Environmental sources along natural cave ripening shape the microbiome and metabolome of artisanal blue-veined cheeses

Elena Alexa
Universidad de León
José F Cobo-Diaz
Universidad de León
Erica Renes
Universidad de León
Tom F O´Callaghan
University College Cork
Kieran Kilcawley
Teagasc - The Irish Agriculture and Food Development Authority
David Mannion
Teagasc
Iwona Skibinska
Teagasc
Lorena Ruiz
CSIC
Abelardo Margolles
CSIC
Paula Fernández-Gómez
Universidad de León https://orcid.org/0000-0001-9647-5251
Adrián Alvarez-Molina
Universidad de León
Fiona Crispie
Teagasc - The Irish Agriculture and Food Development Authority
Paula Puente-Gómez
Universidad de León
Mercedes López
Universidad de León
Miguel Prieto
Universidad de León
Paul Cotter

Teagasc - The Irish Agriculture and Food Development Authority https://orcid.org/0000-0002-5465-9068

Avelino Alvarez-Ordóñez (aalvo@unileon.es)

aalvo@unileon.es

Article

Keywords: cheese, microbiome, metabolome, processing environments, source tracking

Posted Date: December 21st, 2021

DOI: https://doi.org/10.21203/rs.3.rs-1150288/v1

Abstract

Microorganisms colonising processing environments can significantly impact food quality and safety. Here we describe a detailed longitudinal study assessing the impact of cave ripening on the microbial succession and quality markers across different producers of blue-veined cheese. Both the producer and cave in which cheeses were ripened significantly influenced the cheese microbiome and metabolome. The cheese microbiome was significantly determined by the microbiome of caves, which were a source of Brevibacterium, Corynebacterium, Staphylococcus, Tetragenococcus and Yaniella, among others, as demonstrated through source tracking and the characterization of 613 metagenome assembled genomes. Tetragenococcus koreensis and T. halophilus were detected at high abundance in cheese for the first time, associated with the occurrence of various metabolites, and showed high levels of horizontal gene transfer with other members of the cheese microbiome. Overall, we demonstrated that processing environments can be a source of non-starter microorganisms of relevance to ripening of artisanal fermented foods.

Introduction

Artisanal cheeses, including those which are recognised in the EU with Protected Designation of Origin (PDO), are highly valuable premium dairy products due to their unique sensory attributes relative to mass-produced cheeses. Their quality, shelf-life and safety are dependent on a number of factors, but are especially influenced by the diverse microbial communities they harbor, which vary from the core to the rind of the cheese and evolve throughout the cheese-making and ripening process[1,2]. Moreover, they can reflect differences among producers or regions in a manner that can be applied to verify the authenticity of PDO products[3].

Despite the substantial efforts devoted to study microbial succession patterns during the production and ripening of a wide variety of cheeses, certain aspects relating to microbial interactions and their functional implications remain largely unknown[4]. Historically, the microbiome of artisanal cheeses has been characterized using culture-based methods, which have several limitations[3,4]. High throughput sequencing methodologies (e.g., amplicon sequencing or shotgun whole metagenome sequencing) have transformed the way we study microbial ecology in complex ecosystems, such as those of artisanal cheeses[3,5–8], as they provide access to difficult-to-culture or viable but not culturable microorganisms[1,6], and facilitate the mechanistic understanding of community assembly in such a way that allows studying temporal dynamics of microbial communities at an unprecedented level of detail[6]. Nevertheless, there is also a general consensus that metagenomic studies need to be complemented to, for example, investigate the microbial and chemo-sensory relationships from a multi-omics perspective[9].

Some previous studies have highlighted that cheese and other fermented foods can serve as useful models for elucidating the determinants of microbial succession and interactions in complex communities, including understanding the molecular processes and the key metabolically active

microbes that contribute to the development of unique sensorial characteristics[6,7,10]. In the present work the microbiome/metabolome interplay in *Cabrales* cheese, a PDO artisanal Spanish blue-veined cheese, was studied. *Cabrales* cheese is produced in the mountainous area of "*Picos de Europa*" (*Principado de Asturias*, Northern Spain) from raw cow, sheep and/or goat milk. This cheese is ripened for up to 5 months in natural caves above 800 meters altitude, with > 90% humidity and a temperature ranging from 6 to 10ºC, where it develops its unique intense flavor, creaminess texture, and blue-veins from the rind to the core[11]. It is our hypothesis that the unique microenvironments in these natural caves promote the transfer from the environment to the cheese of specific bacteria, yeasts and molds that impart distinctive quality attributes to the ripened blue-veined cheeses. To test this hypothesis, we undertook a large-scale longitudinal study investigating the structure and functional potential of microbial communities of cheeses (cores and rinds) during the ripening process in natural caves by strategically combining classic microbiological methods, whole-metagenome shotgun sequencing and metabolomic analyses. Through this approach, we provide a cutting edge insight into the relationships between traditional cheese-making environments, microorganisms prevailing in cheese and the occurrence of metabolites affecting cheese quality and safety (*i.e.*, volatile compounds and biogenic amines).

Results

Shared characteristics and patterns of succession

The experimental setup included cheeses from three different producers who used the same cave for ripening. In addition, for one producer, cheeses derived from a single production batch were ripened in three different caves, in order to identify microbiome signatures associated with the respective cave microenvironments (Fig. S1).

Regardless of producer or cave used for ripening, cheeses maintained certain shared characteristics and microbial succession patterns. Before cheeses entered the ripening caves (Stage1; 30 days post-manufacture) they showed a mild acidic pH (5.41-6.58), and some degree of drying (water activity of 0.828-0.924). *Cabrales* cheeses are routinely rind-washed throughout ripening in the caves, where the relative humidity is >90%. Consequently, significant (p<0.05) increases in cheese water activity, relative humidity and pH, and decreases in dry matter and salt content occurred thereafter, marked mostly on cheese rinds at intermediate (Stage2) and final (Stage3) stages of ripening (Fig. S2A). Total fat and protein contents remained fairly stable through ripening (Fig. S2A).

Forty-two volatile compounds, belonging to eight chemical classes (Table S1), and seven biogenic amines were detected. While the profile of volatile compounds and biogenic amines was highly variable depending on the producer and ripening cave (as described in detail below), a global view shows that the total content in volatile compounds increased on cheese rinds as ripening proceeded while no significant changes were observed for the cheese cores (Fig. S2B). The biogenic amines content significantly increased with ripening time for both core and rind samples (Fig. S2B).

Regarding the microbiome composition, counts of total aerobic bacteria, lactic acid bacteria, *Enterobacteriaceae*, and yeasts and molds reached maximum levels at Stage1 and then gradually declined during ripening, both for cheese cores and rinds (Fig S3A). At Stage1, cheeses were dominated by lactic acid bacteria, mainly *Lactococcus* and members of the former *Lactobacillus* genus, and by the yeast *Debaryomyces* and the fungi *Penicillium* and *Geotrichum* (Fig. S3B). Whereas fungal populations remained fairly stable along ripening, a significant decrease was evident in the relative abundance of *Lactococcus* and *Lactobacillus*, among others (Fig. S3B), while other taxa, such as *Tetragenococcus*, *Brevibacterium* and *Corynebacterium* emerged with high relative abundances at Stage2 and Stage3, especially at rind level (Fig. S3B).

Metabolome and microbiome differences by producer

A principal component analysis, performed to reflect all of the volatile compounds and biogenic amines, identified that the producer variable had a significant influence on the metabolomic profile of cheeses, both for the core and rind (adonis, p=0.001), with the first two components explaining 80.40 % and 74.80 % of the total variation, respectively (Fig. 1A,B). Focusing on the eight most abundant volatiles, some trends were apparent; higher levels of 2-heptanone, 2-nonanone, 2-heptanol and 2 -pentanone were found in cheeses from Producer3; higher concentrations of ethyl hexanoate and ethyl butanoate were found in Producer2 cheeses; and the abundance of hexanoic and butanoinc acids were lower for Producer3 cheeses (Fig. 1C, Table S2). Regarding amines, tryptamine and tyramine were more abundant on Producer2 and Producer3 cores and cadaverine on Producer1 rinds (Fig. 1D, TableS2).

The cheese producer also had a significant influence on the bacterial and fungal taxonomic profile of samples (Fig 2A,B, Fig S4A,B), both for cheese cores (adonis, p=0.001 for both bacteria and fungi) and rinds (adonis, p=0.001 for bacteria and p=0.018 for fungi). At cheese core level, several bacterial genera within the main ones such as *Lactobacillus*, *Corynebacterium*, and *Enterococcus*, among others, were associated with Producer2, while *Lactobacillus* and *Tetragenococcus* were predominantly linked to Producer3 (Fig. 2A,C, Table S2). At cheese rind level, several bacterial genera were clearly associated with Producer2 (*e.g.*, *Corynebacterium* and *Tetragenococcus*) while *Lactococcus* was associated with Producer1 and Producer3 (Fig. 2B,C). Regarding fungi, core samples were highly abundant in *Penicillium*, but, for Producer2, similar relative abundances of *Geotrichum* and *Penicillium* were observed. At rind level *Penicillium* and *Debaryomyces* showed high abundances in samples from Producer1 and Producer2, and *Debaryomyces* was the dominant genus for Producer3 (Fig. 2D, Table S2).

Metabolome and microbiome differences by ripening cave

For Producer2, the ripening cave also had a significant influence on the metabolomic profile of cheese cores and rinds (adonis, p=0.001), with the first two components explaining 71.3 % and 73.2 % of the total variation, respectively (Fig. 3A,B). Focusing on the eight most abundant volatiles, Cave3 cheeses differed most due to lower levels of ethyl hexanoate and ethyl octanoate, and higher levels of 2-heptanone on cores and rinds and hexanoic acid, butanoic acid and octanoic acid on rinds (Fig. 3C, Table

S3). Regarding amines, tyramine was more abundant in Cave2 cores, tryptamine in Cave2 cores and Cave3 rinds, and cadaverine in Cave1 cores and rinds (Fig. 3D, TableS3).

The cave also had a significant influence, although lower than that found for the variable producer, on the taxonomic profile of cheese cores (adonis, p=0.007 for bacteria and p=0.049 for fungi) and rinds (adonis, p=0.027 for bacteria and 0.653 for fungi) (Fig. 4A,B, Fig. S4C,D). The most relevant differences among caves in the abundance of some particular taxa were those observed for *Lactococcus* and *Lactobacillus*, which were more abundant on core samples from Cave1 and Cave3, respectively, at Stage3; *Tetragenococcus*, *Corynebacterium* and *Staphylococcus*, which were more abundant on rind samples from Cave1, Cave2 and Cave3, respectively (Fig. 4A,C, Table S3); and *Debaryomyces*, which was significantly more abundant on cheese rinds from Cave1 than on those from the other two caves (Fig. 4D, Table S3).

