of the cocoa pulp, by a complex microbial succession, which allows the removal of the pulp and creates specific conditions within the cotyledons. These reactions promote chemical and enzymatic transformations that lead to the acquisition of sensory characteristics desired for chocolate (Lima et al., 2020). As it is a spontaneous process, cocoa beans fermentation may often not work as it should, generating low-quality products and creating a shortage of this commodity in the market (Verce et al., 2021).

Metagenomics has revolutionized our view of microbial ecology with many potential applications for biotechnology (Cuadros-Orellana et al., 2013). Usually, microbiome studies utilize short-read high-throughput sequencing (HTS) platforms (e.g.: Illumina) that generate very high yields but of short read lengths (150–300 base pairs; Fonseca et al., 2022), as the current study. Conversely, long-read HTS platforms (e.g.: Oxford Nanopore Technologies) can sequence long DNA segments (Kbp to Mbp; Tomé et al., 2022). Long-read sequencing may help not only in alignment-based taxonomic and functional assignment due to its increased information content but also in bridging within- and between-genome repetitive sequences (Wommack et al., 2008; Bertrand et al., 2018). Nonetheless, the aforementioned advantages are still impaired by a high error rate if compared to short-read sequencing and; therefore, the lower accuracy of long-read sequencing affects the success rate of current classification methods, as well as there are few algorithms specifically designed to exploit long-read data for metagenomics (Nicholls et al., 2019; Pearman et al., 2020). Furthermore, metagenomic studies using long-read sequencing has been increasing since last years but mainly limited to amplicon-based metagenomics targeted to investigate the structure (taxonomic composition and relative abundance) of microbial communities (Matsuo et al., 2021; Lu et al., 2022), and shotgun metagenomic studies are still quite scarce (Ciuffreda et al., 2021). Therefore, short-read HTS still remains the most advantageous platform to perform shotgun metagenomics.

In order to enable controlled fermentation, extensive research has been carried out to elucidate the processes involved in fermentation and understand which events are desirable and which are unwanted or unnecessary (Schwan and Wheals, 2004; De Vuyst and Leroy, 2020). The most recent efforts have been applying metagenomic technologies to unravel the microbiota involved in fermentation, its succession, and, more importantly, the role of each taxonomic group in the transformations that occur throughout the process (Illegheems et al., 2015; Lima et al., 2020; Verce et al., 2021). Among the main advances of this approach were (1) the confirmation of the wide diversity of fermentative communities, previously under-elucidated by culture-dependent methods, including heretofore unaccounted groups, such as the Enterobacteriaceae (Illegheems et al., 2015); (2) the further clarification of the role of microbial succession and the groups involved in raw material consumption and environmental transformations, including the discovery that the desired transformations may happen even in the absence of some of the microbial groups formerly considered indispensable (Lima et al., 2020) and, (3) with the development of functional potential assessment techniques, an indication of which groups of genes - linked to which metabolic pathways - are associated with each taxonomic group and each set of transformations throughout the fermentation process (Almeida et al., 2020).

Much remains still to be unraveled; however, including information about which features determine the predominance of one taxonomic group over another throughout the fermentation period. In order to deeply investigate functional traits still unexplored and associated with cocoa fermentation, the current study applied shotgun metagenomics (using short-read NGS platform) to determine the enzyme-coding genes of the microbiota present in two different groups of cocoa beans varieties during different fermentation times (between 0 and 144 h). After normalizing their relative abundance, the enzyme-coding genes that were highlighted as significantly more abundant at each time of fermentation were selected, and their potential functionalities were analyzed. Therefore, it was possible to evaluate which groups of enzyme-coding genes were prominent according to the fermented cocoa beans variety and fermentation time, shedding light on the skill sets that ensure the permanence and success of the spontaneous microbiota.

## Materials and methods

### Cocoa samples and analyses

Samples of cocoa (*Theobroma cacao* L.) beans of Forastero variety (FOR) and a mixture of two hybrids (PS1319 and CCN51; MIX) of the fermentation process from the Riachuelo Agroindustry of Mendocá Chocolates<sup>1</sup> (Uruçuca, Bahia, Brazil; Lat-14.7719058; Long-39.0492701) were analyzed. On this farm, all the tree-to-bar processes (from cocoa planting to the chocolate bar) are done. Furthermore, the Good Manufacturing Practices (GMP) standards in the chocolate agroindustry production, and also encompassing the cocoa planting, management, harvesting, and fermentation process are practiced. Thus, both cocoa beans and chocolate have superior quality and are called fine or gourmet chocolate. Pods were washed, dried, opened, and immediately transferred to wooden boxes (45 × 45 × 45 cm, and capacity of 40 kg).

### Field experimental design

Two blocks (FOR and MIX) with three fermentation boxes in each one were sampled at seven distinct times: 0 (cocoa beans just after opening the pods), 24, 48, 72, 96, 120, and 144 h. The first spin was carried out after 48 h of fermentation, followed by spins every 24 h, ending on the seventh day of fermentation. Subsamples of 30 g were collected at five random points in the fermentation mass of each wooden box. Subsequently, mixed (total of 150 g sample), and part of it (100 g) was stored at -20°C for shotgun metagenomics, and part (50 g) was refrigerated (4°C) for physicochemical analyses. The methodology used to Field experimental design was the same as Lima et al. (2020).

### Metagenomic DNA extraction

From the lyophilized pulp fraction of the cocoa beans collected from the wooden boxes, the metagenomic DNA of the samples was extracted according to Cota-Sanchez et al. (2006), and with some modifications, as described by Lima et al. (2020).

An optimized protocol was used based on Cota-Sanchez et al. (2006) with some modifications: 20 g of cocoa mass (pulp + seeds) were placed in sterile glassware with 20 ml of sterilized ultrapure water and vigorously homogenized on a magnetic stirrer for 5 min. Pulp fraction was recovered by decantation and then lyophilized (LP3, Jouan). A total of 0.5 ml of lyophilized with 1 ml of sorbitol buffer 100 mm Tris base, 100 mm Sorbitol, 5 mm EDTA, 2% β-mercaptoethanol and 1% polyvinylpyrrolidone (PVP-40) were placed in a 2.0-ml microtube and centrifuged (Eppendorf 5804R) at 3000g for 10 min. Supernatant was

discarded, and liquid nitrogen was introduced to macerate the pellet with sterile rod. This procedure was repeated with sorbitol buffer until the presence of mucilage was no longer observed.

After the washing step, 500 μl of CTAB (100 mm Tris-HCL, 20 mm EDTA, pH 8.0, 2% CTAB, 1.5 NaCl, 4% PVP-40 and 10 mm 2-β-mercaptoethanol), 800 μl STE buffer (100 mm Tris-HCL, 50 mm EDTA, pH 8.0, 100 mm NaCl, 10 mm β-mercaptoethanol), 50 μl 20% SDS (sodium dodecyl sulfate), 10 μl RNase, 20 μl proteinase K, 50 mg.ml<sup>-1</sup> lysozyme were manually homogenized (inverting the tube) for 7 min. Subsequently it was incubated in water bath at 65°C for 30 min., with homogenizations at 5-min intervals. Afterwards, it was removed from the water bath, adding 415 μl 3 M ice-cold potassium acetate and incubated on ice for 40 min. Later, it was centrifuged at 25,805 g for 20 min. The same volume of ice-cold isopropanol (1: 1) was added, and gently homogenized and incubated at -20°C for 40 min. After, it was centrifuged at 25,805 g for 20 min and supernatant was discarded. Pellet was dried at room temperature, resuspended in 500 μl TE (100 mM Tris, 1 mM EDTA buffer pH 8.0) plus 500 μl of chloroform-isoamyl alcohol (24:1), gently mixed and centrifuged at 10,000 rpm for 10 min. Supernatant was transferred to a 1.5-ml microtube and 65 μl of 3 M sodium acetate and 600 μl of ice-cold isopropanol were added, gently mixed, and incubated overnight at -20°C. After that, it was centrifuged at 25,805 g for 20 min, and the supernatant was discarded. Pellet was washed with 1 ml ice-cold ethanol (96%) and centrifuged at 25,805 g for 5 min. Supernatant was discarded, and pellet was washed again with 1 ml ice-cold ethanol (75%) and centrifuged at 25,805 g for 5 min. Metagenomic DNA was resuspended in 40 μl ultrapure water and stored at -20°C until shotgun metagenomics sequencing.

### Massively parallel sequencing

The metagenomic libraries were prepared according to Illumina's standard protocols, and the sequencing was performed on an Illumina HiSeq X Ten (Novogene Sequencing Laboratory, UC Davis Medical Center). One μg of DNA per sample was used for sequencing, and the raw sequences were deposited in NCBI SRA under Bioproject accession PRJNA552479.

A total amount of 1 μg of DNA per sample was used for the preparation of metagenomic libraries. Libraries were constructed using DNA NEBNext<sup>®</sup> Ultra II kit for insertions of 350 bp. The libraries were analyzed by fragment size using Bioanalyzer Agilent 2,100 and quantified by real time PCR (3 nM). Sample clusters codified by indices were carried out in a cluster generation system cBot according to manufacturer instructions. After cluster generation, sequencing was performed on an Illumina HiSeq X Ten (Novogene Sequencing Laboratory, UC Davis Medical Center), generating 150 bp paired end reads. All the raw sequences were deposited NCBI SRA with Bioproject accession PRJNA552479 (SRR9640346, SRR9640343, SRR9640356, SRR9640355, SRR9640354,

<sup>1</sup> [www.mendoachocolates.com.br](http://www.mendoachocolates.com.br)

SRR9640353, SRR9640352, SRR9640351, SRR9640345,  
SRR9640344, SRR9640348, SRR9640347, SRR9640350,  
SRR9640349).

