npj | biofilms and microbiomes

Article

Published in partnership with Nanyang Technological University

https://doi.org/10.1038/s41522-025-00815-6

Microbiome profiling of Grana Padano and Parmigiano Reggiano cheeses reveals cheese-specific biomarkers, psychobiotic potential, and bioprotective activities

Vincenzo Valentino[1,5], Raffaele Magliulo[1,2,5], Andrea Balivo[1], Alicja Monika Krysmann[3], Chiara Maria Calvanese[1], Alessia Esposito[1], Giuseppina Sequino[1], Alessandro Genovese[1], Davide Porcellato[3], Danilo Ercolini[1,4] & Francesca De Filippis[1,2,4] ✉

Grana Padano (GP), Trentingrana (TG), and Parmigiano Reggiano (PR) are among the finest Italian Protected Designation of Origin (PDO) cheeses. GP, TG, and PR undergo extensive proteolysis during ripening, where the microbiome metabolizes amino acids, producing flavour and bioactive molecules. We explored the microbiome, volatilome, and metaproteome of PDO GP (n = 42), TG (n = 18), and PR (n = 60). Findings revealed diverse microbial communities enriched in proteolytic microbes, associated with cheese-specific processing technology. Correlations between lactic acid bacteria strains and specific volatile compounds were identified in PR. Importantly, we identified genes involved in the production of neuroactive molecules, suggesting potential connections between cheeses consumption and mental health, along with genes related to bacteriocin biosynthesis, possibly enhancing cheese safety, shelf life, and process sustainability. This study provides novel insights into the functional attributes of long-ripened cheeses microbiome, highlighting their potential as sources of psychobiotics and bioprotective strains.

Grana Padano (GP), Trentingrana (TG), and Parmigiano Reggiano (PR) are three of the most well-known Italian hard cheeses, recognized not only for their unique sensory properties but also for their pivotal role in Italian culture and market[1]. All these three cheeses have obtained the label of Protected Designation of Origin (PDO), indicating their strong relationship with specific production areas. The PDO label ensures that there are very strict rules in cheese production, which help to preserve authenticity (EU Regulation 2024/1143). GP and TG (defined as a "variety" of GP) are protected by the same PDO regulation, which is different from PR. According to the regulations, GP is made in the Po River Valley in Northern Italy, while TG originates from Alpine regions. On the contrary, PR is produced in a smaller area across the Emilia-Romagna and Lombardia regions. A key difference lies in cattle feeding: according to the PDO regulation, silage is allowed in GP, but strictly prohibited for TG and PR. All three cheeses use natural whey cultures (NWC) as starters, but their milk processing varies: GP and TG employ partially skimmed milk, processed within 24 h, whereas PR combines evening-skimmed milk with fresh morning milk. Another distinction is the use of lysozyme, added only in GP, to prevent microbial spoilage due to clostridia often present in silage. During production, GP and TG curds are cut into fine granules and cooked at 53–56 °C, while PR undergoes a similar process at a slightly lower temperature (52–54 °C). Additionally, PR curds are manually extracted using the traditional *"spinatura"* technique, which contributes to their characteristic texture.

GP, TG, and PR undergo a very long ripening process, which may exceed 12 months, that is associated with a strong proteolysis. This proteolysis does not only entail protein metabolism, but a complex biochemical network initiated by the cheese microbiome[2], which is key in controlling flavor, texture, and also the nutritional value of these cheeses. Thermophilic lactic acid bacteria (LAB) are active in curd fermentation and during the first phases of ripening. During the maturation, mesophilic LAB take over[3,4]. While thermophilic LAB originate from the NWC, mesophilic LAB are mainly non-starter (NSLAB) and may come from milk or the cheese-making environment[5]. NSLAB persist metabolically active during cheese

[1]Department of Agricultural Sciences, University of Naples Federico II, Portici, Italy. [2]NBFC—National Biodiversity Future Center, Palermo, Italy. [3]Faculty of Chemistry, Biotechnology and Food Science, Norwegian University of Life Sciences (NMBU), Ås, Norway. [4]Task Force on Microbiome Studies, University of Naples Federico II, Naples, Italy. [5]These authors contributed equally: Vincenzo Valentino, Raffaele Magliulo. ✉e-mail: francesca.defilippis@unina.it

ripening, and their activity leads to the production of several compounds, some of which may be directly responsible for flavors, while others are bioactive compounds associated with various health-promoting properties[6,7].