The cave microbiome strongly shapes the rind cheese microbiome

The characterisation of the microbial communities prevailing in some primary sources that could determine the cheese microbiome (i.e., milk, curd, different factory processing environments and cave environments) showed that milk was dominated by *Pseudomonas*, followed by *Acinetobacter*, *Staphylococcus* or *Lactococcus*, at different levels depending on the producer (Fig. S5). Similar profiles but with a higher abundance of some lactic acid bacteria, such as *Lactococcus*, were found in curd samples (Fig. S5). *Brevibacterium*, *Psychrobacter*, *Pseudomonas, Acinetobacter, Penicillium*, *Debaryomyces* and *Kluyveromyces* were the dominant genera on food processing environments (both food contact and non food contact) of the three cheese producing plants (Fig. S5). *Brevibacterium, Corynebacterium* and *Debaryomyces* were the dominant genera in cave food contact surfaces, while *Penicillium* and a wide range of non-dominant bacterial and fungal genera prevailed on the caves´ non food contact surfaces (Fig. S5).

A source tracking analysis revealed that the bacterial composition of cheese cores, and of cheese rinds at Stage1, was mainly determined, for Producer1 and Producer2, by the curd microbiota, and for Producer3 by the microbiota of food contact environments from the processing plant (Fig. 5). The main bacterial genera traced back to the curd and dairy food contact surfaces were *Lactococcus*, *Lactobacillus* and, in the case of Producer2, *Hafnia* (Fig. 5). Over subsequent stages of ripening, the curd microbiota still represented an important source for the cheese core microbiome, although the cave environment had an important role in shaping the microbiome of cheese cores on Cave2 at Stage2 and Stage3, mainly being a source of *Tetragenococcus* and *Corynebacterium* for Producer1 and Producer3, and *Corynebacterium* for Producer2 (Fig. 5). Other less abundant genera (*e.g., Brachybacterium*, *Alkalibacterium*, *Staphylococcus*) were also traced back to the cave environments. Regarding cheese rinds, their bacterial microbiota at Stage2 and Stage3 were traced back mainly to the cave environments, which were revealed as a likely source of *Brevibacterium*, *Corynebacterium* and *Tetragenococcus* (on Stage3) and other minority genera (on Stage2), for Producer1; *Tetragenococcus* in Cave1 and Cave3, together with

Corynebacterium in Cave2, for Producer2; and *Tetragenococcus*, *Brevibacterium* and other minority genera (on Stage2) and *Tetragenococcus* (on Stage3) for Producer3 (Fig. 5).

A source tracking analysis undertaken on the fungal communities showed that food contact surfaces from Cave2 were a source of *Debaryomyces* in rind samples from Producer1 and Producer3, and milk was a source of *Geotrichum* in core samples from Producer2 (Fig. S6).

MAGs analysis confirmed the relevance of cave environments as a source of bacteria

A total of 110 high-quality metagenome assembled genomes (MAGs), 227 medium quality MAGs with completeness > 90%, and 276 medium-quality MAGs with completeness of 50-90% were obtained. They were classified into 70 genera, with the main genera represented being *Lactococcus* (103 MAGs), *Corynebacterium* (74 MAGs), *Lactobacillus* (68 MAGs), *Tetragenococcus* (54 MAGs), *Bifidobacterium* (30 MAGs), *Staphylococcus* (28 MAGs), *Brevibacterium* (28 MAGs) and *Yaniella* (22 MAGs). Cheese samples were the ones with the highest number of MAGs (413 out of 613 MAGs), while raw materials, processing environments at factory level and cave environments yielded 43, 74 and 83 MAGs, respectively (Fig. 6A, Fig. S7).

MAGs from several putative new species were obtained on samples from cave surfaces or cheeses. These included 14 *Brevibacterium* sp. MAGs most closely related to *B. sandarakinum,* 8 *Corynebacterium* sp. MAGs related to *C. nuruki*, 6 *Corynebacterium* sp. MAGs related to *C. glyciniphilum*, 2 *Psychrobacter* sp. MAGs and one *Tetragenococcus* sp. MAG. related to *T. koreensis*. Remarkably, despite the low relative abundance of *Yaniella* found at read level, a total of 22 MAGs from this genus, closely related to the recently discovered Candidatus *Yaniella excrementavium*, were obtained (Fig. 6A, Fig. S7). Furthermore, 3 MAGs assigned to the former genus *Lactobacillus*, obtained from cheese core samples from Producer3 at Stage1, were not assigned at species level, with *L. selangorensis* and *L. camelliae* being the most closely related species (Fig. 6A).

It is worth noting that MAGs assigned to *Lactococcus*, a dominant genus at Stage1, presented phylogenetic differences by producer, with *Lactococcus lactis* MAGs from Producer2 clustering separately from MAGs from other producers, *L. piscium/carnosum* MAGs being exclusively found for Producer2 and *L. raffinolactis* and *L. garvieae* for Producer3 (Fig. 6A). On the other hand, different closely related MAGs from *Corynebacterium casei*, *Staphylococcus equorum*, *Brevibacterium* sp., *Tetragenococcus halophilus*, *T. koreensis* and *Yaniella* sp. were obtained from both cheese and cave environmental samples. Moreover, in the case of cheese MAGs from *Brevibacterium*, *Tetragenococcus* and *Yaniella,* they were only obtained from cheeses at Stage2 and Stage3 (Fig. 6A, Fig. S7). Similar results were observed at read level, with the absence of these genera at Stage1, while they were within the main ones found at Stage2 and Stage3 (Fig. S8). It was also interesting to see that two different clusters of *S. equorum*, with differentiated functional potential, were detected colonizing food processing environments within factories and cheeses at Stage1, and cave environments and cheeses at Stage2 and Stage3, respectively (Fig. 6A, Fig. S9).

The analysis of the functional potential of MAGs evidenced that *Lactococcus*, *Lactobacillus*, *Staphylococcus* and *Tetragenococcus* had a relatively similar functional profile, which was characterised by the high abundance of functions related to the metabolism of carbohydrates, while *Corynebacterium*, *Brevibacterium* and *Yaniella* MAGs had a very different functional profile, characterised by the high abundance of different functions mainly related to protein and fatty acids metabolism (Fig. 6B).

The cheese microbiome is associated with cheese quality markers

Correlation analyses undertaken with respect to cheese samples from stages 2 and 3 of ripening revealed that, among the species for which MAGs were generated, at core level, *Lactococcus* (*L. lactis*, *L. piscium*) *Lactobacillus* (mainly *L. paracasei*) and *Bifidobacterium* (*B. crudilactis, B. mongoliense*) were those that had the strongest positive correlations with the abundance of various volatile compounds and biogenic amines. The most remarkable associations found were for *L. lactis* with pentanoic and hexanoic acids; *L. piscium* with 3-methyl-butanal, propanoic acid, ethyl-butanoate, propyl-butanoate, trimethylpyrazine, ethyl-octanoate, putrescine and cadaverine; *L. paracasei* with 3-methyl-butanal, propyl-butanoate, trimethylpyrazine and isobutyl-decanoate; *B. crudilactis* with 3-methyl-butanal and trimethylpyrazine; and *B. mongoliense* with trimethylpyrazine, isoamyll-hexanoate, isopentyl-octanoate, isobutyl-decanoate and tryptamine (Fig. 7A,B). At rind level, these species also showed positive correlations with the abundance of certain volatiles and amines, but other representative taxa, traced back to the cave environment, such as *Tetragenococcus* (mainly *T. halophilus*), *Staphylococcus equorum*, and *Corynebacterium* (*C. flavescens*, *C. variabile*) were also strongly associated with some cheese volatile and amine markers. The most remarkable associations were found for *T. halophilus* with propyl-butanoate, 2-heptanol, trimethylpyrazine, 2-pentylbutyrate, 2-nonanone, 2-nonanol, ethyl-octanoate, 2-undecanone, ethyl-decanoate, isobutyl-decanoate, ethyl dodecanoate, putrescine and tyramine; *S. equorum* with 2-heptanone, 2-heptanol, 2-octanone, 2-nonanol, γ-hexalactone, 2-undecanone, isopenthyl-octanoate, isoamyl-hexanoate and putrescine; and *C. variabile* and *C. flavescens* with pentanoic and hexanoic acids (Fig.7B). Interestingly, both at cheese core and cheese rind level, with the exception of *Brevibacterium* and the former *Lactobacillus* genus, species from the same genera showed highly different microbiome-volatilome correlation patterns while species belonging to the same genera, with the exception of *Corynebacterium* and *Tetragenococcus*, had similar correlations with amine concentrations (Fig. 7).

The cheese microbiome is rich in Horizontal Gene Transfer events

In order to identify among the main bacterial taxa cues of microbiome acclimatisation to cave and cheese microenvironments, signs of horizontal gene transfer (HGT) events were seeked in the available MAGs. A total of 23,001 HGT events, containing 67,411 coding regions, were detected between 56 taxa above species level, with *Lactococcus*, *Tetragenococcus* and *Staphylococcus* being the main genera involved (59.3, 52.0 and 23.3 % of total HGT events, respectively), and *Lactococcus-Tetragenococcus*, *Staphylococcus-Tetragenococcus* and *Lactococcus-Staphylococcus* as the main HGT pairs (29.6, 10.9

and 9.4 % of total HGT events, respectively) (Fig. 8A). HGT events were identified from MAGs belonging to all surfaces sampled, but cheese core and rind samples at stages 2 and 3 of ripening contributed the most (56.6 % of HGT events) (Fig. 8B).

Up to 35.3 % of the HGT events were associated with plasmid sequences, according to plasflow analysis, while only 0.5 % carried relaxase encoding genes, associated with plasmid mobilization. Moreover, 11.7, 8.1 and 0.9 % harboured prophages, transposases and integrases, respectively. Although 55.5 % of the coding regions could not be assigned to ko codes of the KEGG Orthology database, 10.6, 7.2 and 5.4 % were assigned at level2 to the groups *Protein families: signaling and cellular processes*, *Carbohydrate metabolism* and *Protein families: genetic information processing*, respectively. At level3 of KO classification, *Prokaryotic defense system*, *Galactose metabolism*, *Glycolysis/Gluconeogenesis*, *Purine metabolism* and *Fructose and mannose metabolism* were within the dominant functions (Fig. 8C). Finally, 68 HGT-associated contigs were longer than 20kb, containing from 12 to 45 coding regions (CDS) (Fig. 8D), and clustered into 26 groups. *Lactococcus-Tetragenococcus* and *Alkalibacterium-Tetragenococcus* were the dominant pairs within these longest HGT events, with 14 and 4 clusters, respectively (Table S4). Half of the obtained HGT-clusters contained genes related to mobile genetic elements such as plasmids, transposases, phages and integrons; 3 clusters harboured genes related to β-lactams resistance, and up to 9 clusters contained genes related to protein, carbohydrate or lipid metabolism (Table S4).