## Bioinformatics analysis

For preprocessing the data, we firstly used FASTQC v. 0.11.4 (Andrews, 2010) for quality analysis; CUTADAPT v. 1.18 (Martin, 2011) for adapter removal; BOWTIE2 (v. 2.3.4.3; Langmead and Salzberg, 2012) for Alignment in order to filter for pairs of reads not mapping on *Theobroma cacao* genome (NCBI Criollo cocoa genome V2); and, finally, FLASH v. 1.2.11 (Magoc and Salzberg, 2011) for merging filtered reads. For the taxonomic assignment, the strategies used were the same as in Lima et al. (2020).

Data pre-processing was as follows: (1) FASTQC (v0.11.4; Andrews, 2010) for quality analysis; (2) Adapter removal with CUTADAPT (v. 1.18; Martin, 2011); (3) Alignment using BOWTIE2 (v. 2.3.4.3; Langmead and Salzberg, 2012) to filter for pairs of reads not mapping on *Theobroma cacao* (genome from NCBI; Criollo cocoa genome V2); (4) Filtered reads were merged using FLASH (v1.2.11; Magoc and Salzberg, 2011).

For taxonomic assignment, two strategies were used: (a) the first one using all reads against a database, and (b) another using only the 16S and 18S rDNA reads. The first taxonomic analysis was performed by CENTRIFUGUE (v. 1.0.3; Kim et al., 2016) and KAIJU (v. 1.6.2; Menzel et al., 2016) and used a database composed of genomes selected from RefSeq and Genbank according to the following rules: (i) the genome was not excluded from RefSeq, except if it is from single cell, (ii) it is the latest version, (iii) only one genome per taxid, (iv) bacterial genomes were downloaded from RefSeq, (v) archaeal, fungal and viral genomes were downloaded from Genbank. The following rules were used for maintaining bacterial, fungal, and archaeal genomes: (i) genome was obtained from single cell, or (ii) it is a reference or representative genome, or (iii) it is related to type material, or (iv) it is a complete genome; (v) all viral genomes were downloaded from Genbank.

The genomes were download in September 2018 (Bacteria: 12,361; Archaea: 544; Fungi: 1,536; Virus: 14,132), and two indexes were generated, one for CENTRIFUGUE and another for KAIJU. The index for KAIJU used the same genomes as that of CENTRIFUGUE but, as KAIJU uses amino acids instead of nucleotides, the reads were translated in their six frames. Furthermore, due to RAM limitations, amino-acids sequences equal or less than 30 aa were discarded for KAIJU indexing. Reads were classified using both programs, and the results were merged, giving priority to CENTRIFUGUE results. The second taxonomic analysis was based on the 16S and 18S ribosomal subunits. Using the reads merged by FLASH, the ribosomal subunits sequences were extracted, classified, and mapped in OTUs using MAPSEQ (v. 1.2.3; Matias Rodrigues et al., 2017). The classification results of MAPSEQ were filtered using a combined score above 0.4 as recommended by the MAPSEQ authors. OTUs classified as metazoans or plants were removed.

## Functional prediction and quantification

For the assembly, after the previous processing of the reads, they were trimmed using Trimmomatic (v0.38; Bolger et al., 2014) with parameters “LEADING:15 TRAILING:15 SLIDINGWINDOW:5:15 MINLEN:50.” To further reduce the complexity of the metagenome, each sample had their reads binned by their assigned phylum, with the bins of each FOR samples merged with their respective bins and the same for the bins of MIX libraries. Then, each bin was assembled using Spades (v3.13.1; Nurk et al., 2017) with the parameter “--meta.” For the quantification, Salmon (v0.9.1; Patro et al., 2017) was used to generate an index with a merged file using all assemblies, and the contigs were quantified using the “--meta” parameter, to change the behavior of the software to quantify metagenomic reads, with the reads without adapters and before trimming. The TPM (Transcripts Per Million) value that is given by Salmon, which is normalized by the length of the contig, and the library size will be called here as CPM (Contigs Per Million) to avoid confusion as these are the values for each genomic contig.

The contigs from procaryote reads had their proteins predicted using Prodigal (v2.6.3; Hyatt et al., 2010) with the parameter “-p meta.” For the eukaryote contigs, their proteins were predicted using Augustus (v3.3.3; Stanke et al., 2008) with the parameter “--species=generic.” For functional annotation of the predicted proteins, the software Interproscan (v5.44-79; Jones et al., 2014) was used to predict metabolic pathways. Then, the proteins that were annotated with an EC number had the CPM from their respective contigs used as their quantification. For each sample, when there were multiple identical EC numbers from the same protein, they were counted as one, if there were different genes with the same EC number, then they were summed. All the complete scripts are in the [Supplementary material 1](#).

## Statistical analyses

In order to examine variation in enzyme-coding genes composition for each sample of FOR and MIX samples, we first used a Principal Component Analysis (PCA), transforming input variables to zero mean and unit variance (Husson et al., 2017). Then we performed non-metric multidimensional scaling (NMDS) of individual samples, using Hellinger transformation and the Bray-Curtis dissimilarity index (Legendre and Gallagher, 2001). We first visually examined the extent to which there was a structure in the ordination, and then we used PERMANOVA (Anderson, 2001) to test for significant clustering with respect to different varieties and fermentation times, using 999. We finally wanted to check whether multivariate dispersion could be different between varieties using PERMDISP (Anderson, 2001). All the analyses were performed in R (R Core Team, 2020), using the packages *FactoMineR*, *Vegan*, and *ggplot2*, and the complete script is in the [Supplementary material 2](#).

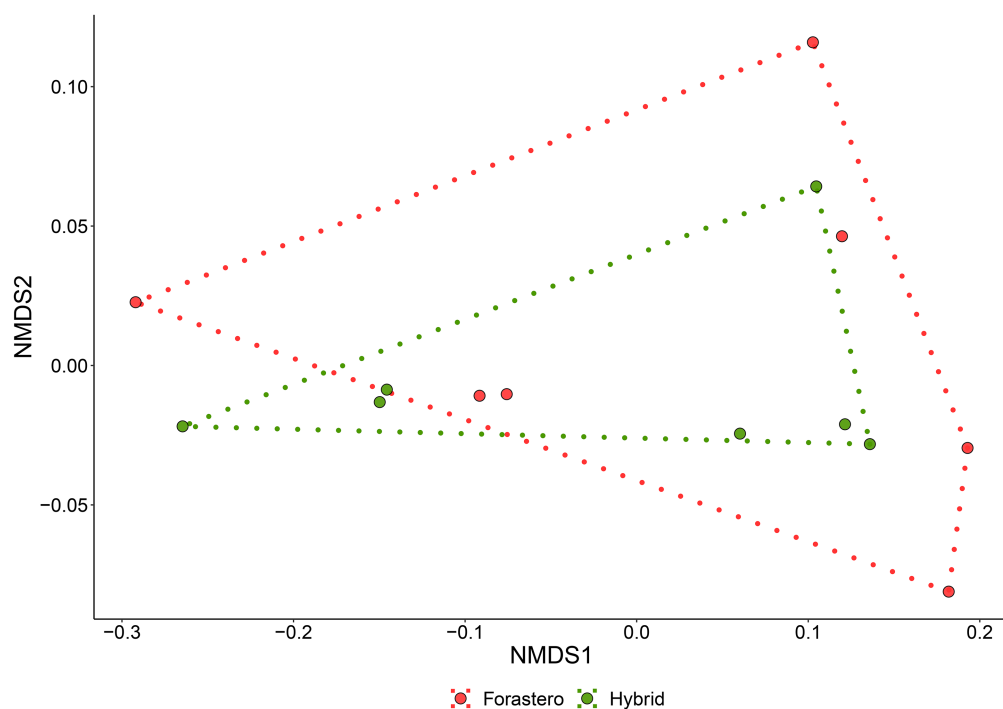


FIGURE 1

Permutational Dispersion plot of cocoa varieties samples in distinct fermentation times. Black circles for FOR (Forastero) and red triangles for MIX (mixture of two cocoa beans hybrid varieties).

The classification of the enzyme-coding genes associated with the main metabolic pathways (lipid, protein, carbohydrate, nucleotide, and micromolecule metabolisms) was based on enzyme EC and corresponding KEGG id using the website.<sup>2</sup> The variation of the abundance of enzyme-coding genes along the fermentation time was analyzed and visualized using color matrices. The results were compared over time [T1 (24–48 h) × T2 (72–96 h) and T2 × T3 (120–144 h)] and were considered significant when a value of  $p$  less than 0.05 and a fold change higher than 2.5 were found. Unique enzyme-coding genes have been present at all times; however, in very low abundances. Only the statistically significant shared enzyme-coding genes were then analyzed.