The cheese microbiome plays a key role in flavor development, driving biochemical reactions during fermentation and ripening. Lipolysis of fatty acids leads to the formation of free fatty acids, methyl ketones, and secondary alcohols, all of which are important contributors to cheese aroma[8]. Moreover, microbial enzymes break down milk proteins, releasing peptides and amino acids that contribute to taste and aroma[9,10]. This proteolytic activity is particularly prevalent in long-aged cheeses, where the catabolism of amino acids leads to the formation of various volatile compounds, such as 2- and 3-methylbutanal, methanethiol, and ethyl esters, which play an active role in the aroma of cheeses like Parmigiano[11,12]. In long-ripened cheeses, extensive proteolysis shapes the peptide profiles, while differences in the microbial community further influence the flavor development[13].

Beyond sensory properties, these types of cheeses often contain bioactive peptides with antihypertensive, antioxidant, and immunomodulatory functions[14]. Notably, PR can harbor angiotensin-converting enzyme inhibitory peptides, which may help regulate blood pressure[15]. The presence of neuroactive compounds was also reported, including opioid-like peptides in GP[14] and anxiolytic peptides in PR[16]. This aligns with the reawakening interest in fermented foods and psychobiotic microbes, defined as live organisms conferring a health benefit in people affected by psychiatric illness, which can influence mental health when consumed in adequate amounts[17]. The connection between diet, gut microbiome, and mental disorders, including autism, multiple sclerosis, and schizophrenia, is now a major research focus with broad public implications[18,19].

In this work, we attempted to identify PDO cheese taxonomic and functional markers and explored the role of the cheese microbiome in shaping the volatile compounds that contribute to the aroma of long-ripened Italian PDO cheeses. Furthermore, we shed light on the cheese microbiome metabolic activities potentially linked with health outcomes, including the potential production of neuroactive compounds impacting the gut-brain axis. Our results provide further insights into the potential of the microbial communities inhabiting long-ripened cheeses and may therefore have wide applications, in both the dairy industry and public health.

Results

Taxonomic profiles of GP, PR, and TG cheese microbiome reveal the dominance of non-starter LAB (NSLAB)

The microbiome profiles of GP, PR, and TG cheeses were dominated by *Lacticaseibacillus paracasei* and *Lc. rhamnosus*, cumulatively accounting for >40% of the total microbial abundance (Fig. 1A). In particular, *Lc. rhamnosus* was the most abundant species overall (average relative abundance