Discussion

Our study found that two different groups of microorganisms prevailed in *Cabrales* PDO cheese. The first group comprises various taxa (e.g., *Lactococcus*, *Lactobacillus*, *Bifidobacterium*, *Penicillium*, *Geotrichum, Debaryomyces*) that show a high abundance in the cheeses from the initial stages of production, prior to the cheeses entering the ripening caves. These microorganisms very likely originate from the raw materials, the starter cultures used and/or the dairy plant processing environments. This group of microbes, with the exception of *Debaryomyces*, occurred at high abundances in cheese cores over the whole ripening period and are associated with the occurrence of various volatile compounds in the cheese core, highlighting in some cases intraspecific phylogenetic diversity depending on the cheese producer. However, as ripening progressed within the cave this group of microbes was gradually displaced, especially on cheese rinds, by the second group of microbes, which comprises several taxa (e.g., *Tetragenococcus*, *Corynebacterium*, *Brevibacterium*, *Yaniella*, *S. equorum*) not found (or present at very low abundance) in milk, curd, or environments within the processing plant and cheeses at the initial stages of ripening. These microbes are associated with the occurrence of some volatile compounds in rinds, linked to unique aroma notes, as discussed below in greater detail, and are very likely sourced within the cave environments, as evidenced by SourceTracker predictions and the close genomic relationships among some strains, such as some representatives of *S. equorum*, *T. korrensis, T. halophilus and C. casei*, as revealed by analysis of MAGs from cheese and cave samples.

The role of some members of the *Cabrales* cheese microbiome in the ripening process is clear based on previous studies for these sort of cheeses. Indeed, *Lactococcus*, *Lactobacillus*, *Staphylococcus*, *Corynebacterium*, *Brevibacterium, Penicillium*, *Geotrichum* or *Debaryomyces* have all previously been detected at high abundance in blue-veined rind washed cheeses, and in some cases also in cheese processing environments[5,6,12], with very clear contributions to cheese fermentations. Moreover, some of these taxa, such as *Corynebacterium* or *Brevibacterium*, play a particularly important role in aroma development during cheese ripening[5,12,13]. In our study, new strong microbiome-volatilome associations were uncovered for these taxa. Indeed, *Lactococcus* showed strong positive correlations with the abundance of carboxylic acids, which impart cheese, rancid, pungent, sharp, sweaty, malty/chocolate, fruity or buttery flavour notes, depending on strain to strain variability[14]; and *L. piscium, B. crudilactis* and *L. paracasei* strongly correlated with the occurrence of propyl butanoate and ethyl octanoate, responsible for sweet, fruity, and floral notes[15]; and of 3-methyl butanal, responsible for the generation of malty, cheesy, dark-chocolate notes[15].

On the other hand, other taxa of the *Cabrales* cheese microbiome have received much less attention in the past or have been detected at high relative abundance in cheese in our study for the first time. Specifically, this is the case of one low abundant species of *Yaniella* and two species from *Tetragenococcus* (*T. koreensis* and *T. halophilus*), found in very high abundance in the latest stages of *Cabrales* cheese ripening. Although the presence of *Yaniella* in food related environments has been recently described[2,16], its relevance during cheese ripening remains unclear. Here we describe a total of 22 *Yaniella* sp. MAGs (8 of them from cave environments), closely related to the recently discovered *Candidatus Yaniella excrementavium*, for which further research is needed to understand its contribution to the quality and safety of *Cabrales* cheese. *Tetragenococcus* has been only found, at low relative abundances, in Brie cheese[17] and two Mexican cheeses[18]. The two *Tetragenococcus* species here detected have a functional profile similar to that of some traditional lactic acid bacteria (e.g., *Lactococcus*, *Lactobacillus*) and possibly displace them due to their better acclimatization to the microenvironments associated with our study. *Tetragenococcus* are halo-alkaliphilic bacteria and the relatively high salt content and progressive increase in pH along ripening, especially at rind level, can provide ideal conditions to promote their growth[19]. Remarkably, the dominance of one of the two *Tetragenococcus* species over the other depended on both the producer and ripening cave, and when both *Tetragenococcus* species co-occurred, *T. halophilus*, with a higher optimum pH for growth of 6.1-7.0[17], showed tropism for the cheese rind and *T. koreensis*, with an optimum pH for growth of 5.7-6.5[17], for the cheese core. Interestingly, these taxa were associated with the co-occurrence of various metabolites or quality markers in the cheese, which suggests that they significantly contribute to some of the quality attributes of this distinctive cheese. Specifically, *T. koreensis* was associated with a high content of carboxylic acids in core cheese samples and *T. halophilus* showed strong positive correlations with some ketones (2-nonanone, 2-undecanone), secondary alcohols (2-nonanol, 2-heptanol), and esters (propyl-butanoate, 2-pentylbutyrate, ethyl-octanoate, ethyl-decanoate, ethyl dodecanoate, isobutyl-decanoate, among others), which originate from β-oxidation of fatty acids and/or transamination and

esterification reactions and are responsible for strong musty and blue cheese notes[15]. Moreover, the identification of these taxa in cheeses from a very restricted geographical region suggests that they may be used as microbiome markers for PDO authentication purposes.

Finally, the results provide evidence of high levels of horizontal gene transfer among MAGs belonging to different bacterial genera, suggesting that these events can mediate strain adaptation to the cheese/cave microenvironments, as has been previously proposed for cheese rind microbiomes[20]. A relatively high proportion of such HGT-associated sequences harboured mobile genetic elements. Likewise, high levels of phage and transposase associated genes have been previously found on MAGs obtained from Swiss Gruyère cheese[21]. Remarkably, extensive HGT was observed for *Tetragenococcus* with *Lactococcus* and *Staphylococcus*, and the fact that a many associated genes were related to carbohydrate metabolism functions (e.g., galactose metabolism, glycolysis/gluconeogenesis, and fructose and mannose metabolism) indicates that these transfer events may have mediated the adaptation of *Tetragenococcus* to the cheese dairy environments.

Overall, our study highlights that cave environments represented an important source of non-starter microorganisms with a relevant role in the ripening of these artisanal blue-veined cheeses, and identifies among them novel taxa and taxa not previously regarded as being dominant components of the cheese microbiome (*Tetragenococcus* spp.), providing very valuable information to the authentication of this PDO artisanal cheese in the future.

Materials And Methods

Cheese making, sampling strategy and sample collection

Cabrales cheese is the most popular traditional blue-veined cheese in Spain and has a Protected Designation of Origin (PDO) status since 1981. This cheese is produced in “Picos de Europa" (a region in Northern Spain with altitudes above 800 metres) from raw cow, sheep or goat milk or a mixture of two or all three types of milk. The traditional cheese-making process of *Cabrales* cheese involves curdling mixtures of evening and morning milk at 28 - 30ºC using rennet exclusively of animal origin. Curds are cut to hazelnut grain size and placed in cylindrical moulds at room temperature, being turned upside down several times in order to drain them off without applying pressure. Then, dry salt is added to the cheese surface and cheeses are taken to a ripening chamber where they remain at room temperature for 15 - 30 days. After this time, cheeses are placed in natural caves, which are characterized for being deep, with a north-facing entry, and for having at least two openings with flowing water in order to create an airstream. These conditions lead to very high humidity (> 90 %), with temperatures ranging from 6°C to 10°C. Cheeses must remain in the caves for two to five months and during this time they are periodically turned and washed.

In each of the three cheese processing facilities the following samples were taken for analysis: 200 mL of milk and curd samples (three replicates) during the cheese-making process, swab samples from the

processing environments entering in contact with the cheese (e.g., work tables, cheese vat, cheese moulds), swab samples from non-food contact surfaces (e.g., drains, floors, walls), and three cheeses (2.0 kg each) taken 30 days after cheese making, just before leaving the factory to the ripening cave/s.

In the natural caves, the following samples were taken: swab environmental samples (food contact and non-food contact surfaces, including drains if available), and three cheeses (2.0 kg each cheese) for each of the producers at both the intermediate (90 to 120 days of ripening) and final (130 to 190 days of ripening) ripening stages.

Environmental samples were collected by swabbing using sterile sponges pre-moistened with 10 mL of neutralizing buffer, which were placed individually in sterile bags (3M, Minnesota, USA).

Appropriate personal protective equipment and gloves were used during all samplings to avoid cross-contamination. At each collection point, the samples were placed in a cooling box containing ice packs and transported to the laboratory within two to three hours. A detailed description of the samples taken is provided in Fig. S1.

Upon arrival at the lab, the cheese rinds were scraped with a sterile knife until the cheese paste was visible (≃ 0.7 cm). A different sterile knife was used to prepare cheese core samples in order to avoid any possible cross contamination from the rind.

Culture-based microbiological analyses

For microbiological analyses, 10 g of cheese (rind and core, separately), milk or curd were homogenized with 90 mL of Buffered Peptone Water (BPW; Merck, Germany) for 2 min in a Stomacher 400 Lab Blender (Seward Medical, London, UK), while 10 mL of BPW was directly added to the sterile bags containing swab environmental samples. Serial decimal dilutions were then prepared in BPW and spread in triplicate on the following media: (i) Plate Count Agar (PCA; Pronadisa, Spain) for mesophilic bacteria, incubated for 24±2 h at 37ºC; (ii) De Man Rogosa Sharpe agar (MRS; Merck) for lactic acid bacteria (LAB), incubated for 24±2 h at 30°C; (iii) Oxytetracycline Glucose Yeast Extract Agar (OGYEA; Pronadisa) for yeasts and molds, incubated for 5 days at 25°C; and (iv) Violet Red Bile Glucose Agar (VRBGA; Pronadisa) for *Enterobacteriaceae*, incubated for 24±2 h at 37ºC. The results were expressed as means and standard deviations of log CFU (colony-forming units) per gram from three independent replicates.

Proximate analysis

Milk samples were analyzed for their content in total solids, fat, protein, lactose and somatic cells by using Milkoscan FT2 (Foss Electric, Hillerod, Denmark).

pH and titratable acidity of milk, curd and blue-veined cheese samples were determined according to AOAC International (1990b) methods.

In the case of cheese rind and cheese core samples, total solids, fat, protein and salt contents were determined according to the International standards 004 (FIL-IDF 2004), 221 (FIL-IDF 2008), 20-1 (FIL-IDF 2001), and 935.43 (AOAC 1990a), respectively. Water activity (aw) was analyzed instrumentally using an Aqua Lab Dew Point Analyzer CX-2 (Decagon Devices, Pullman, WA, USA). All samples were analysed in duplicate.

Biogenic amines analysis

Biogenic amines in cheese samples were determined following the method described by Redruello et al.[22], as follows: one gram of each cheese sample was homogenized with 10 mL of 0.1 M HCl-0.2% 3,3´thiodipropionic acid (Merck, Madrid, Spain) and then centrifuged at 5000 *g* for 20 min. This mixture was kept in an ultrasonic bath Bransonic 221 (Branson Ultrasonics S.A, Danbury, USA) for 30 min and then centrifuged at 5000 *g* for 20 min. The resultant supernatant was deproteinized by passing it through ultrafiltration inserts (Amicon Biomax 5K; Merck) and centrifuging at 3500 *g* for 60 min. Then, 20 µL were derivatized with diethyl ethoxymethylenemalonate (DEEMM; Merck). L-2-aminoadipicacid (Merck) was used as an internal standard. The chromatography system consisted of an H-Class Acquity UPLCTM system (Waters, Massachusetts, USA) coupled to a photodiode array detector. The separation of biogenic amines was carried out using a Waters Acquity UPLCTM BEH C18 column (1.7 µm particle size, 100 mm × 2.1 mm I.D.) held at 35 ºC. The mobile phase was made up of 25 mM acetate buffer pH 6.7 plus 0.02 % sodium azide (eluent A), methanol (eluent B) and acetonitrile (eluent C). One µL sample was injected and the flow rate was set up at 0.45 mL/minute according to the linear gradient used by Redruello et al.[22]. The target compounds were identified by their retention times and their spectral characteristics at 280 nm, and were quantified using the internal standard method. Data were acquired and analyzed using the software Empower 2 (Waters). Each sample was analysed in duplicate.