## Results

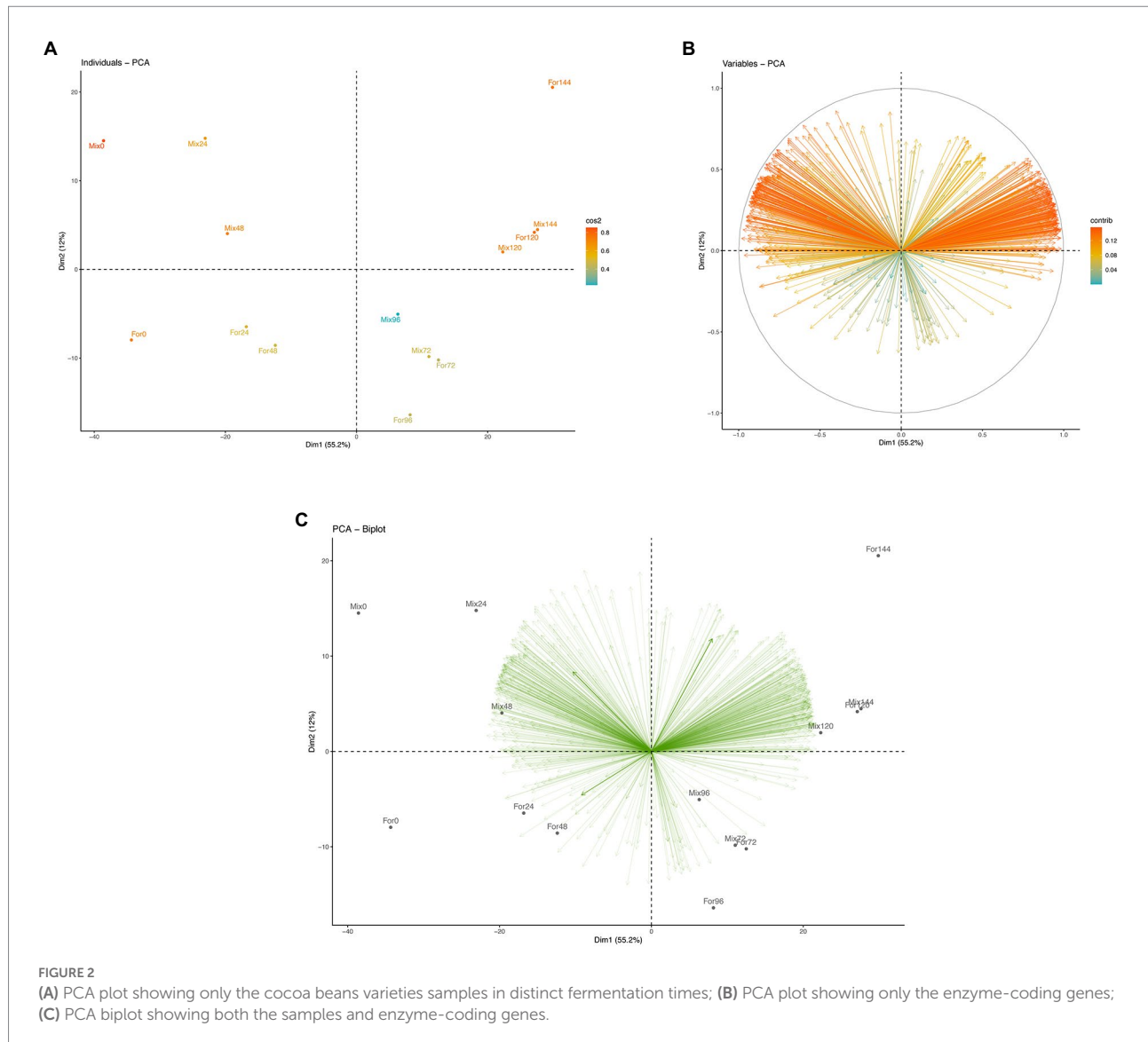
The non-metric multidimensional scaling (NMDS; Supplementary material 3) jointly with a permutational dispersion (PERMDISP) test (Figure 1) clearly indicated that (except for the T0 samples, which were not evaluated hereafter) the multivariate dispersion of the two cocoa beans varieties (FOR and MIX) was not significantly different, and, thus, they can be treated as duplicates of one same entity. Furthermore, the Principal

Component Analysis (PCA; Figures 2A–C) showed a clearly distinct ordination of both varieties of cocoa beans on different fermentation times 24–48 h, 72–96 h, and 120–144 h, which were hereafter named as T1, T2, and T3, respectively.

Most of the shared statistically significantly enzyme-coding genes in each T1, T2, and T3 times were directly associated with the dominant core microbiome fungal genus *Hanseniaspora* (T1) and bacterial genus *Acetobacter* (T2). *Hanseniaspora* is described as one of the most common fungal genus, besides *Saccharomyces* and *Pichia*, as revealed by both culture-dependent and culture-independent studies (Arana-Sanchez et al., 2015; De Vuyst and Weckx, 2016; Schwenninger et al., 2016; Koffi et al., 2017; Pereira et al., 2017; Mota-Gutierrez et al., 2018; Papalexandratou et al., 2019; Lima et al., 2020). *Acetobacter* has already been documented in cocoa fermentation, in different regions and by different methods (Camu et al., 2008; Garcia-armisen et al., 2010; Papalexandratou et al., 2011c; Pereira et al., 2012, 2013; Crafacek et al., 2013; Ramos et al., 2014; Hamdouche et al., 2015; Schwenninger et al., 2016; Visintin et al., 2016; Moreira et al., 2017; Gabaza et al., 2019; Lima et al., 2020). These data suggest that *Acetobacter* is also a core bacterial genus in cocoa bean fermentation worldwide.

Moreover, these aforementioned enzyme-coding genes were also related to other relatively abundant bacteria or fungi, such as *Pantoea*, *Lactobacillus*, *Frateuria*, *Candida*, and *Rhizopus* in T1; *Komagataeibacter*, *Frauteria*, *Gluconobacter*, *Pantoea*, and

<sup>2</sup> <https://www.brenda-enzymes.org/>



*Sphingomonas* in T2; and *Komagataeibacter* in T3 (Supplementary material 4).

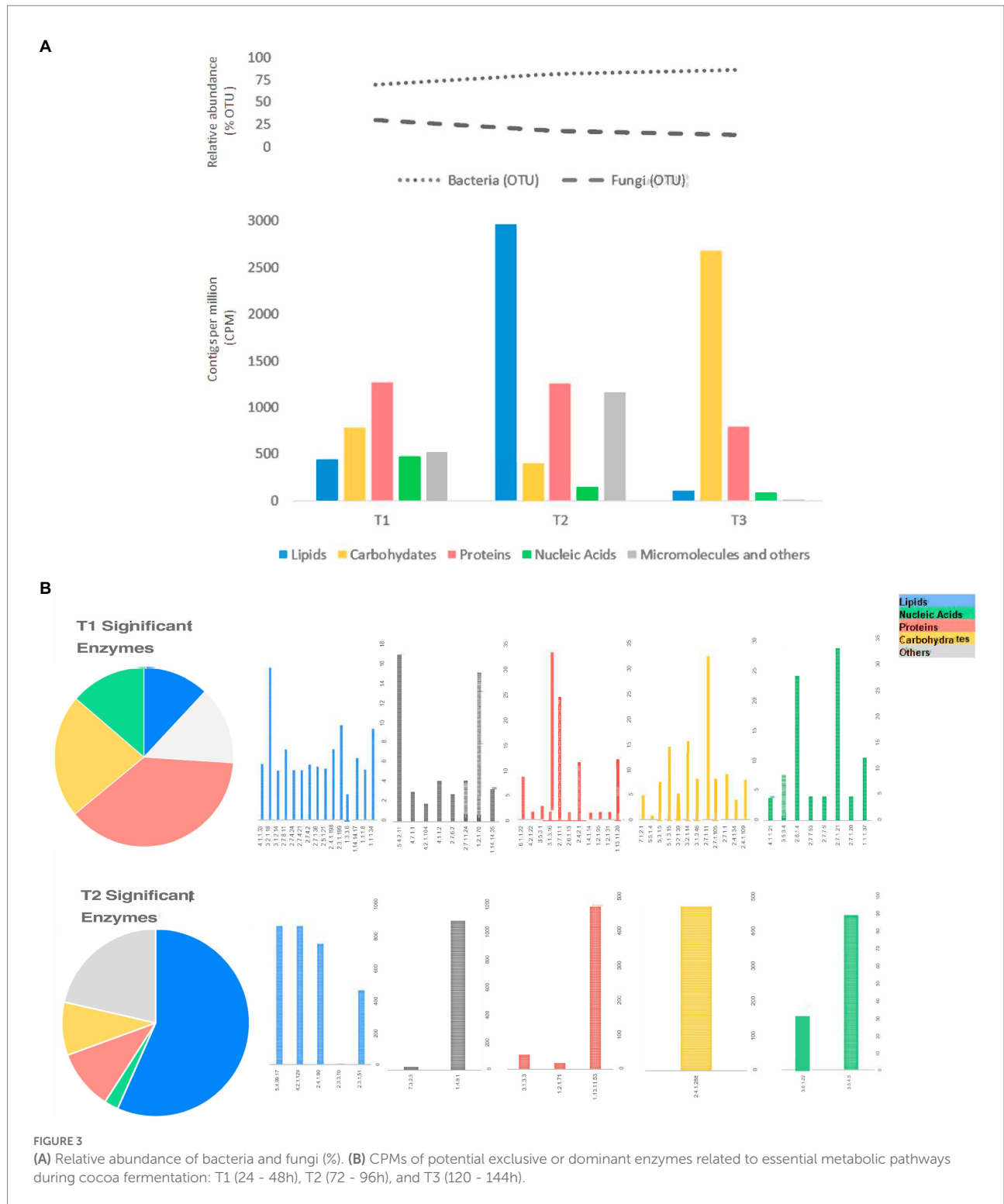
Regarding the genes encoding enzymes, there was a greater abundance for those related to the metabolism of proteins and carbohydrates within 48 h of fermentation. This coincides with the presence of a high concentration of bacteria and yeasts present in the fermentation medium whose results were previously found by our research group and reported in Lima et al. (2020). The abundance of genes related to lipids metabolism is evident in T2 followed by genes related to carbohydrate metabolism in T3; this set of genes could be related to the stress response to which the microorganisms are subjected since it coincides with changes in the environment in relation to pH and temperature that occur during the fermentation process of almonds (Figure 3).

Among the significant genes with greater abundance at different fermentation times are those involved in the metabolism of proteins and carbohydrates in T1, followed by the genes

involved in the metabolism of lipids in T2 and carbohydrates in T3 stand out (Figure 4).