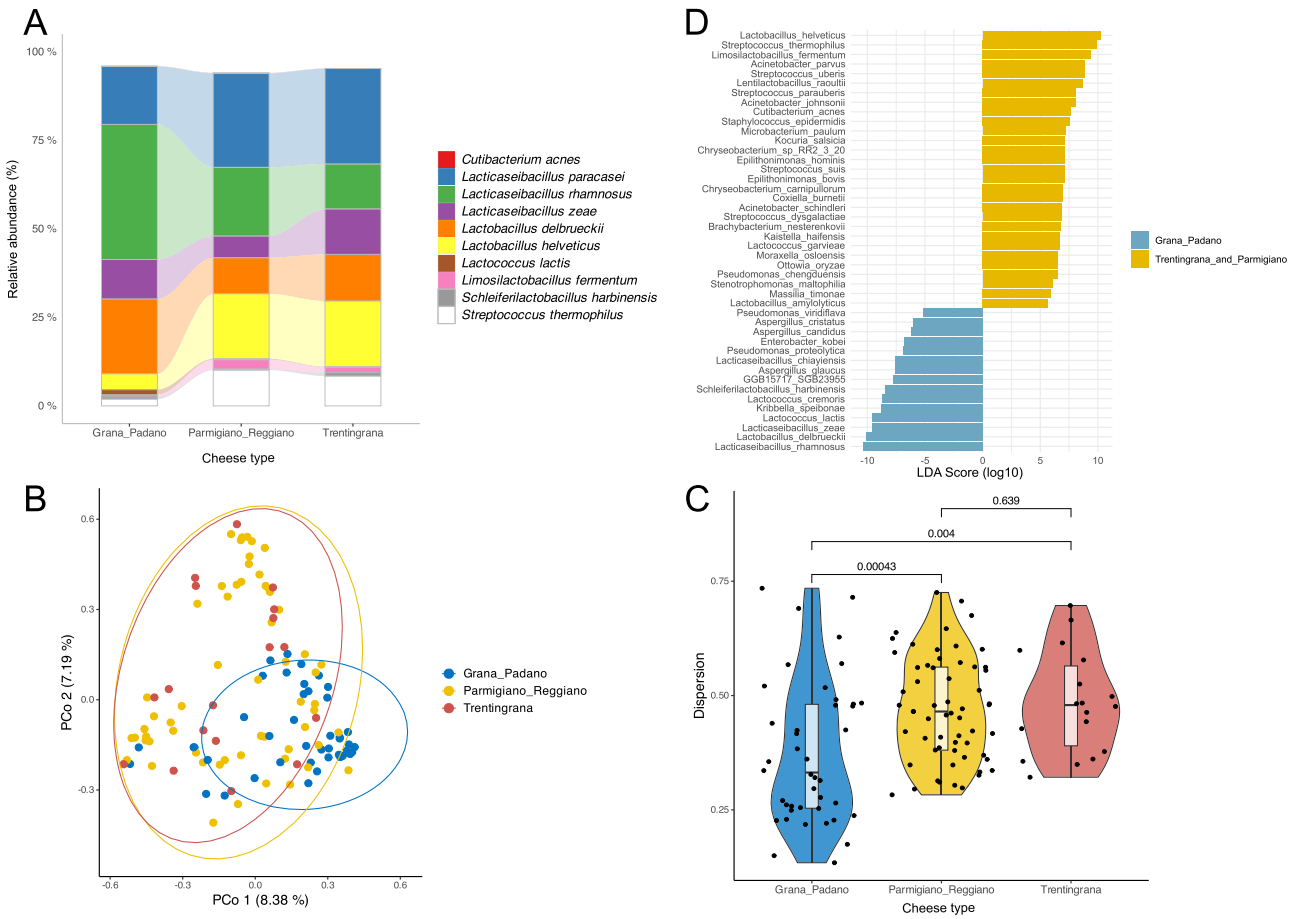

Fig. 1 | Taxonomic composition of the cheese microbiome. A Connected barplot of the top 10 species showing the highest median abundance in GP, PR, and TG cheeses. Average abundances of species in each cheese type are reported. **B** Principal Coordinate Analysis (PCoA) based on the pairwise Bray-Curtis distance computed on the microbiome taxonomic profiles. Each dot represents a cheese sample. Dots are color-coded according to the cheese type. The values in brackets indicate the percentage of variance explained by each principal coordinate. **C** Boxplots showing the beta dispersion (on the *y*-axis), i.e., the Bray-Curtis distance between each sample and the centroid of its group. Boxes represent the interquartile range (IQR) between the first and third quartiles, and the line inside represents the median. Whiskers indicate the lowest and the highest values within 1.5 × IQR from the first and third quartiles, respectively. Medians are compared between the groups through the Wilcoxon rank sum test. **D** Marker species of GP (in blue) or PR/TG (in yellow) groups according to the Linear Discriminant Analysis Effect Size (LEfSE) analysis. Values on the *x*-axis indicate the $\log_{10}$ of the Linear Discriminant Analysis score.