Volatile compounds analysis

Volatile-compounds profiling of the blue-veined cheese samples was carried out using the headspace solid-phase microextraction gas chromatography-mass spectrometry method (HS-SPME-GC–MS) as described by Panthi et al.[23], with some modifications. Briefly, 0.5 g of each cheese sample was placed in a 20-mL amber screw-capped headspace vial with a silicone/PTFE liner (Apex Scientific Ltd., Co. Kildare, Ireland). Extraction of volatile compounds was performed using a single SPME fiber 50/30 µm Divinylbenzene/Carboxen/Polydimethylsiloxane (DVB/CAR/PDMS; Agilent Technologies Ireland Ltd., Cork, Ireland), which was exposed to the HS for 20 min at a depth of 21 mm at 40 °C under the condition of pulsed agitation, 5 s on and 2 s off at 350 rpm, using a Shimadzu AOC-5000 injection system (Mason Technology, Dublin, Ireland). The fiber was retracted, injected via a merlin microseal, in splitless mode, into the inlet of a Shimadzu 2010 Plus GC (Mason Technology) and desorbed for 2 min at 250°C. Separation of volatile compounds was carried out using a DB-624 UI column (1.80 µm particle size, 60 m x 0.32 mm I.D.; Agilent Technologies Ltd.). Helium was used as a carrier gas at a constant flow of 1.2mL/min. The temperature of the column oven was set as follows: the initial temperature was 40 °C, then increased at 5°C/min to 230 °C (15 min hold), and further increased at 15°C/minute to 260°C (5

minutes hold), yielding a total GC run time of 65 min. The detector was a Shimadzu TQ8030 MSD triple quadrupole mass spectrometer (Mason Technology) and it was used in single quadrupole mode. The ion source temperature was 220°C, the interface temperature was set at 260°C, and the MS mode was electronic ionization (70 V), with the mass range scanned between 35 and 250 amu. A set of external standards were run at the start and end of the sample set and abundances were compared to known amounts to ensure that both the SPME extraction and MS detection were performing within specifications. In addition, autotune of the GCMS was performed before the analysis to ensure optimal GC-MS performance. Post-analysis processing of the obtained total ion chromatograms was carried out using GCMSsolution (Shimadzu, Japan) and batch processing of samples was performed using R software (RStudio. Inc., Boston, MA). Volatile compounds were identified by: i) mass spectrum comparison to the National Institute of Standards and Technology (NIST) 2011 mass spectral library, ii) the automated mass spectral deconvolution and identification system (AMDIS), iii) an internal library created in GC–MS Solutions software (Shimadzu, Japan) with target and qualifier ions using linear retention indices as described by van Den Dool and Kratz[24], and vi) linear retention indexes matching against peer-reviewed publications, where possible, to confirm compound identification. Each cheese sample was analysed in triplicate.

Whole metagenome sequencing analyses

DNA extraction

Prior to DNA extraction, curd, cheese, and milk samples were prepared by homogenising 10 g or mL in 90 ml of BPW, and centrifuging at 6500 *g* for 8 min. Then, pellets were washed with 50 mL of BPW, followed by a new centrifugation at 6500 *g* for 8 min to harvest the associated microbiota. For the environmental samples, pools of five swabs from each sample category were mixed with 10 mL of BPW. After thorough homogenisation, the BPW solutions were centrifuged at 6500 *g* for 8 min. Pellets from all samples were kept at -80 ºC until further use. Total metagenomic DNA was extracted from the pellets by using the DNeasy PowerSoil PRO kit (Qiagen GmbH, Germany) following the manufacturer's instructions, but conducting a double elution step with 25 µL of 10 mM Tris-HCl, in order to improve DNA yields.

Library construction and shotgun sequencing

Extracted DNA was employed to prepare 150 bp paired-end sequencing libraries using the Illumina Nextera XTLibrary Preparation Kit (Illumina Inc., San Diego, CA). Sequencing was performed on the Illumina NextSeq 500 platform using a NextSeq 500/550 High Output Reagent kit v2 (300 cycles), in accordance with the standard Illumina sequencing protocols.

Reads quality filtering and annotation

Quality filtering of raw reads was performed with AfterQC v0.9.6[25] using default parameters. Reads matching to cow and sheep genomes were removed using a combination of bowtie2 v2.3.4.1[26], samtools v1.9[27] and bedtools v2.29.0 (https://bedtools.readthedocs.io/en/latest/#) according to a pipeline

previously published (http://www.metagenomics.wiki/tools/short-read/remove-host-sequences). The taxonomic assignment of filtered reads was done using Kraken2 v2.0.8-beta[28] with the maxikraken2-1903_140GB database (https://lomanlab.github.io/mockcommunity/mc_databases.html) for bacterial taxonomy, while a custom database with kraken2-microbial database plus 20 reference fungal genomes from species found often in cheese environments (Table S5) was employed for fungal taxonomy assignation, due to the absence of major fungal species on the maxikraken2-1903_140GB database.

Reads assembly and MAGs binning

Each sample was independently subjected to *de-novo* metagenomic assembly through metaSPAdes v3.13[29] using default parameters. Filtered reads were mapped against contigs higher than 1000 bp obtained from the same sample using bowtie2 v2.3.4.1[26] with the *--very-sensitive-local* parameter. The *jgi_summarize_bam_contig_depths* script, from MetaBAT2 v2.12.1[30], was used to calculate contigs depth values, mandatory for per-sample contig binning based on tetranucleotide frequency and the contig abundance protocol, which was performed using MetaBAT2 and the option '-m 1500' [30].

The quality of MAGs was estimated using the lineage-wf workflow of CheckM (v1.1.3)[31] and cmseq (https://github.com/SegataLab/cmseq), to classify MAGs as high quality MAGs (>90% completeness, <5% contamination, <0.5 % non-synonymous mutations) and medium quality MAGs (>50 completeness, <5% contamination) for further analysis.

MAGs analysis

MAGs were assigned to taxonomy groups using the CAT/BAT pipeline v.5.1.2[32] and CAT_prepare_20200618 database. Those MAGs assigned to the bacterial genera *Lactococcus*, *Lactobacillus*, *Tetragenococcus*, *Staphylococcus*, *Corynebacterium*, *Brevibacterium* and *Yaniella* were taxonomically assigned at species level using as reference the available genomes from these genera on the NCBI (ftp://ftp.ncbi.nlm.nih.gov/genomes/), which were selected and downloaded using the download_genomes pipeline (https://github.com/JoseCoboDiaz/download_genomes). The MAGs and NCBI genomes for each selected genus were employed for the calculation of Average Nucleotide Identity (ANI) values by using dRep v2.6.2[33]. Taxonomic assignment at species level was done by manual inspection of ANI phylogenetic trees obtained by dRep software. Moreover, the *Mdb.txt* output file obtained from all ANI calculations was transformed into a distance matrix, which was employed for the construction of a phylogenetic tree plot by *ggtree* R-package, using the UPGMA clustering algorithm (*hclust* method="average").

The functional annotation of MAGs with completeness higher than 90% was performed by using eggNOG-mapper v2.1.5[34] and the MMseqs algorithm[35]. The "KEEG_ko" column from the eggNOG-mapper output was employed to build the count matrix of KOs per MAG, where the information of the 3 hierarchical levels of the KEGG database were added, and only those KOs belonging to *09100 Metabolism* were kept for further analysis.

Detection of Horizontal Gene Transfer (HGT) events

A custom blast database was built with all MAGs obtained. A Blastn analysis of the MAGs fasta files, containing all MAGs sequences correctly labeled on the headers to discriminate to which MAGs belonged each contig, was performed against the MAGs database with a 99.9% identity cut-off. The blastn output was filtered to remove hits where query and subject belonged to the same genus and to remove duplicated hits. Hits with alignment length shorter than 500bp were also discarded for further analysis. HGT sequences were extracted by bedtools using the query sequences position from the blastn output. eggNOG-mapper v2.1.5[34] was employed, as previously described for MAGs, to functionally characterise HGT sequences. Detection of CDS related to integrons, transposases or phages was done by searching "integron", "transposase" and "phage" on the eggNOG-mapper output file, while detection of plasmids was done by searching "relaxase" and by employing PlasFlow (https://github.com/smaegol/PlasFlow) as a second approach. The clustering of HGT sequences was done using vsearch[36] and 0.99 as identity cutoff. Chord diagrams for HGT events were plotted using the circlize R-package.

Diversity and taxonomy plots and statistical analysis

Only those taxa belonging to the kingdoms Bacteria and Fungi were used for further analyses, which were executed in parallel.

The β-diversity was estimated by Principal Coordinates Analysis using Bray-Curtis dissimilarity and the *vegdist* function. Within-group dispersion was evaluated through the *betadisper* function. Both functions are located in the R package *vegan* (https://github.com/vegandevs/vegan). The effects of cave, producer and sample type (which includes ripening stage and cheese part for cheese samples, and food contact *vs* non food contact surfaces for environmental samples) on sample dissimilarities were determined by permutational multivariate analysis of variance using distance matrices (PERMANOVA) with the *adonis* function in the R package *vegan* (https://github.com/vegandevs/vegan). The *compare_means* function in the R package *ggpubr* was used to include statistically significant differences on boxplot figures, which were plotted by using the R package *ggplot2* (https://github.com/tidyverse/ggplot2).

Comparisons between multiple groups of samples for taxonomy were performed by using the Kruskal Wallis test and the *post hoc* Wilcoxon signed-rank test. P-values were adjusted through the Benjamini & Hochberg method[37] and significance was established at $p < 0.05$.

Correlation analyses were performed using the Spearman correlation coefficient, calculated with the *rcorr* function in the R package *Hmisc* (https://hbiostat.org/R/Hmisc/).

All analyses were carried out using R version 3.6.2 (https://cran.r-project.org/).

Source tracking analysis

Genera relative abundance matrices were employed to attribute the main sources of the cheese microbial populations by using SourceTracker2[38] with the *–per_sink_feature_assignments* parameter. The source-

to-sink genera contribution dataset was reorganized by using in-house scripts (https://github.com/JoseCoboDiaz/piecharts_tax_st2) to plot a matrix of piecharts that indicates the influence of each source to each sink (by the size of the piechart) and the percent influence of each genera in each source-to-sink pair.

Data availability

The raw reads fastq files from shotgun metagenome sequencing of the 133 samples and the genome fasta files from the 613 medium and high quality MAGs obtained are available under NCBI BioProject ID PRJNA764004, with physical BioSample accession numbers from SAMN21465692 to SAMN21465824 and SRA accession numbers from SRR15927537 to SRR15927669 for raw reads fastq files, and BioSample accesion numbers from SAMN23026855 to SAMN23027467 for MAGs-genome fasta files.