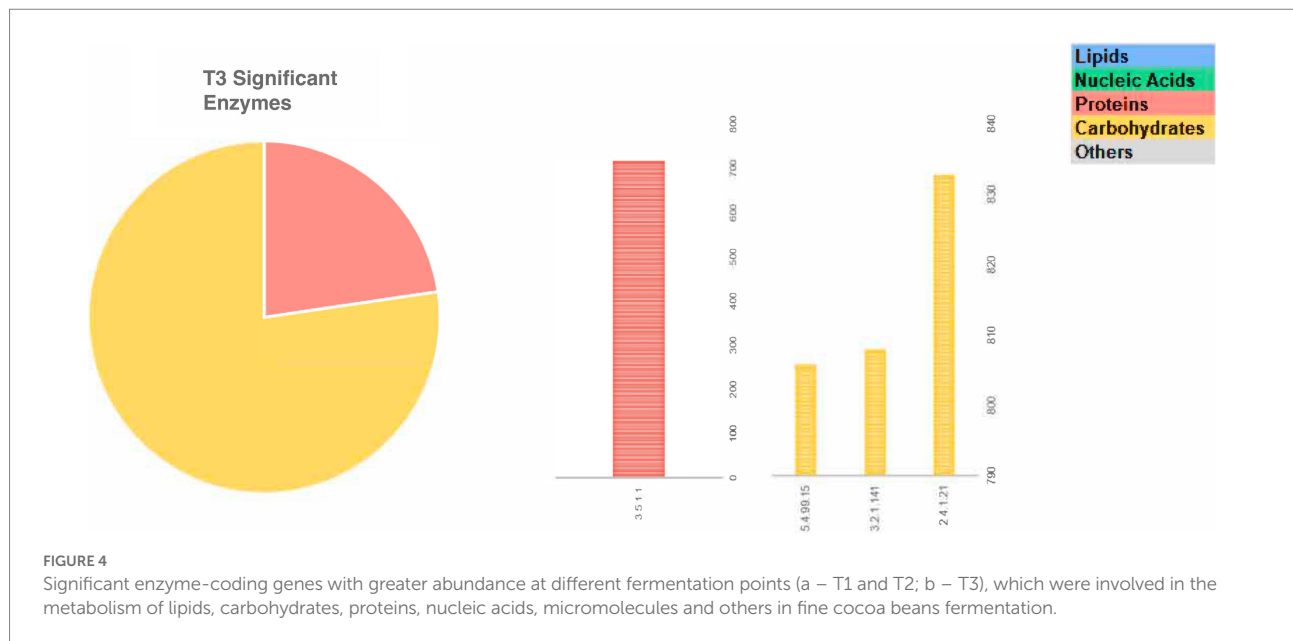
In general, at T1, a higher number of genes was found significantly more abundant, for all five categories evaluated (metabolism of lipids, carbohydrates, proteins, and amino acids, nucleic acids, and others). The categories with the highest number of significantly more abundant genes were 'lipid metabolism' and 'carbohydrate metabolism', with 15 representatives each, although the 'lipid related' genes accounted for just about 12% of the total abundance where 'carbohydrate related' genes reached almost 30%.

About lipid metabolism, of the 15 highlighted lipid-related genes at T1, 4 encode enzymes of the mevalonate pathway for the synthesis of isoprenoid precursors, such as farnesyl (1.1.1.34; 2.7.1.36; 4.1.1.33; 2.7.4.2), together representing 26% of the abundance of this category; 2 encode enzymes in the biosynthesis of steroids, specifically those that transform farnesyl into squalene and this into its epoxide (2.5.1.21; 1.14.14.17–11.5% of the



abundance); 3 encode enzymes in the metabolism of glycerophospholipid, related to the myoinositol phosphate metabolism pathway (1.1.1.8; 2.4.1.198; 2.7.8.11–19.5% of the abundance); and 2 are involved in the phosphatidylinositol signaling system (2.7.4.21; 2.7.4.24–10% of the abundance). One gene encodes an enzyme for the synthesis of fatty acids (3.1.2.14)

with the possibility of leading to the elongation of fatty acids or the metabolism of glycerophospholipids; 1 encodes an enzyme from the fatty acid elongation pathway (2.3.1.199), 1 encodes an enzyme of the fatty acid beta-oxidation/degradation pathway (1.3.3.6) and 1 encodes an enzyme in the sphingolipid metabolism (3.2.1.18). The latter gene was the most abundant in this category,



alone representing about 15% of the total abundance of genes encoding enzymes of the metabolism of lipids. At T2, the number of significantly more abundant genes dropped to just 13. Of these, five were related to lipid metabolism, which stood out in this stage of fermentation, corresponding to more than 58% of the total abundance found. Among the five highlighted genes, two stood out, with about 30% of the total abundance, each: both encoding enzymes of the metabolism of hopanoids – pentacyclic triterpenes (5.4.99.17 and 4.2.11.22), one encodes a glycosylceramide synthesizing enzyme in the sphingolipid metabolism (2.4.1.80–25% of the abundance), one encodes an enzyme in the metabolism of glycerophospholipid (2.3.1.51–15.65% of the abundance) and one, with much lower abundance than the others (2.3.3.10–0.13% of the abundance), encodes a hydroxy-methyl-glutaryl-CoA synthase, involved in several metabolic pathways, including the formation of the carbon skeleton for terpenoids.

Thus, despite lipid metabolism having not been directly related to the generation of chocolate flavor in cocoa fermentation to date, there are strong indications that it is essential for the survival of strains in the hostile environment that is installed along the process, which indirectly influence flavor formation.

Regarding carbohydrate metabolism, in T1, of the genes involved in carbohydrate metabolism, most are those involved in the glycolytic pathway such as 6-phosphofructokinase (2.7.1.11–20.5% of the abundance) and hexokinase (2.7.1.1–5.7%) and H<sup>+</sup>-exporting transporter (7.1.2.1–3%) involved in ADP phosphorylation in phosphorylation oxidative. The malate dehydrogenase gene (1.1.1.37) that catalyzes the conversion of malate to oxaloacetate in the tricarboxylic acid cycle (TCA) was found with an abundance of 5.75%. Regulatory genes for glycolysis and gluconeogenesis (6-phosphofructo-2-kinase, 2.7.1.105–5.3%) as well as those involved in the metabolism pathway of pentoses fructose-2,6-bisphosphate 2-phosphatase (3.1.3.46–5.3%),

phosphoglycerate mutase (5.4.2.11–15.25%), and glucose-6-phosphate 1-epimerase also are present (5.1.3.15–9.3%). The oxalate decarboxylase is involved in glyoxylate metabolism in the conversion of oxalate to formate (4.1.1.2–3.73%). Genes involved in chitin metabolism (3.2.1.14–10%) were also found and genes involved in the cell wall metabolism of microorganisms such as dolichyl-phosphate-mannose-protein (2.4.1.109–5%), inositol-3-phosphate synthase (5.5.1.4–0.5%), and 1, 3-beta-glucan synthase (2.4.1.34–2.4%). Genes involved in the interconversion of aldose and ketose sugars and in the metabolism of glycogen also appear (xylose isomerase - 5.3.1.5 and glucan endo-1,3-beta-D-glucosidase - 3.2.1.39) with 4.9 and 3.4% of the abundance, respectively.

In T2 the gene which encodes to arabinogalactan biosynthesis of peptidoglycan in mycobacterium is predominantly present (2.4.1.288). Nonetheless, at T3 the abundance of the gene could be involved in the trehalose disaccharides pathway (3.2.1.141, 5.4.99.15—both with 33% of the abundance each) and glycogen production (2.4.1.21–34%).

About protein and amino acid metabolism, considering the abundance of all the 11 genes present in fermentation at T1 related to amino acids and protein metabolism, the genes involved in serine/threonine dephosphorylation (3.1.3.16) and phosphorylation (2.7.11.1) are the most abundant, with 36 and 27%, respectively. The genes involved in the cysteine metabolic pathway (1.3.11.20, 4.2.1.22) account for 15% while for arginine and proline metabolism (2.6.1.13, 3.5.3.1) was 4.8%, lysine (1.2.1.31, 1.2.1.95) was 3.3% and glutamate (1.4.1.14) just 1.5% of abundance. The mitogen-activated protein kinase (MAPK; 2.7.11.24) signal transduction pathways, which are among the most widespread mechanisms of cellular regulation, and asparagine-tRNA ligase (6.1.1.22) were also present with 3.2 and 9.2% of abundance, respectively. Among the total of genes present



at T2, 11% of them correspond to genes related to the metabolism of proteins and amino acids and, of these, 88.9% abundance was found for the metabolism of cysteine and methionine (1.13.11.53) in the conversion of the methylterpene in methylpropionate and less abundance to arginine (1.2.1.71–3.5%) and serine (3.1.3.3–7.9%) pathway. However, at T3, of the total gene, 1 was to the proteins and amino acid metabolic pathway, and the asparagine gene (3.5.1.1–23% of abundance) involved in the conversion of L-asparagine to aspartate was found.

At T1, when compared to T2, genes encoding enzymes involved in nucleic acid metabolism accounted for 12.2% of the significantly more abundant genes. Six genes encoding enzymes related to the metabolism of purines and pyrimidines (2.7.1.21; 2.4.2.1; 3.5.3.4; 2.7.7.53; 2.7.1.20; 4.1.1.21). At T2, only one gene encoding an enzyme of the metabolism of nucleic acids was highly abundant: 3.5.4.5 - related to the metabolism of pyrimidines. The representativeness of this group of genes at T2 dropped notably, to about 1.8% of the total of the significantly more abundant genes, indicating the deceleration in DNA/microorganism duplication throughout fermentation. In addition to genes related to encoding metabolism enzymes from the four main groups of biomolecules, significantly more abundant genes encoding enzymes from other metabolic groups were also found. At T1, this group of genes represented about 12% of the genes considered. Among the genes in this group, participants in the metabolism of B-complex vitamins (2.8.1.4; 2.7.6.2 - thiamine and 6.3.1.20 - lipoic acid - 43% of the total), of porphyrins (1.2.1.70–35% of the total), of sulfur metabolism (1.14.14.35; 2.7.7.5–13.45% of the total), phosphorus (4.7.1.1) and nitrogen (4.2.1.104) were found. At T2, the genes of this group got to represent almost 23% of the highlighted genes abundance, the second most represented group, and almost 95% of the abundance was due to the gene encoding the enzyme methylamine dehydrogenase (1.4.9.1) of the metabolism of methane, although genes encoding enzymes of nicotinic acid and sulfur metabolism were also highlighted.