Declarations

Acknowledgements

The laboratory of AAO, ML and MP have received funding by the Ministry of Science and Innovation of the Spanish Government, under grant numbers AGL2016-78085-P and MCIN/AEI/10.13039/501100011033, and the European Commission under the European Union´s Horizon 2020 research and innovation program under grant agreement No 818368. AAM is grateful to the Ministry of Science, Innovation and Universities of the Spanish Government for awarding him a pre-doctoral grant (FPI BES-2017-079753). PFG is grateful to Junta de Castilla y León and the European Social Fund (ESF) for awarding her a pre-doctoral grant (BOCYL-D-15122017-4). B. Redruello is acknowledged for her assistance in biogenic amines analysis. The authors wish to acknowledge the cheese producers for their availability and support throughout the study.

References

1. Yeluri Jonnala, B. R., McSweeney, P. L. H., Sheehan, J. J. & Cotter, P. D. Sequencing of the cheese microbiome and its relevance to industry. *Frontiers in Microbiology* **9**, (2018).

2. Kamilari, E., Tomazou, M., Antoniades, A. & Tsaltas, D. High throughput sequencing technologies as a new toolbox for deep analysis, characterization and potentially authentication of protection designation of origin cheeses? *International Journal of Food Science* **2019**, (2019).

3. Penland, M. *et al.* Linking Pélardon artisanal goat cheese microbial communities to aroma compounds during cheese-making and ripening. *International Journal of Food Microbiology* **345**, (2021).

4. Mayo, B., Rodríguez, J., Vázquez, L. & Flórez, A. B. Microbial interactions within the cheese ecosystem and their application to improve quality and safety. *Foods* **10**, (2021).

5. Quijada, N. M. *et al.* Austrian raw-milk hard-cheese ripening involves successional dynamics of non-inoculated bacteria and fungi. *Foods* **9**, 1–21 (2020).

6. Wolfe, B. E., Button, J. E., Santarelli, M. & Dutton, R. J. Cheese rind communities provide tractable systems for in situ and in vitro studies of microbial diversity. *Cell* **158**, 422–433 (2014).

7. Bertuzzi, A. S. *et al.* Omics-based insights into flavor development and microbial succession within surface-ripened cheese. *mSystems* **3**, (2018).

8. Walsh, A. M., Macori, G., Kilcawley, K. N. & Cotter, P. D. Meta-analysis of cheese microbiomes highlights contributions to multiple aspects of quality. *Nature Food* **1**, 500–510 (2020).

9. Afshari, R., Pillidge, C. J., Dias, D. A., Osborn, A. M. & Gill, H. Cheesomics: the future pathway to understanding cheese flavour and quality. *Critical Reviews in Food Science and Nutrition* **60**, 33–47 (2020).

10. Walsh, A. M. *et al.* Microbial succession and flavor production in the fermented dairy beverage kefir. *mSystems* **1**, 1–17 (2016).

11. Flórez, A. B. & Mayo, B. Microbial diversity and succession during the manufacture and ripening of traditional, Spanish, blue-veined Cabrales cheese, as determined by PCR-DGGE. *International Journal of Food Microbiology* **110**, 165–171 (2006).

12. Wadhawan, K., Steinberger, A. J., Rankin, S. A., Suen, G. & Czuprynski, C. J. Characterizing the microbiota of wooden boards used for cheese ripening. *JDS Communications* **2**, 171–176 (2021).

13. Frétin, M. *et al.* Bacterial community assembly from cow teat skin to ripened cheeses is influenced by grazing systems. *Scientific Reports* **8**, 1–11 (2018).

14. Thierry, A. *et al.* Strain-to-strain differences within lactic and propionic acid bacteria species strongly impact the properties of cheese–A review. *Dairy Science and Technology* **95**, 895–918 (2015).

15. Bertuzzi, A. S., McSweeney, P. L. H., Rea, M. C. & Kilcawley, K. N. Detection of volatile compounds of cheese and their contribution to the flavor profile of surface-ripened cheese. *Comprehensive Reviews in Food Science and Food Safety* **17**, 371–390 (2018).

16. Salazar, J. K. *et al.* Metagenomics of pasteurized and unpasteurized Gouda cheese using targeted 16S rDNA sequencing. *BMC Microbiology* **18**, 1–13 (2018).

17. Unno, R. *et al.* Lactic acid bacterial diversity in Brie cheese focusing on salt concentration and pH of isolation medium and characterisation of halophilic and alkaliphilic lactic acid bacterial isolates. *International Dairy Journal* **109**, 104757 (2020).

18. Morales, F., Morales, J. I., Hernández, C. H. & Hernández-Sánchez, H. Isolation and partial characterization of halotolerant lactic acid bacteria from two Mexican cheeses. *Applied Biochemistry and Biotechnology* **164**, 889–905 (2011).

19. Justé, A. *et al.* Genetic and physiological diversity of *Tetragenococcus halophilus* strains isolated from sugar- and salt-rich environments. *Microbiology* **154**, 2600–2610 (2008).

20. Bonham, K. S., Wolfe, B. E. & Dutton, R. J. Extensive horizontal gene transfer in cheese-associated bacteria. *eLife* **6**, 1–23 (2017).

21. Somerville, V. *et al.* Long read-based de novo assembly of low complex metagenome samples results in finished genomes and reveals insights into strain diversity and an active phage system. *BMC Microbiology* **19**, 143 (2019) doi:10.1186/s12866-019-1500-0.

22. Redruello, B. *et al.* A fast, reliable, ultra high performance liquid chromatography method for the simultaneous determination of amino acids, biogenic amines and ammonium ions in cheese, using diethyl ethoxymethylenemalonate as a derivatising agent. *Food Chemistry* **139**, 1029–1035 (2013).

23. Panthi, R. R. *et al.* Influence of herd diet on the metabolome of Maasdam cheeses. *Food Research International* **123**, 722–731 (2019).

24. Van Den Dool, H., & Dec. Kratz, P. A generalization of the retention index system including linear temperature programmed gas-liquid partition chromatography. *Journal of Chrmoatography A*, **11**, 463-471 (1963).

25. Chen, S. *et al.* AfterQC: Automatic filtering, trimming, error removing and quality control for fastq data. *BMC Bioinformatics* **18**, (2017).

26. Langmead, B., Wilks, C., Antonescu, V. & Charles, R. Scaling read aligners to hundreds of threads on general-purpose processors. *Bioinformatics* **35**, 421–432 (2019).

27. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

28. Wood, D. E., Lu, J. & Langmead, B. Improved metagenomic analysis with Kraken 2. *Genome Biology* **20**, 257 (2019).

29. Nurk, S., Meleshko, D., Korobeynikov, A. & Pevzner, P. A. MetaSPAdes: A new versatile metagenomic assembler. *Genome Research* **27**, 824–834 (2017).

30. Kang, D. D. *et al.* MetaBAT 2: An adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ* **2019**, 1–13 (2019).

31. Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM: Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Research* **25**, 1043–1055 (2015).

32. Von Meijenfeldt, F. A. B., Arkhipova, K., Cambuy, D. D., Coutinho, F. H. & Dutilh, B. E. Robust taxonomic classification of uncharted microbial sequences and bins with CAT and BAT. *Genome Biology* **20**, 1–14 (2019).

33. Olm, M. R., Brown, C. T., Brooks, B. & Banfield, J. F. DRep: A tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. *ISME Journal* **11**, 2864–2868 (2017).

34. Cantalapiedra, C. P., Hernández-Plaza, A., Letunic, I., Bork, P. & Huerta-Cepas, J. eggNOG-mapper v2: Functional annotation, orthology assignments, and domain prediction at the metagenomic scale. *Molecular Biology and Evolution* 1–5 (2021) doi:10.1093/molbev/msab293.

35. Mirdita, M., Steinegger, M., Breitwieser, F., Söding, J. & Levy Karin, E. Fast and sensitive taxonomic assignment to metagenomic contigs. *Bioinformatics* **37**, 3029–3031 (2021).

36. Rognes, T., Flouri, T., Nichols, B., Quince, C. & Mahé, F. VSEARCH: A versatile open source tool for metagenomics. *PeerJ* **2016**, 1–22 (2016).

37. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B.* 57 (1), 289–300 (1995).

38. Knights, D. *et al.* Bayesian community-wide culture-independent microbial source tracking. *Nature Methods* **8(9),** 761–763 (2011).

Figures

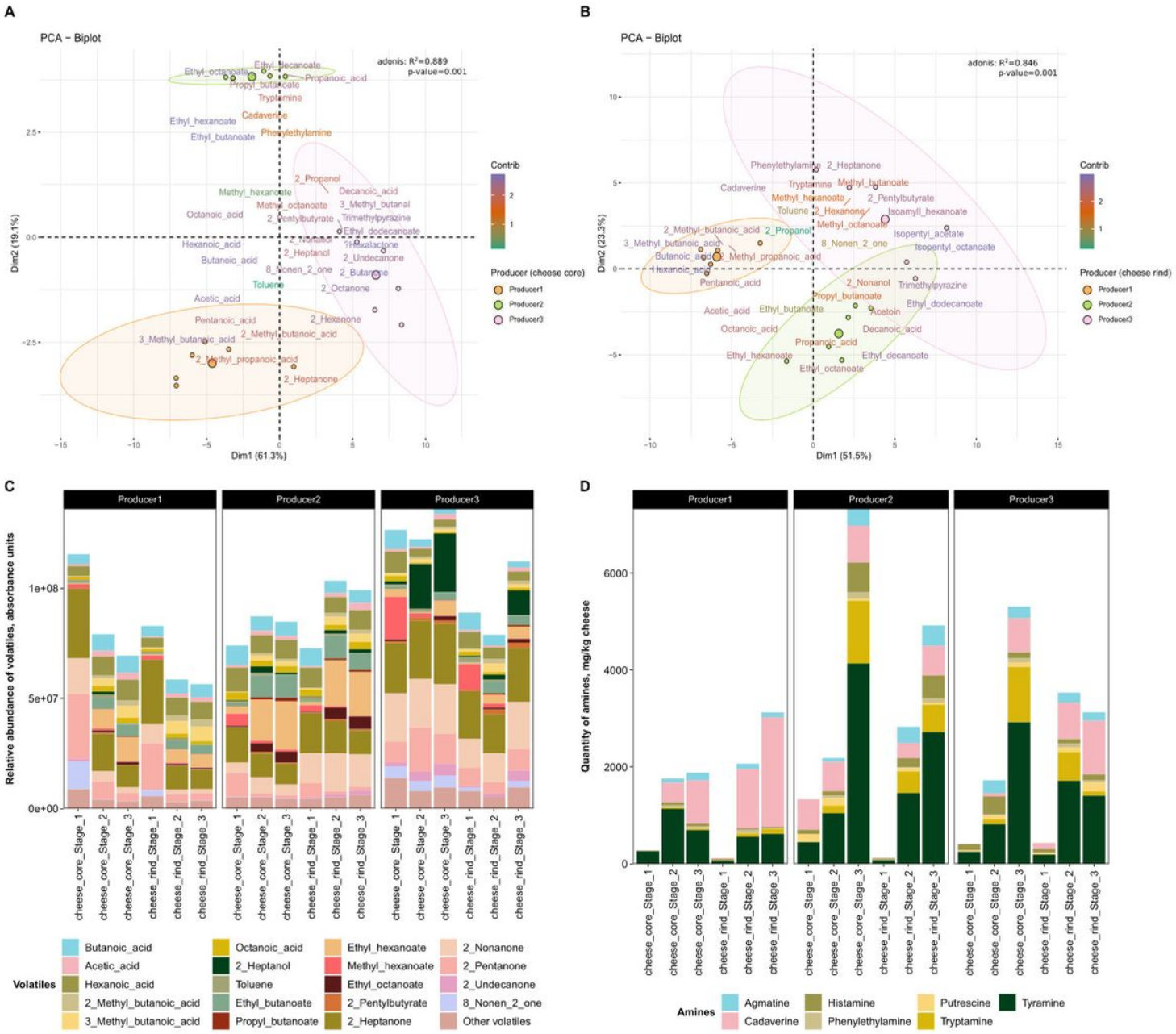

Figure 1

Metabolome of blue-veined cheeses during ripening in the communal cave.