## Discussion

Based on the genes encoding enzymes, there was a greater abundance for those related to the metabolism of proteins and carbohydrates after 48 h of fermentation, with the greater presence of genes involved with the metabolism of fructose and glucose. These results are probably due to the abundance of bacteria of the genus *Gluconobacter* present in the cocoa fermentation, where they were the most abundant in the 0 h time. In addition, this bacterial genus was present in a relatively high abundance up to 48 h of fermentation (T1), being after that time overlapped by the most abundant presence of bacteria of the genus *Acetobacter* (Lima et al., 2020). The abundant presence of yeast of the genus *Hanseniospora* throughout the fermentation process also explains an abundance of genes related to the metabolism of carbohydrates at T1 and T3. According to Gálvez et al. (2007), *Hanseniospora* is

associated with the intense metabolism of sugars, generating ethanol, and exhibiting pectinolytic activity.

During cocoa beans fermentation, amino acids are precursors to produce different flavor compounds (Spinnler, 2012). We found genes coding for serine and threonine at T1 with greater abundance (63%) and at T2 (7.9%), these would be involved in the intermediary metabolic pathway to produce the amino acid valine, which is a precursor of saturated fatty acids. Studies with the genome of *Saccharomyces cerevisiae*, *Bacillus subtilis* subsp. *subtilis* str. 168, *Limosilactobacillus fermentum* IFO 3956 and *Acetobacter acetii* were used to build a metabolic pathway for L-leucine, L-phenylalanine, L-tyrosine, L-tryptophan, valine, and L-threonine showing that these amino acids are crucial as flavor and odor precursors and bioactive compounds in fine cocoa (Fernández-Niño et al., 2021).

The methionine may be involved in sulfur aroma compounds; these metabolites are also present in other fermentation processes such as in the production of cheese, wine, and fruit ripening (Spinnler, 2012). In our study, the presence of this gene was found in T2 (89% of the abundance) as well as cysteine, arginine, and serine, while at T3 the asparaginase represented the total of genes found for this metabolism. Similar to our results, the presence of some amino acid genes involved in flavor and aroma were found by Illegheems et al. (2015) that showed it could be associated with lactic acid bacteria (LAB) during the cocoa bean fermentation process.

Analyses carried out in other studies during cocoa fermentation showed that the peak concentration of peptides and amino acids is related to the increase in fermentation days (3–4 days) and the metabolic changes during fermentation associated with proteolysis, such as the production of peptides, amino acids, and polyphenols, are important to the formation of the typical flavor and aroma during the roasting of cocoa beans (Mayorga-Gross et al., 2016; Herrera-Rocha et al., 2021). Similarly, in the study, the abundance of protein-related genes was found at the beginning of fermentation, at T1 (24 - 48 h), which could be related to proteolysis and release of amino acids and peptides. Nevertheless, these compounds were not analyzed, and additional experiments would be necessary to elucidate this hypothesis in the cocoa beans fermentation process.

The presence of many genes involved in carbohydrate metabolism was expected since at the beginning of the fermentation process there is a great availability of this compound. The cocoa pulp is rich in simple carbohydrates (glucose and fructose), and sucrose (Verse Herrera-Rocha et al., 2021). The fact that many genes are presently involved in different pathways for the metabolization of carbohydrates is corroborated by the work described by Pothakos et al. (2020) in the Ecuadorian coffee fermentation process. The authors also verified several enzyme genes related to glycolysis and pentose-phosphate pathways which would be favoring the release of intermediates for the formation of lactic acid or acetic acid by the bacteria present at this time of fermentation. Besides that, the acetyl-CoA produced could be converted into ethanol. The production and consumption of

ethanol are reflected in the production of acetic acid by AAB (Verge et al., 2021), which explains the abundance of *Acetobacter* bacteria in the 24h of fermentation and remains until the end of the fermentation (Lima et al., 2020).

The abundance of genes encoding enzymes that are involved in the production of secondary metabolites was evident in T1 probably due to a response to the stress of the medium that was modified mainly by the action of AAB bacteria after 24h of fermentation, with consequent acidification of the medium and increased temperature (Lima et al., 2020). There are also reports in the literature relating the tolerance of AAB to the constitution of phospholipids, glycolipids, and carotenoids such as tetrahydroxybacteriohopane (THBH) in their cell membrane (Qiu et al., 2021).

The presence of the squalene-hopene cyclase gene is related to the formation of THBH in AAB cell membranes and this gene is abundantly present in T1 and T2 in the present work, which could confirm the involvement of THBH in resistance by AAB in the presence of high concentrations of acetic acid. In addition, the resistance of these bacteria to high levels of acetic acid is also due to the maintenance of low levels of intracellular acetic acid that occurs through the ABC transporter (Nakano et al., 2006; Qiu et al., 2021).

The abundant presence of yeast of the genus *Hanseniaspora* throughout the fermentation process (Lima et al., 2020) also explains an abundance of genes related to the metabolism of carbohydrates in T1 and T3. Studies indicate that there could be changes in the yeast cell wall related to fermentative processes that are halotolerant or osmotolerant, where greater flexibility of the cell wall, related to the presence of mannans, could be involved in the greater resistance of these microorganisms to these extreme environments compared to the microorganisms that have more rigid cell walls, which tend to be less halotolerant (Dakal et al., 2014). Thus, the abundance of genes “dolichyl-phosphate-mannose-protein” and “1,3-beta-glucan synthase” found in T1 could be involved in the modulation of  $\beta$ -D-glucan and cell wall mannans yeast synthesis which can undergo changes in responses to osmotic stress, considering the presence of high concentrations of sugar in the medium in the initial fermentation times.

Additionally, the presence of the abundance of genes (1->4)-alpha-D-glucan 1-alpha-D- glucosylmutase and 4-alpha-D-((1->4)-alpha-D-glucan) trehalose trehalohydrolase that catalyzes the trehalose synthesis in T3, could be related to the stress response to which microorganisms were exposed, mainly to osmotic and thermal stress. The presence of trehalose has been described as involved in bacterial and yeast responses to different environmental stress, such as resistance to osmotic pressure or in relation to extreme temperatures (De Virgilio et al., 1994; Iturriaga et al., 2009). Experiments carried out by Reina-Bueno et al. (2012) with *Chromohalobacter salexigens* mutants showed that trehalose synthesis is regulated by osmotic stress at the transcriptional level of the trehalose-6- phosphate synthase gene and in heat stress there is a post-transcriptional regulation, leading to an osmotic and thermoprotection. In general, it can be stated that the genes

related to lipid metabolism, significantly more abundant at T1, are related to the synthesis of squalene, glycerophospholipids, and sphingolipids. According to Lima et al. (2020), after T1 (24-48h) the microbial richness, evenness, and Shannon diversity drastically decrease, indicating the reduction in the number and diversity of the microorganisms in this process. At T2 the significantly more abundant genes are related to the synthesis of glucosyl-ceramide and hopanoids, the latter straight derived from squalene (Belin et al., 2018). All these products can be directly related to microbial resistance to hostile environments, whether due to changes in pH or to elevated temperatures.

Hopanoids are branched cyclic triterpene compounds derived from squalene, which are similar to sterols. Similarly, to these compounds, hopanoids can associate with other lipids in the cell membrane of prokaryotes, altering their fluidity and permeability. Hopanoids concentration in the cell membrane of bacteria can vary from less than 1% to more than 90%, in peripheral, intra-, and extracellular membranes, depending on the species and on the environmental conditions, and their presence leads to the condensation, thickening, and reduction of permeability of these membranes. Not only the concentration of hopanoids but also the type and the occurrence of interactions with sphingolipids, forming the so-called ‘lipid rafts’, interfere in the properties of the membranes, promoting resistance without significant loss of fluidity. Hopanoids are related to the survival of bacteria under stressful conditions. Their concentration rises at high temperatures and in the presence of acidic pH, such as found during cocoa fermentation. Apparently, the presence of hopanoids in membranes reduces the loss of cations and protons to the acidic environment, as well as protects the membrane structure against thermolysis in a heated environment (Belin et al., 2018; Santana-Molina et al., 2020).

Bacteria and fungi also use sphingolipids and ceramides to modulate the melting point of cell membranes to overcome high temperatures. Sphingolipids show a higher melting point than membrane phospholipids and their composition may be changed according to the temperature of the environment to ensure fluidity and permeability of cell membranes. In yeasts, the acquisition of thermotolerance depends on the fast accumulation of ceramides and sphingolipids such as glucosyl- ceramide (although this is not produced by *S. cerevisiae*) and glycosyl-inositol-phosphoryl-ceramide. The accumulation of these products is related to the long-term thermal resistance in fungi (Fabri et al., 2020).

Our results seem to disagree with the data published by Herrera-Rocha et al. (2022) and also by Servent et al. (2018). In the former, the authors concluded that there is no change in the content of lipids during fermentation, including glycerophospholipids, the second most abundant group of lipids in fermenting cocoa. In the latter, the authors found evidence of the release of fatty acids, possibly by lipases - not found in the present study - but did not evidence change in the total lipid content, nor did they mention the presence of glycerophospholipids. However, Servent et al. (2018) made it clear that the behavior of lipids in cocoa was dependent on the fermentation origin, as demonstrated by PCA. This leads us to

believe that similar to other transformations that occur during cocoa fermentation, the changes in the lipid fraction are directly dependent on the composition of the microbiota involved and indicates that, in the case of the present study, some change in the composition of glycerophospholipids is to be expected. As there are no details on the fermentative microbiota in [Herrera-Rocha et al. \(2022\)](#), nor in [Herrera-Rocha et al. \(2021\)](#) nor in [Servent et al. \(2018\)](#), the proof of this hypothesis will depend on additional experiments.