Biplots of principal component analysis of metabolites (including biogenic amines and volatiles) found in cheese core **(A)** and cheese rind **(B)** samples ripened in the communal cave. The biplots show the most significant metabolites contributing to the metabolome of cheeses ripened in Cave2 (communal cave) by the three producers. Colored concentration ellipses (size determined by a 0.95-probability level) show the observations grouped by producer. Contrib (Contribution) indicates the average value of percentages of contribution to the 2 first axis for each variable.

Barplots presenting volatile compounds **(C)** Absorbance units (AU) (quantitatively minor volatile compounds were grouped as “Other volatiles”); and biogenic amines content **(D)**. Each bar represents the mean value (n=3) for blue-veined cheese samples (cores and rinds separately) belonging to Cave2 (communal cave) for each Producer (Producer1, Producer2 and Producer3) during three ripening stages (Stage1, Stage2 and Stage3).

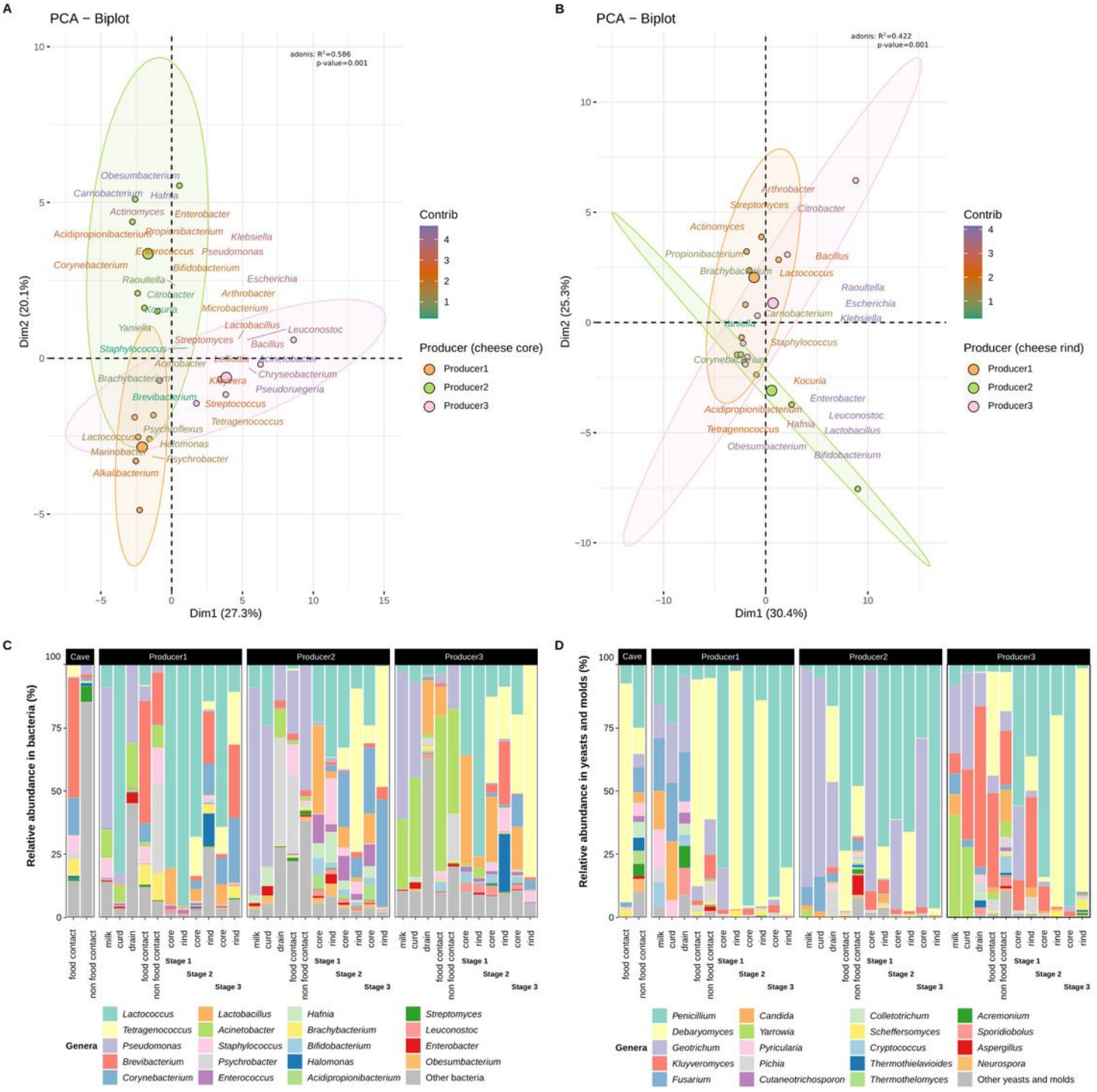

Figure 2

Microbiome of blue-veined cheeses during ripening in the communal cave.

Biplots of principal component analysis of the forty most abundant bacterial genera found in the cheese core **(A)** and cheese rind **(B)** samples ripened in the communal cave depending on the Producer. The biplots show the most significant bacterial genera contributing to the microbiome of cheeses ripened in Cave2 (communal cave) by the three producers. Colored concentration ellipses (size determined by a 0.95-probability level) show the observations grouped by producer.

Barplots representing the relative abundance of the 19 most relevant bacterial **(C)** and yeasts and molds **(D)** genera. Other minority bacteria and fungi genera are grouped into “Other”. Each bar represents the average value (n=3) for blue-veined cheese samples from each Producer.

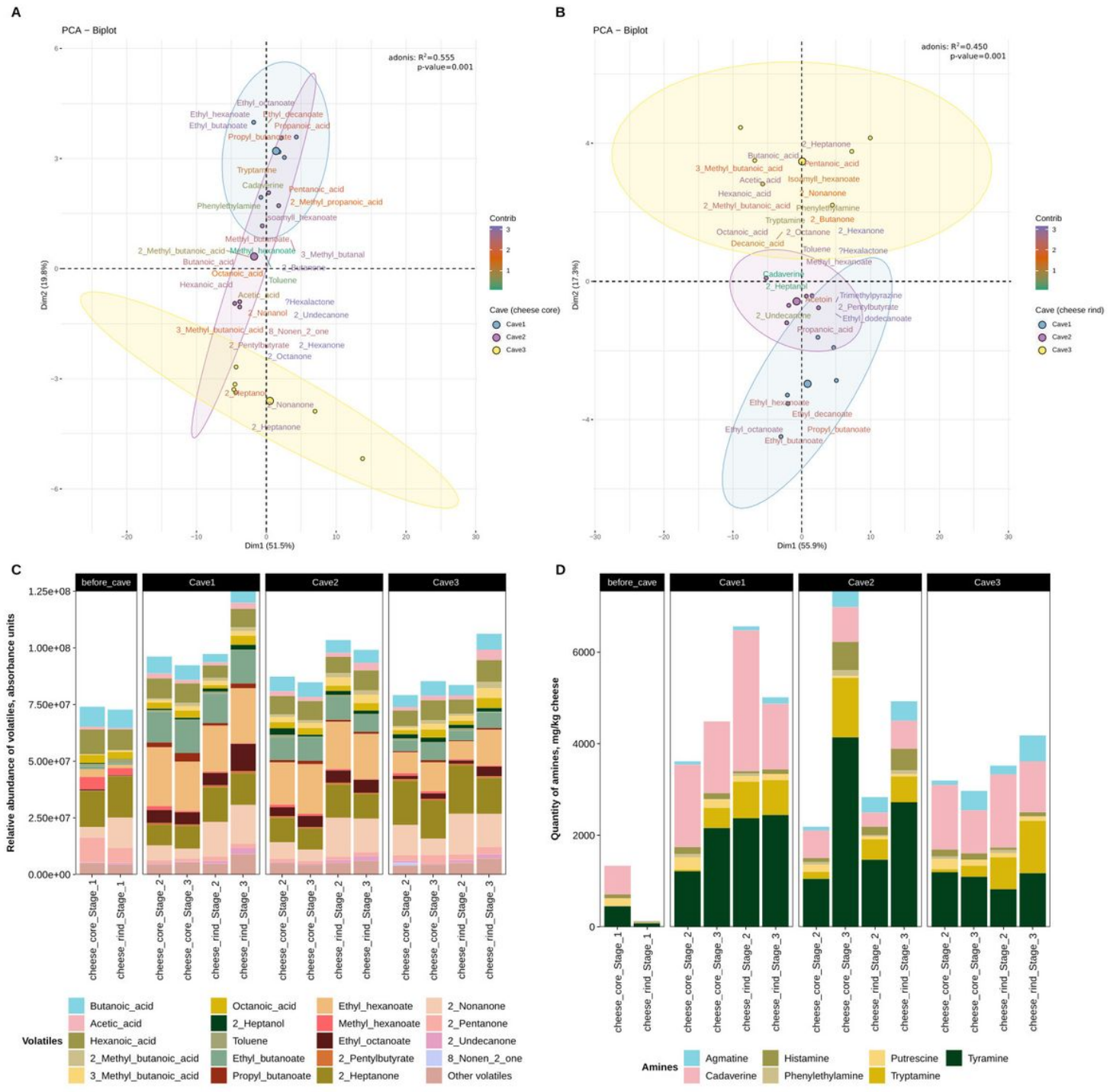

Figure 3

Metabolome of blue-veined cheeses from Producer2 during ripening in the three different caves.

Biplots of principal component analysis of metabolites (including biogenic amines and volatiles) found in cheese core **(A)** and cheese rind **(B)** samples ripened by Producer2 in the three different caves. The biplots show the most significant metabolites contributing to the metabolome of cheeses in the three different caves. Colored concentration ellipses (size determined by a 0.95-probability level) show the

observations grouped by cave. Contrib (Contribution) indicates the average value of percentages of contribution to the 2 first axis for each variable.

Barplots presenting volatile compounds **(C)** Absorbance units (AU) (quantitatively minor volatile compounds were grouped as “Other volatiles”) and biogenic amines content **(D)**. Each bar represents the mean value (n=3) for blue-veined cheese samples (cores and rinds separately) before cave ripening (Stage1) and during ripening (Stage2 and Stage3) in the three different natural Caves (Cave1, Cave2, Cave3).

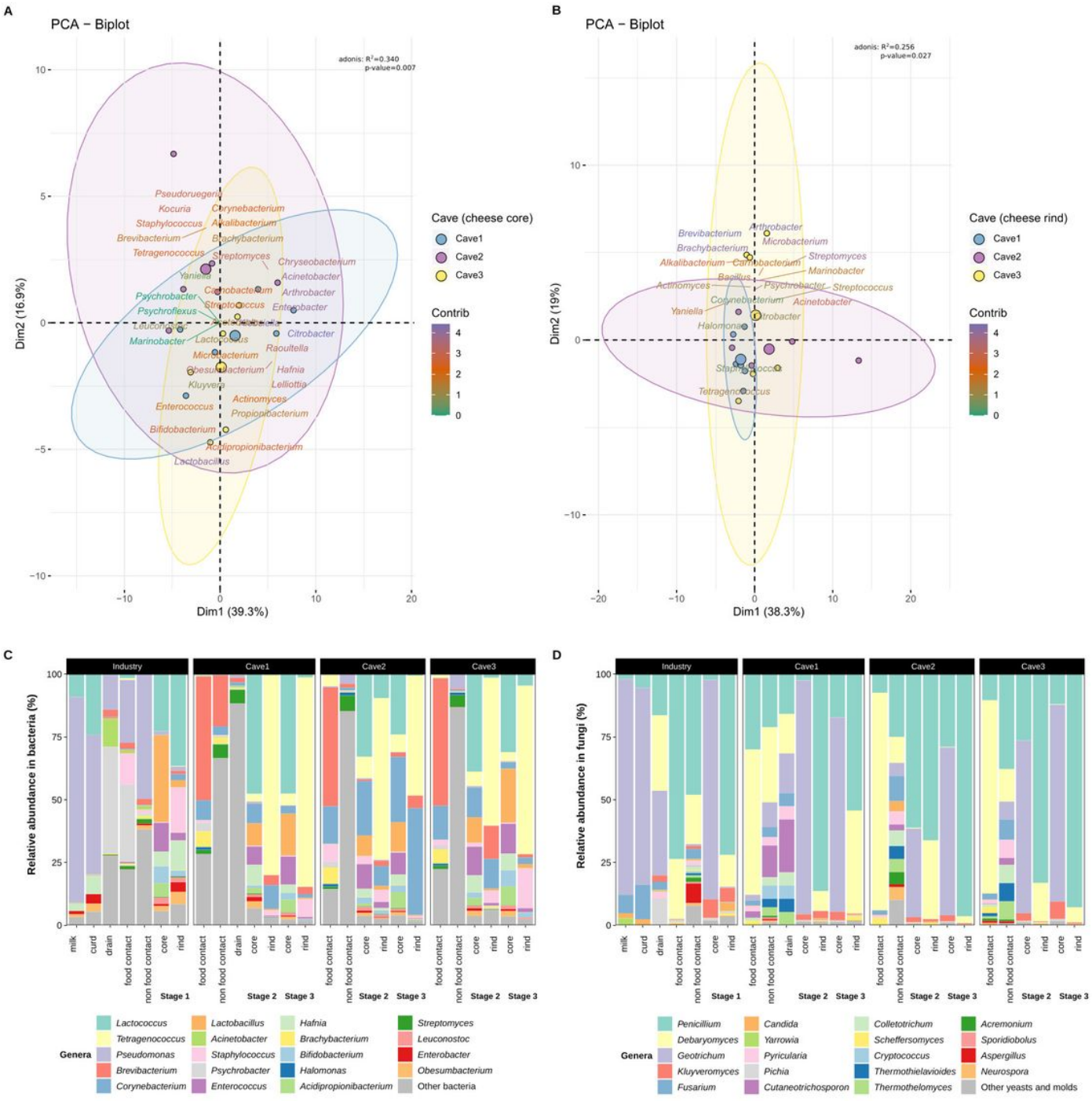

Figure 4

Microbiome of blue-veined cheeses from Producer2 during ripening in the three different caves.

Biplot of principal component analysis of the forty most abundant bacterial genera found in the cheese core **(A)** and cheese rind **(B)** samples ripened in the three different caves. The biplots show the most significant bacterial genera contributing to the cheese microbiome in the three different caves. Colored concentration ellipses (size determined by a 0.95-probability level) show the observations grouped by cave.

Barplots representing the relative abundance of the 19 most relevant bacterial **(C)** and yeasts and molds **(D)** genera. Other minority bacteria and fungi genera are grouped into “Other”. Each bar represents the average value (n=3) for blue-veined cheese samples in each Cave.

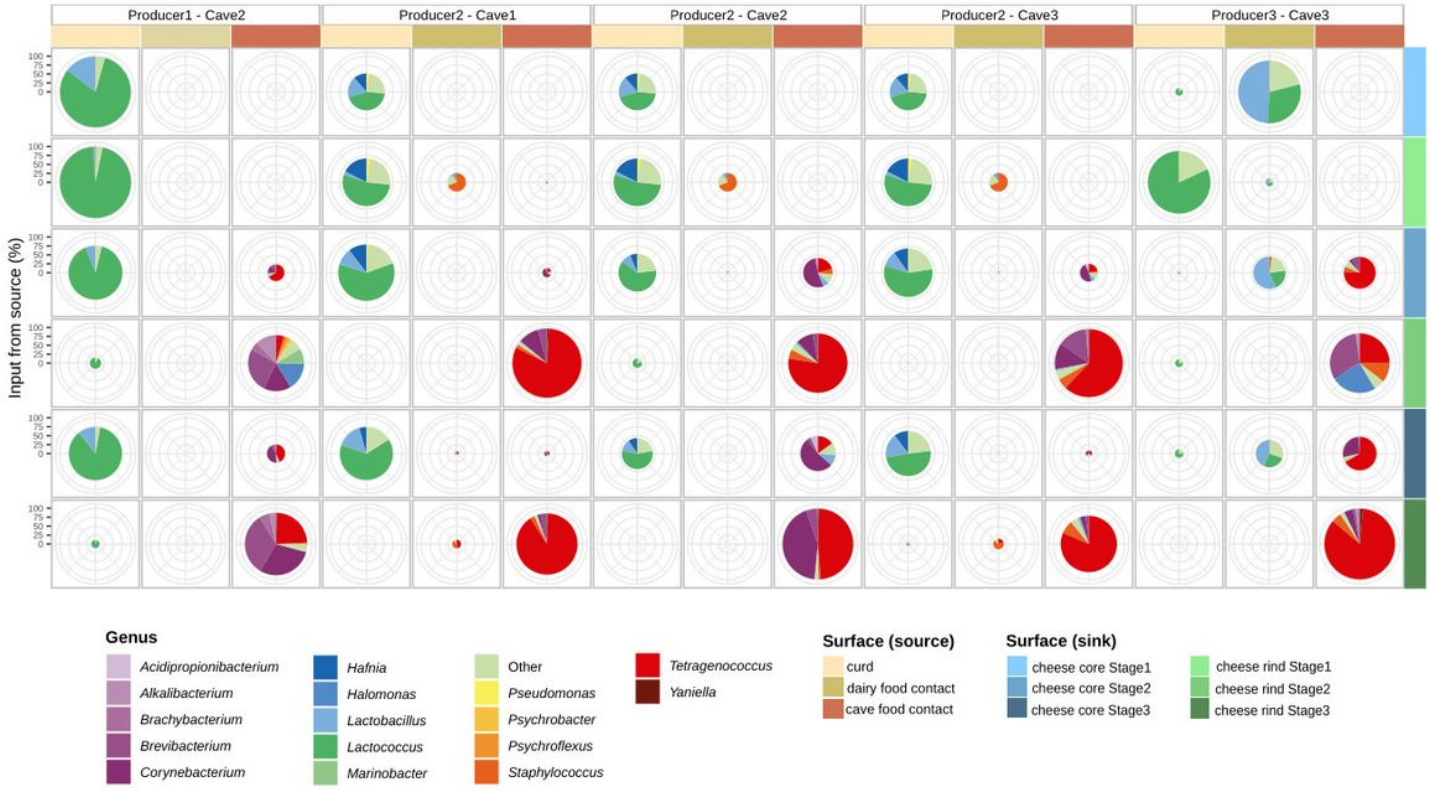

Figure 5

Taxonomical (bacteria) source attribution of cheese samples calculated by SourceTracker2 software.

Piechart plots represent bacterial sources (on columns, *source* samples) for the cheese bacterial community composition (on rows, *sink* samples). The ratio of the piechart is proportional to the percentage of *source* sample influence on *sink* sample (indicated on the y-axis). The colors within each

piechart indicate the percentage of genus influence for each source-sink pair. Only *source* samples and 16 main genera with significant influence were represented, while other genera were grouped as “Other”.

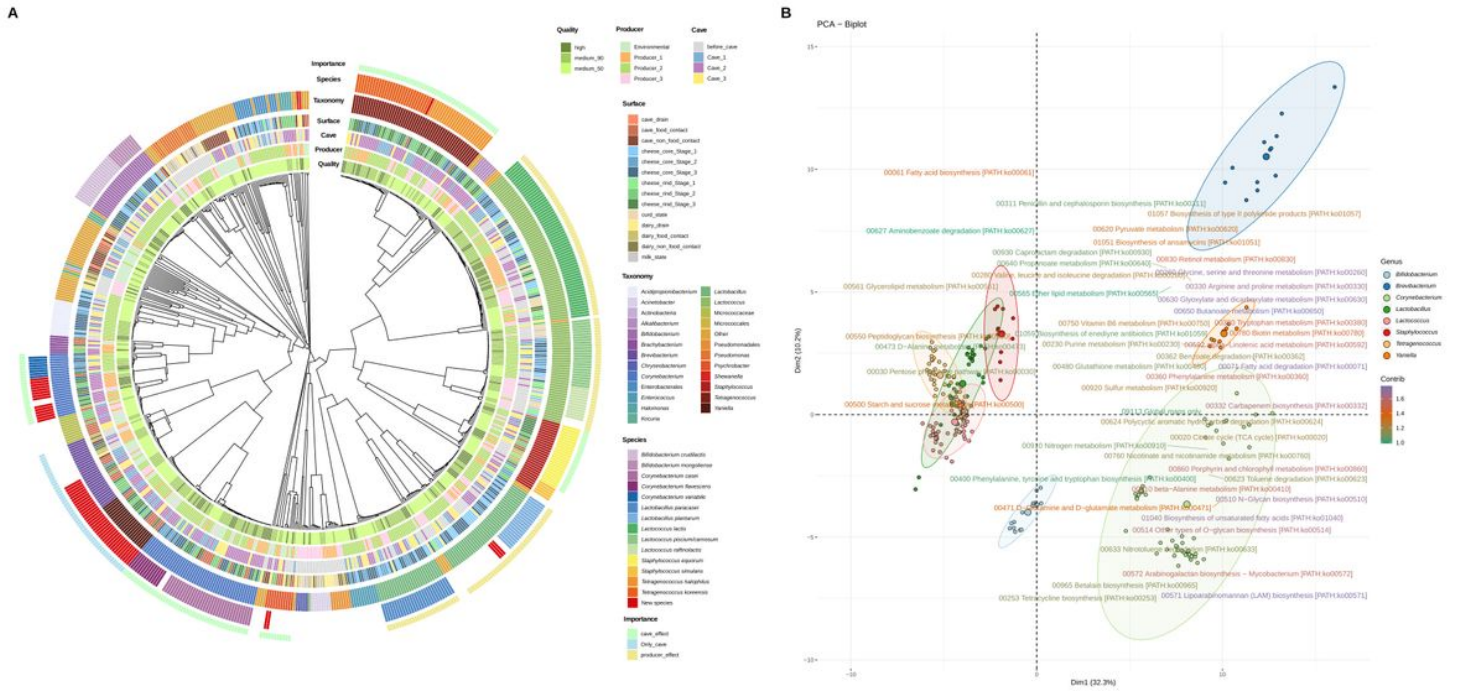

Figure 6

Phylogenetic tree of MAGs obtained by per-sample assembly and binning of metagenomic reads.