The results obtained indicate that the genomic machinery involved in resistance to hostile environments, such as acidic pH and high temperature, is essential for the microbiota of cocoa fermentation, regardless of the cocoa varieties used as substrate and of the initial fermentation microbiota, providing a significant advantage to those resistant species and ensuring their survival in this specific process/environment.

The *de novo* synthesis of purine and pyrimidine bases is essential for cell multiplication, which may explain the high representation of genes related to this function in the initial stages of fermentation—the metabolism of nucleic acids at T1 stood out as the 3rd largest group of genes among those significantly more abundant, only behind the metabolism of carbohydrates and proteins which, according to [Almeida et al. \(2020\)](#), represent the main portion of the functionalities found in the metagenome of cocoa spontaneous fermentation. However, also according to the same authors, it is possible that the metabolism of nucleic acids works as a secondary source of energy, related to the pentose-phosphate pathway.

Among the genes encoding enzymes for the metabolism of biomolecules other than carbohydrates, proteins, lipids, and nucleic acids, those related to the metabolism of B-complex vitamins were remarkable at T1. Two genes, corresponding to 38.6% of the total, 1 encoding an enzyme at the beginning of the thiamine synthesis pathway from cysteine and another encoding an enzyme at the end of this pathway, for the generation of thiamine diphosphate (TPP), were found among the genes significantly more abundant. In addition, 1 gene encoding a ligase from the metabolism of lipoic acid was also evidenced. Thiamine and lipoic acid are important coenzymes of energy metabolism. Their presence is essential for the activity of the pyruvate-dehydrogenase complex and is also necessary for other decarboxylation activities in the metabolism of carbohydrates, amino acids, and fatty acids. According to [Agyirifo et al. \(2019\)](#), yeasts, such as those of the genera *Saccharomyces* and *Hanseniaspora*, found in spontaneous cocoa beans fermentations, can produce large amounts of thiamine, enriching the medium, which can be beneficial for the subsequent development of lactic acid bacteria, especially of the genus *Lactobacillus*, dependent on thiamine from the environment to grow ([Carr et al., 2002](#)). It is interesting to note that lipoate synthesis always occurs as bound to proteins, from which it can be separated by specific enzymes (lipoamidases; [Cronan et al., 2005](#)). Some organisms can use protein-free lipoic acid through the activity of the enzyme lipoate-protein ligase, of which the encoding gene was found enriched at T1, and that uses ATP to bind lipoate to proteins and re-establish

its activity ([Cronan et al., 2005](#)). The ability to synthesize this enzyme could mean a competitive advantage for a species to establish itself in an adverse environment.

Glutamyl-tRNA reductase (1.2.1.70), encoded by another enriched gene at T1, makes up the first part of the metabolic pathway for the synthesis of 5-aminolevulinate from glutamate. This metabolite can generate several porphyrins, including protoporphyrin IX, which generates the 'heme' group, essential for the formation of cytochromes a and c and various oxidative enzymes, such as peroxidases and catalases, among others. This variety of applications can explain the representativeness of this gene, accounting for more than 1/3 of the total. Verse [Herrera-Rocha et al. \(2021\)](#) related different porphyrin-active proteins (oxygen-dependent coproporphyrinogen-III oxidase; peroxiredoxin) to oxygen-sensing and stress during cocoa fermentation.

The gene encoding the enzyme sulfate adenylyltransferase (2.7.7.5) was also found enriched in T1. This enzyme catalyzes the formation of APS (adenylyl-sulfate) from sulfate captured from the environment. APS is the starting compound for different forms (assimilative and dissimilative) of sulfur metabolism, and also participates in the synthesis of methionine in eukaryotes (*S. cerevisiae*). The other gene encoding an enzyme linked to sulfur metabolism, dimethylsulfone monooxygenase (1.14.14.35), found enriched at T1, generates methane sulfonate, which, in turn, generates sulfite and sulfide, the latter destined for the synthesis of cysteine. To our knowledge, sulfur metabolism has not been mentioned/explored in recent work on cocoa fermentation microbiota.

The gene encoding the enzyme alpha-D-ribose 1-methylphosphonate 5-phosphate C-P lyase (4.7.1.1) was also found enriched at T1. This enzyme is part of the C-P lyase pathway, which involves an enzyme complex capable of transforming phosphonates (organic molecules containing stable covalent bonds between phosphorus and carbon, which are difficult to lyse and use) into phosphates, that can be used by cell metabolism. Some bacteria have this enzymatic machinery that gives them an advantage in the use of phosphorus from the environment in forms other than soluble phosphate ([Manav et al., 2018](#)). Possibly, the presence of these genes provides a competitive advantage for the permanence/predominance of these bacteria in the adverse environment of cocoa fermentation.

The gene encoding cyanase also stood out among those significantly more abundant in T1. The enzyme cyanate lyase (4.2.1.104) degrades cyanate to carbon dioxide and ammonia using bicarbonate as a co-substrate and allows bacteria to use cyanate as the sole source of nitrogen in environments with low availability of this element, also providing a competitive advantage to the microorganisms that establish themselves in the spontaneous fermentation of cocoa ([Linder, 2019](#)).

At T2 the gene encoding the amicyanin enzyme (1.4.9.1) was substantially noticed. This enzyme is part of a complex of 3 electron-transporting proteins (methylamine dehydrogenase, blue copper protein amicyanin, and cytochrome c551i) that leads to the

deamination of methylamine generating formaldehyde and ammonia (Chen et al., 1994), in the methane cycle. It is possible that this gene was highlighted because of the activity of C-P-lyase (4.7.1.1) at T1 that generates methane.

Also highlighted in T2, although in much less abundance, it was reported the gene encoding the enzyme NAD<sup>+</sup> diphosphatase (3.6.1.22). Its activity generates NAD<sup>+</sup> and deamino-NAD<sup>+</sup> from nicotinamide-D-ribonucleotide and nicotinate-D-ribonucleotide, respectively. These are extremely important coenzymes in general and energy metabolism. According to Agyirifo et al. (2019), yeasts, characteristic of the beginning of cocoa fermentation, are producers of nicotinic acid and nicotinamide, which are the basis for the predominant microorganisms in T2 to synthesize the coenzymes necessary for their metabolism.

The changes/differences in abundance of the genes across T1, T2, and T3 can be stated on changes in microbiome conditions during the period of fermentation. Nutrients and metabolites, pH, temperature could be impacting on species diversity and their metabolism.

Another gene encoding a sulfur metabolism enzyme also stood out among the significantly more abundant genes – the ABC-type sulfate transporter (73.2.3), a transporter protein responsible for the active uptake of sulfate from the environment, feeding the sulfur pathway. Its activity is immediately followed by the activity of the sulfate adenylyltransferase enzyme (2.7.7.5), of which the encoding gene was found enriched at T1, reinforcing the importance of sulfur metabolism for the permanence/predominance of cocoa fermentation microbiota over time.

The results obtained help to better understand the community metabolism of microbial succession throughout the spontaneous fermentation of different varieties of fine cocoa beans unraveling potential enzymes involved in that process, which could improve cocoa processing practices and more stable quality end products.

## 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

CL: conceptualization, methodology, validation, investigation, writing—original draft, writing—review and editing, and visualization. AV: formal analysis and writing—review and editing. GC: methodology, software, formal analysis, writing—original draft, and writing—review and editing. FL: software, formal analysis, writing—review and editing. RS: methodology, formal analysis, writing—original draft, writing—review and editing. CR: methodology and writing—review and editing. LP: methodology and formal analysis. LV: writing—review and editing and project administration. GP, VA, and RB: writing—review and editing. AC,

AU, and MK: conceptualization, methodology, validation, investigation, writing—original draft, writing—review and editing, and visualization. CS: project administration and funding acquisition. AG-N: conceptualization, methodology, validation, investigation, writing—original draft, writing—review and editing, visualization, project administration, and funding acquisition. All authors contributed to the article and approved the submitted version.

## Funding

This work was funded by the Coordination of Superior Level Staff Improvement (CAPES, Grant PROCAD 88881.068458/2014–01). The funder had no role in study design, data collection, analysis, and decision to publish or prepare the manuscript. AG-N receives a research grant for productivity from the National Council for Scientific and Technological Development (CNPq), Brazil (no. 310764/2016–5).

## Acknowledgments

The authors would like to thank all the colleagues that contributed directly or indirectly to this work: the Graduate Program in Biotechnology of the State University of Feira de Santana (UEFS), the Graduate Programs in Microbiology, and in Bioinformatics of Federal University of Minas Gerais (UFMG), and, especially Raimundo Mororó, at Riachuelo agroindustry of Mendoá Chocolates. We are also grateful to the Research Support Foundation of the state of Minas Gerais (FAPEMIG) and the Research Support Foundation of the state of Bahia (FAPESB).

## 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.2022.994524/full#supplementary-material>

## References

- Agyirifo, D. S., Wamalwa, M., Otwe, E. P., Galyuon, I., Runo, S., Takrama, J., et al. (2019). Metagenomics analysis of cocoa bean fermentation microbiome identifying species diversity and putative functional capabilities. *Heliyon* 5:e02170. doi: 10.1016/j.heliyon.2019.e02170
- Almeida, O. G. G., Pinto, U. M., Matos, C. B., Frazilio, D. A., Braga, V. F., von Zeska-Kressa, M. R., et al. (2020). Does quorum sensing play a role in microbial shifts along spontaneous fermentation of cocoa beans? An *in silico* perspective. *Food Res. Int.* 131:109034. doi: 10.1016/j.foodres.2020.109034
- Anderson, M. J. (2001). A new method for non-parametric multivariate analysis of variance. *Austral Ecol.* 26, 32–46. doi: 10.1111/j.1442-9993.2001.01070.pp.x
- Andrews, S. (2010). FastQC A Quality Control tool for High Throughput Sequence Data. Available at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
- Arana-Sanchez, A., Segura-garcía, L. E., Kirchmayr, M., Orozco-avila, I., Lugocervantes, E., and Gschaedler-mathis, A. (2015). Identification of predominant yeasts associated with artisan Mexican cocoa fermentations using culture-dependent and culture-independent approaches. *World J. Microbiol. Biotechnol.* 31, 359–369. doi: 10.1007/s11274-014-1788-8
- Belin, B. J., Busset, N., Giraud, E., Molinaro, A., Silipo, A., and Newman, D. K. (2018). Hopanoid lipids: from membranes to plant–bacteria interactions. *Nat. Rev. Microbiol.* 16, 304–315. doi: 10.1038/nrmicro.2017.173
- Bertrand, D., Shaw, J., Narayan, M., et al. (2018). Nanopore sequencing enables high-resolution analysis of resistance determinants and mobile elements in the human gut microbiome. *bioRxiv* [Preprint]. doi: 10.1101/456905
- 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
- Camu, N., Gonzalez, A., De Winter, T., Van Schoor, A., De Bruyne, K., Vandamme, P., et al. (2008). Influence of turning and environmental contamination on the dynamics of lactic acid bacteria and acetic acid bacteria populations involved in spontaneous cocoa bean heap fermentation in Ghana. *Appl. Environ. Microbiol.* 74, 86–98. doi: 10.1128/AEM.01512-07
- Carr, F. J., Chill, D., and Maida, N. (2002). The lactic acid bacteria: a literature survey. *Crit. Rev. Microbiol.* 28, 281–370. doi: 10.1080/1040-840291046759
- Chen, L., Durlay, R. C., Mathews, F. S., and Davidson, V. L. (1994). Structure of an electron transfer complex: methylamine dehydrogenase, amicyanin, and cytochrome c551i. *Science* 264, 86–90. doi: 10.1126/science.8140419
- Ciuffreda, L., Rodríguez-Pérez, H., and Flores, C. (2021). Nanopore sequencing and its application to the study of microbial communities. *Comput. Struct. Biotechnol. J.* 19, 1497–1511. doi: 10.1016/j.csbj.2021.02.020
- Cota-Sanchez, J. H., Remarchuk, K., and Ubayasena, K. (2006). Ready-to-use DNA extracted with a CTAB method adapted for herbarium specimens and mucilaginous plant tissue. *Plant molecular biology reporter*, 24. *Article* 24, 161–167. doi: 10.1007/BF02914055
- Crafack, M., Mikkelson, M. B., Saerens, S., Knudsen, M., Blennow, A., Lowor, S., et al. (2013). Influencing cocoa flavor using *Pichia kluyveri* and *Kluyveromyces marxianus* in a defined mixed starter culture for cocoa fermentation. *Int. J. Food Microbiol.* 167, 103–116. doi: 10.1016/j.ijfoodmicro.2013.06.024
- Cronan, J. E., Zhao, X., and Jiang, Y. (2005). Function, attachment and synthesis of lipoic acid in *Escherichia coli*. *Adv. Microb. Physiol.* 50, 103–146. doi: 10.1016/S0065-2911(05)50003-1
- Cuadros-Orellana, S., Leite, L. R., Smith, A., Medeiros, J. D., Badotti, P. L. C., Vaz, A. B. M., et al. (2013). Assessment of fungal diversity in the environment using metagenomics: a decade in review. *Fungal Genet. Biol.* 3, 110–123. doi: 10.4172/2165-8056.1000110
- Dakal, T. C., Solieri, L., and Giudici, P. (2014). Adaptive response and tolerance to sugar and salt stress in the food yeast *Zygosaccharomyces rouxii*. *Int. J. Food Microbiol.* 185, 140–157. doi: 10.1016/j.ijfoodmicro.2014.05.015
- De Virgilio, C., Hottiger, T., Dominguez, J., Bolle, T., and Wiemken, A. (1994). The role of trehalose synthesis for the acquisition of thermotolerance in yeast. *Eur. J. Biochem.* 219, 179–186. doi: 10.1111/j.1432-1033.1994.tb19928.X
- De Vuyst, L., and Leroy, F. (2020). Functional role of yeasts, lactic acid bacteria and acetic acid bacteria in cocoa fermentation processes. *FEMS Microbiol. Rev.* 44, 432–453. doi: 10.1093/femsre/fuaa014
- De Vuyst, L., and Weckx, S. (2016). The cocoa bean fermentation process: from ecosystem analysis to starter culture development. *J. Appl. Microbiol.* 121, 5–17. doi: 10.1111/jam.13045
- Fabri, J. H. T. M., Sá, N. P., Malavazi, I., and Poeta, M. D. (2020). The dynamics and role of sphingolipids in eukaryotic organisms upon thermal adaptation. *Prog. Lipid Res.* 80:101063. doi: 10.1016/j.plipres.2020.101063
- Fernández-Niño, M., Rodríguez-Cubillos, M. J., Herrera-Rocha, F., Anzola, J. M., Cepeda-Hernández, M. L., Mejía, J. L. A., et al. (2021). Dissecting industrial fermentations of fine flavour cocoa through metagenomic analysis. *Sci. Rep.* 11:8638. doi: 10.1038/s41598-021-88048-3
- Fonseca, P. L., Skaltsas, D., da Silva, F. F., Kato, R. B., de Castro, G. M., García, G. J., et al. (2022). An integrative view of the Phyllosphere Mycobiome of native rubber trees in the Brazilian Amazon. *Journal of Fungi.* 8:373. doi: 10.3390/jof8040373
- Gabaza, M., Joossens, M., Cnockaert, M., Muchuweti, M., Raes, K., and Vandamme, P. (2019). Lactococci dominate the bacterial communities of fermented maize sorghum and millet slurries in Zimbabwe. *Int. J. Food Microbiol.* 289, 77–87. doi: 10.1016/j.ijfoodmicro.2018.09.001
- Gálvez, S. L., Loiseau, G., Paredes, J. L., Barel, M., and Guiraud, J. P. (2007). Study on the microflora and biochemistry of cocoa fermentation in the Dominican Republic. *Int. J. Food Microbiol.* 114, 124–130. doi: 10.1016/j.ijfoodmicro.2006.10.041
- García-Armisen, T., Papalexandratou, Z., Hendryckx, H., Camu, N., Vrancken, G., De Vuyst, L., et al. (2010). Diversity of the total bacterial community associated with Ghanaian and Brazilian cocoa bean fermentation samples as revealed by a 16 S rRNA gene clone library. *Appl. Microbiol. Biotechnol.* 87, 2281–2292. doi: 10.1007/s00253-010-2698-9
- Hamdouche, Y., Guehi, T., Durand, N., Kedjebo, K. B. D., Montet, D., and Meile, J. C. (2015). Dynamics of microbial ecology during cocoa fermentation and drying: towards the identification of molecular markers. *Food Contr.* 48, 117–122. doi: 10.1016/j.foodcont.2014.05.031
- Herrera-Rocha, , Cala, M. P., Aguirre Mejía, J. L., Rodríguez-López, C. M., Chica, M. J., Olarte, H. H., et al. (2021). Dissecting fine-flavor cocoa bean fermentation through metabolomics analysis to break down the current metabolic paradigm. *Sci. Rep.* 11:21904. doi: 10.1038/s41598-021-01427-8
- Herrera-Rocha, , Cala, M. P., León-Inga, A. M., Aguirre Mejía, J. L., Rodríguez-López, C. M., Florez, S. L., et al. (2022). Lipidomic profiling of bioactive lipids during spontaneous fermentations of fine-flavor cocoa. *Food Chem.* 397:133845. doi: 10.1016/j.foodchem.2022.133845
- Husson, F., Le, S., and Pages, J. (2017). *Exploratory Multivariate Analysis by Example Using R, Chapman, and Hall. 2nd Edn.* Chapman and Hall/CRC.
- Hyatt, D., Chen, G.-L., LoCasio, P. F., Land, M. L., Larimer, F. W., and Hauser, L. J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11:119. doi: 10.1186/1471-2105-11-119
- Illegheems, K., Weckx, S., and De Vuyst, L. (2015). Applying meta-pathway analyses through metagenomics to identify the functional properties of the major bacterial communities of a single spontaneous cocoa bean fermentation process sample. *Food Microbiol.* 50, 54–63. doi: 10.1016/j.fm.2015.03.005
- Iturriaga, G., Suárez, R., and Nova-Franco, B. (2009). Trehalose metabolism: from osmoprotection to signaling. *Int. J. Mol. Sci.* 10, 3793–3810. doi: 10.3390/ijms10093793
- Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., et al. (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30, 1236–1240. doi: 10.1093/bioinformatics/btu031
- Kim, D., Song, L., Breitwieser, F. P., and Salzberg, S. L. (2016). Centrifuge: rapid and sensitive classification of metagenomic sequences. *Genome Res.* 26, 1721–1729. doi: 10.1101/gr.210641.116
- Koffi, O., Samagaci, L., Goualie, B., and Niamke, S. (2017). Diversity of yeasts involved in cocoa fermentation of six major cocoa-producing regions in Ivory Coast. *Eur. Sci. J.* 13, 496–516. doi: 10.19044/esj.2017.v13n30p496
- Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923
- Legendre, P., and Gallagher, E. D. (2001). Ecologically meaningful transformations for ordination of species data. *Oecologia* 129, 271–280. doi: 10.1007/s004420100716
- Lima, C. O. C., Vaz, A. B. M., Castro, G. M., Lobo, F., Solar, R., Rodrigues, C., et al. (2020). Integrating microbial metagenomics and physicochemical parameters and a new perspective on starter culture for fine cocoa fermentation. *Food Microbiol.* 93:103608. doi: 10.1016/j.fm.2020.103608
- Linder, T. (2019). Cyanase-independent utilization of cyanate as a nitrogen source in ascomycete yeasts. *World J. Microbiol. Biotechnol.* 35:3. doi: 10.1007/s11274-018-2579-4
- Lu, J., Zhang, X., Zhang, X., Wang, L., Zhao, R., Liu, X. Y., et al. (2022). Nanopore sequencing of full rRNA operon improves resolution in Mycobiome analysis and reveals high diversity in both human gut and environments. *Mol. Ecol.* doi: 10.1111/mec.16534