(A) Phylogenetic tree constructed with the ANI distance matrix generated by using the dRep software. Concentric circles indicate, from inside to outside, the quality of MAGs; producer, cave and surface origin of the sample; genus and species classification of the MAGs; and importance of the MAGs. Only the 24 most abundant taxa at genus level (or above) are indicated, while the rest are grouped in “Other”. Some taxa are classified in higher taxonomic levels (phylum, family or order) according to the deep level of taxonomic assignment obtained by using the CAT/BAT software. Indicated species were taxonomically re-assigned by building additional phylogenetic trees employing reference genomes (RefSeq database) or other representative genomes from NCBI. The last circle indicates clusters of interest for cave and producer effects on the microbiome. Only new species within the genera of interest (*Brevibacterium*, *Corynebacterium*, *Lactobacillus, Tetragenococcus* and *Yaniella*) were marked with red color on the “species” circle.

(**B**) Principal Coordinate Analysis of the MAGs functional composition by level 3 of KEGG Orthology classification. Only functional groups belonging to *09100 Metabolism* were employed. Only MAGs of interest due to their importance related to the producer and/or cave influence were employed.

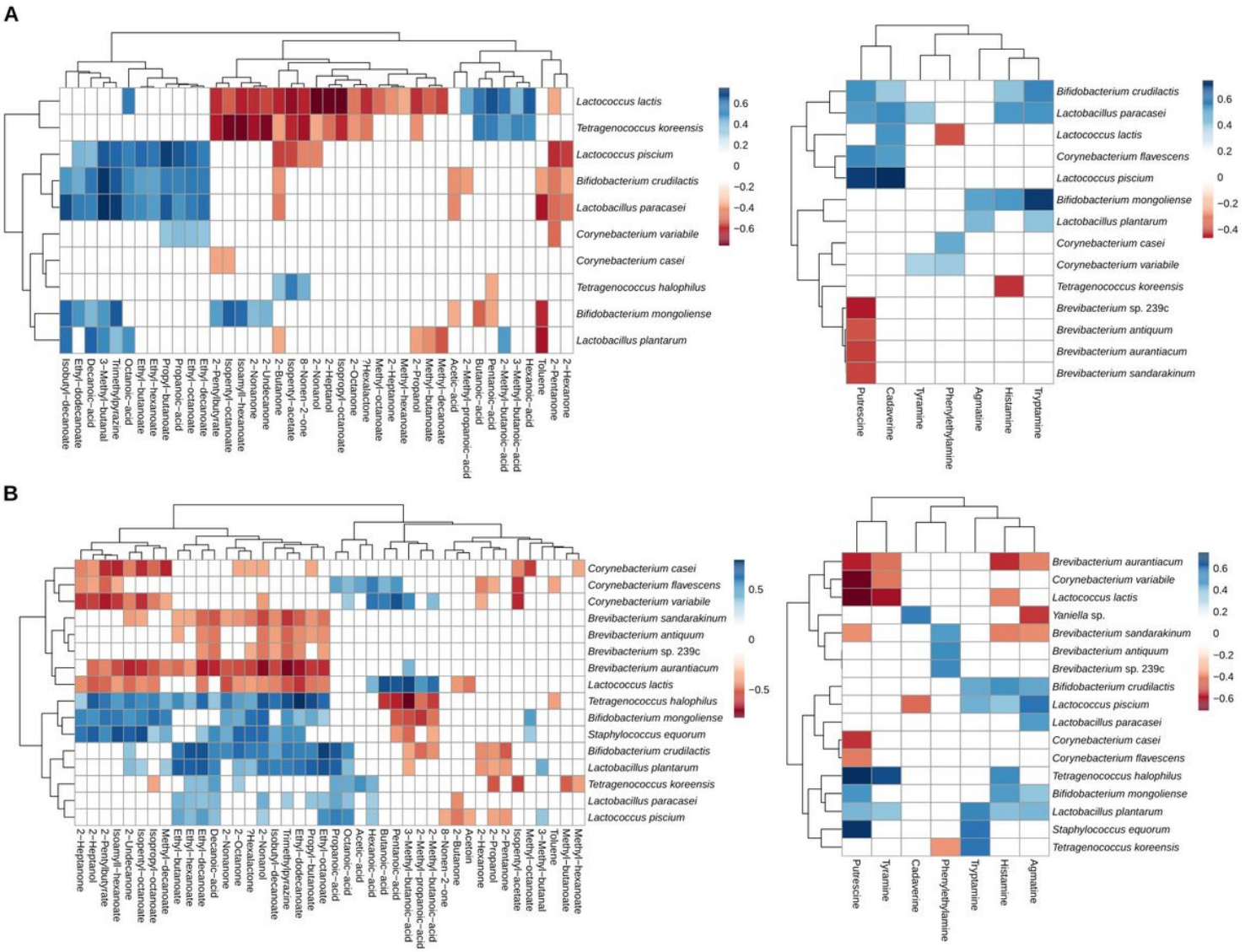

Figure 7

Spearman correlation between abundance of metabolites and bacterial species.

Heatmaps representing significant ($p<0.05$) Spearman correlations between the abundance of volatiles and biogenic amines analyzed and species abundance at read level for cheese core **(A)** and cheese rind **(B)** samples. Only those species of special interest with metagenomic assembled genomes (MAGs) obtained by using the metaSPAdes + METABAT2 pipeline were employed for correlation analyses. Cheese samples from Stage1 (before cave ripening) were discarded for correlation analyses.

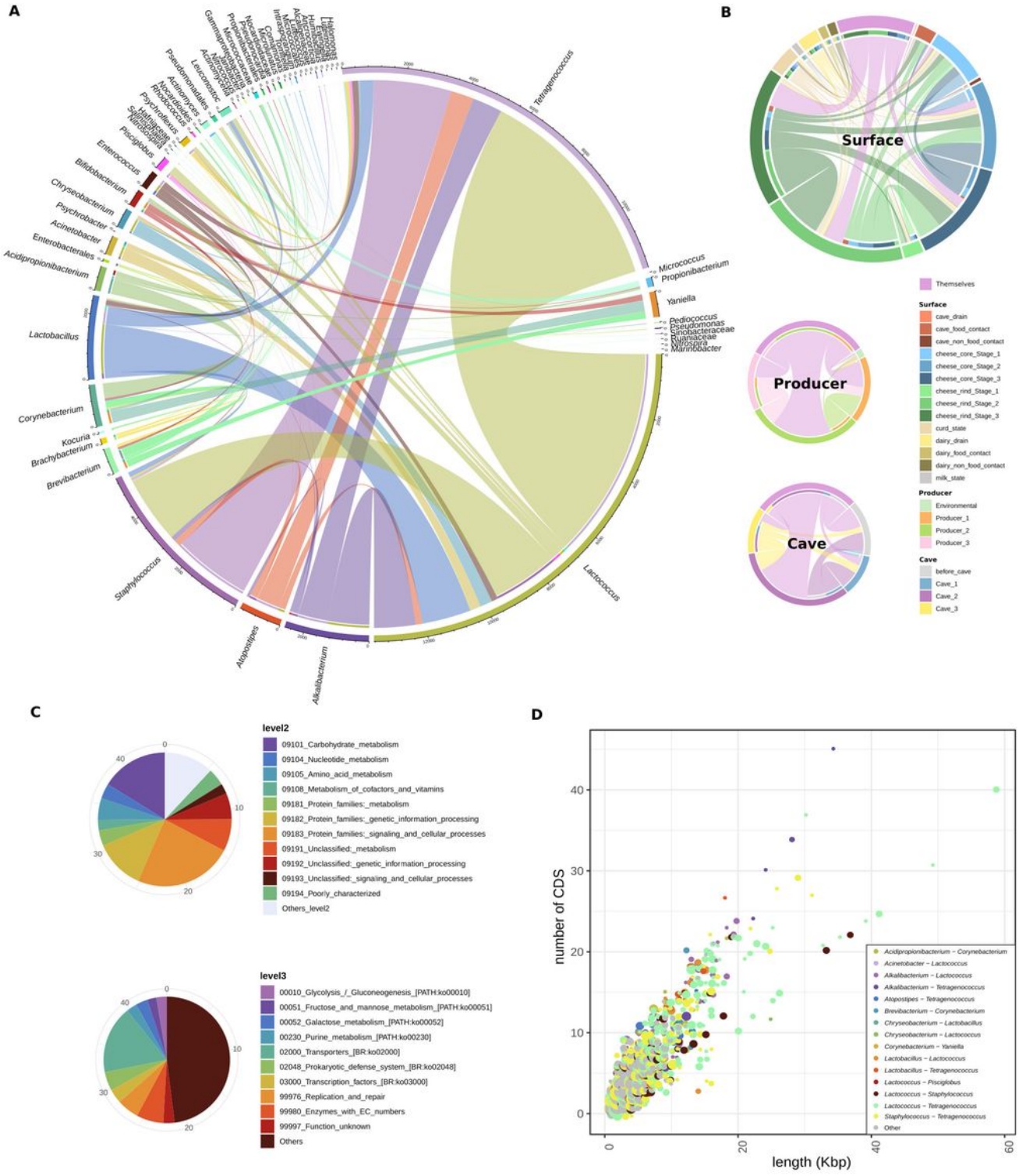

Figure 8

Horizontal Gene Transfer (HGT) events.

HGT events (identity > 99.9%, length > 500 bp) detected between MAGs belonging to different genera **(A)** and classified by surface, producer and cave **(B).** The thickness of the circle internal lines is proportional to the number of HGT events detected for each genera pair. The “Themselves” group corresponds with

HGT events between different genera from the same surface, producer or cave group. (**C)** Percentage of coding regions (CDS) detected within HGT sequences belonging to different functions of the level2 and level3 of KEGG Orthology functional classification. Only 45% of CDS detected, with KO code assigned, were represented. (**D)** Relation between the length of HGT events and the number of CDS detected in them. Circle colors indicate the genera pair for each HGT event, while the size of the circles is square-root proportional to the number of HGT events with same length and CDS number.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- FigureS1.pdf
- FigureS2.pdf
- FigureS3.pdf
- FigureS4.pdf
- FigureS5.pdf
- FigureS6.pdf
- FigureS7.pdf
- FigureS8.pdf
- FigureS9.pdf
- TableS1.xlsx
- TableS2.xlsx
- TableS3.xlsx
- TableS4.xlsx
- TableS5.xlsx