- Magoc, T., and Salzberg, S. L. (2011). FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* 27, 2957–2963. doi: 10.1093/bioinformatics/btr507
- Manav, M. C., Sofos, N., Hove-Jensen, B., and Brodersen, D. E. (2018). The Abc of phosphonate breakdown: a mechanism for bacterial survival. *BioEssays* 40:e1800091. doi: 10.1002/bies.201800091
- Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* 17, 1–10. doi: 10.14806/ej.17.1.200
- Matias Rodrigues, J. F., Schmidt, T. S. B., Tackmann, J., and Von Mering, C. (2017). MAPseq: highly efficient k-mer search with confidence estimates, for rRNA sequence analysis. *Bioinformatics* 33, 3808–3810. doi: 10.1093/bioinformatics/btx517
- Matsuo, Y., Komiya, S., Yasumizu, Y., Yasuoka, Y., Mizushima, K., Takagi, T., et al. (2021). Full-length 16S rRNA gene amplicon analysis of human gut microbiota using MinION™ Nanopore sequencing confers species-level resolution. *BMC Microbiol.* 21:35. doi: 10.1186/s12866-021-02094-5
- Mayorga-Gross, A. L., Quirós-Guerrero, L. M., Fourny, G., and Vaillant, F. (2016). An untargeted metabolomic assessment of cocoa beans during fermentation. *Food Res. Int.* 89, 901–909. doi: 10.1016/j.foodres.2016.04.017
- Menzel, P., Lee Ng, K., and Krogh, A. (2016). Fast and sensitive taxonomic classification for metagenomics with Kaiju. *Nat. Communications* 7:11257. doi: 10.1038/ncomms11257
- Moreira, I. M. V., Vilela, L. F., Miguel, M. G. C. P., Santos, C., Lima, N., and Schwan, R. F. (2017). Impact of a microbial cocktail used as a starter culture on cocoa fermentation and chocolate flavor. *Molecules* 22:766. doi: 10.3390/molecules22050766
- Mota-Gutierrez, J., Botta, C., Ferrocino, I., Giordano, M., Bertolino, M., Dolci, P., et al. (2018). Dynamics and biodiversity of bacterial and yeast communities during fermentation of cocoa beans. *Appl. Environ. Microbiol.* 84:18. doi: 10.1128/AEM.01164-18
- Nakano, T., Suzuki, K., Fujimura, T., and Shinshi, H. (2006). Genome-wide analysis of the ERF gene family in Arabidopsis and Rice. *Plant Physiol.* 140, 411–432. doi: 10.1104/pp.105.073783
- Nicholls, S. M., Quick, J. C., Tang, S., and Loman, N. J. (2019). Ultra-deep, long-read nanopore sequencing of mock microbial community standards. *Gigascience* 8:giz043. doi: 10.1093/gigascience/giz043
- Nurk, S., Meleshko, D., Korobeynikov, A., and Pevzner, P. A. (2017). metaSPAdes: a new versatile metagenomic assembler. *Genome Res.* 27, 824–834. doi: 10.1101/gr.213959.116
- Papalexandratou, Z., Camu, N., Falovy, G., and De Vuyst, L. (2011c). Comparison of the bacterial species diversity of spontaneous cocoa bean fermentations carried out at selected farms in Ivory Coast and Brazil. *Food Microbiol.* 28, 964–973. doi: 10.1016/j.fm.2011.01.010
- Papalexandratou, Z., Kaasik, K., Kauffmann, L. V., Skorstengaard, A., Bouillon, G., Espensen, J. L., et al. (2019). Linking cocoa varieties and microbial diversity of Nicaraguan fine cocoa bean fermentations and their impact on final cocoa quality appreciation. *Int. J. Food Microbiol.* 304, 106–118. doi: 10.1016/j.ijfoodmicro.2019.05.012
- Patro, R., Duggal, G., Love, M. I., Irizarry, R. A., and Kingsford, C., (2017). Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* 14, 417–419. doi: 10.1038/nmeth.4197
- Pearman, W. S., Freed, N. E., and Silander, O. K. (2020). Testing the advantages and disadvantages of short- and long-read eukaryotic metagenomics using simulated reads. *BMC bioinformatics* 21:220. doi: 10.1186/s12859-020-3528-4
- Pereira, G. V. M., Alvarez, J. P., Neto, D. P. C., Socol, V. T., Tanobe, V. O. A., Rogez, H., et al. (2017). Great intraspecies diversity of *Pichia kudriavzevii* in cocoa fermentation highlights the importance of yeast strain selection for flavor modulation of cocoa beans. *LWT - Food Sci. Technol. (Lebensmittel-Wissenschaft - Technol.)* 84, 290–297. doi: 10.1016/j.lwt.2017.05.073
- Pereira, G. V. M., Magalhaes, K. T., Almeida, E. G., Coelho, I. S., and Schwan, R. F. (2013). Spontaneous cocoa bean fermentation carried out in a novel-design stainless steel tank: influence on the dynamics of microbial populations and physico-chemical properties. *Int. J. Food Microbiol.* 161, 121–133. doi: 10.1016/j.ijfoodmicro.2012.11.018
- Pereira, G. V. M., Miguel, M. G. C. P., Ramos, C. L., and Schwan, R. F. (2012). Microbiological and physicochemical characterization of small-scale cocoa fermentations and screening of yeast and bacterial strains to develop a defined starter culture. *Appl. Environ. Microbiol.* 78, 5395–5405. doi: 10.1128/AEM.01144-12
- Pothakos, V., De Vuyst, L., Zhang, S. J., De Bruyn, F., Verce, M., Torres, J., et al. (2020). Temporal shotgun metagenomics of an Ecuadorian coffee fermentation process highlights the predominance of lactic acid bacteria. *Current Research in Biotechnology* 2, 1–15. doi: 10.1016/j.crbiot.2020.02.001
- Qiu, X., Zhang, Y., and Hong, H. (2021). Classification of acetic acid bacteria and their acid resistant mechanism. *AMB Express* 11:29. doi: 10.1186/s13568-021-01189-6
- R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Ramos, C. L., Dias, D. R., Miguel, M. G. C. P., and Schwan, R. F. (2014). Impact of different cocoa hybrids (*Theobroma cacao* L.) and *S. cerevisiae* UFLA CA11 inoculation on microbial communities and volatile compounds of cocoa fermentation. *Food Res. Int.* 64, 908–918. doi: 10.1016/j.foodres.2014.08.033
- Reina-Bueno, M., Argandon, M., Salvador, M., Rodríguez-Moya, J., Iglesias-Guerra, F., Csonka, L. N., et al. (2012). Role of Trehalose in salinity and temperature tolerance in the model halophilic bacterium *Chromohalobacter salexigens*. *PLoS One* 7:e33587. doi: 10.1371/journal.pone.0033587
- Santana-Molina, C., Rivas-Marin, E., Rojas, A. M., and Devos, D. P. (2020). Origin and evolution of polycyclic triterpene synthesis. *Mol. Biol. Evol.* 37, 1925–1941. doi: 10.1093/molbev/msaa054
- Schwan, R. F., and Wheals, A. E. (2004). The microbiology of cocoa fermentation and its role in chocolate quality. *Crit. Rev. Food Sci. Nutr.* 44, 205–221. doi: 10.1080/10408690490464104
- Schwenninger, S. M., Leischtfeld, S. F., and Gantenbein-demarchi, C. (2016). High-throughput identification of the microbial biodiversity of cocoa bean fermentation by MALDI-TOF MS. *Lett. Appl. Microbiol.* 63, 347–355. doi: 10.1111/lam.12621
- Servent, A., Boulanger, R., Davrieux, F., Pinot, M. N., Tardan, E., Forestier-Chiron, N., et al. (2018). Assessment of cocoa (*Theobroma cacao* L.) butter content and composition throughout fermentations. *Food Res. Int.* 107, 675–682. doi: 10.1016/j.foodres.2018.02.070
- Spinnler, H. E. (2012). “Flavours from amino acids,” in *Food Flavors. Chapter 4*. ed. H. Jelen, 16.
- Stanke, M., Diekhans, M., Baertsch, R., and Haussler, D. (2008). Using native and syntactically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics* 24, 637–644. doi: 10.1093/bioinformatics/btn013
- Tomé, L. M., da Silva, F. F., Fonseca, P. L., Mendes-Pereira, T., Azevedo, V. A., Brenig, B., et al. (2022). Hybrid assembly improves genome quality and completeness of *Trametes villosa* CCMB561 and reveals a huge potential for lignocellulose breakdown. *Journal of Fungi*. 8:142. doi: 10.3390/jof8020142
- Verce, M., Schoonejans, J., Aguirre, C. H., Molina-Bravo, R., De Vuyst, L., and Weckx, S. (2021). A combined metagenomics and Metatranscriptomics approach to unravel Costa Rican cocoa box fermentation processes reveals yet unreported microbial species and functionalities. *Frontiers. Microbiology* 12:641185. doi: 10.3389/fmicb.2021.641185
- Visintin, S., Alessandria, V., Valente, A., Dolci, P., and Cocolin, L. (2016). Molecular identification and physiological characterization of yeasts, lactic acid bacteria and acetic acid bacteria isolated from heap and box cocoa bean fermentations in West Africa. *Int. J. Food Microbiol.* 216, 69–78. doi: 10.1016/j.ijfoodmicro.2015.09.004
- Wommack, K. E., Bhavsar, J., and Ravel, J. (2008). Metagenomics: read length matters. *Appl. Environ. Microbiol.* 74, 1453–1463. doi: 10.1128/AEM.02181-07