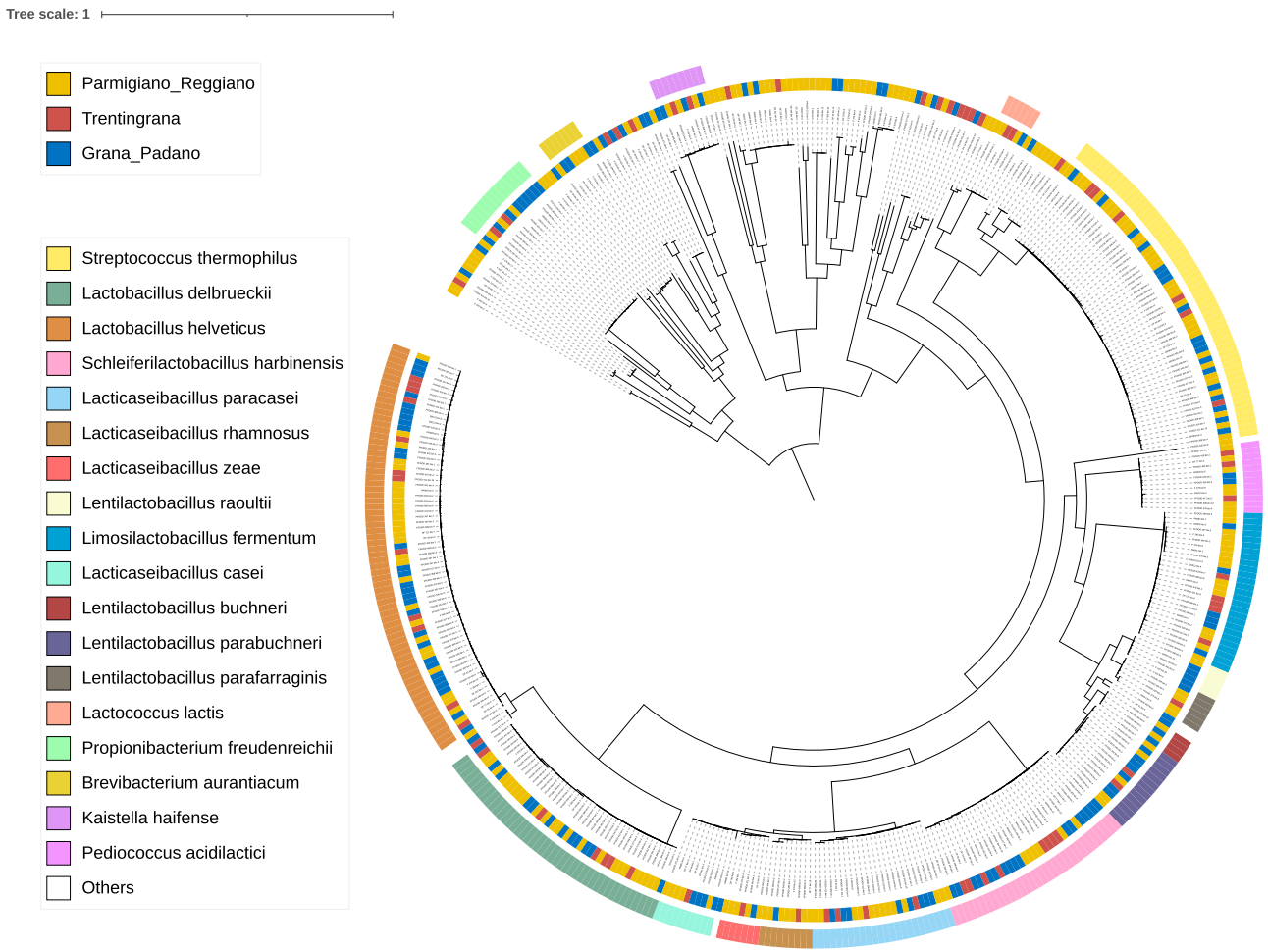

Fig. 2 | Taxonomic assignment of metagenome-assembled genomes (MAGs). Phylogenetic tree of all the 452 MAGs reconstructed from 120 samples of GP, TG, and PR cheeses. The phylogenetic tree was generated by aligning genes predicted from MAGs to a set of 120 bacterial marker genes. Taxonomic classification was performed using the release95 database of GTDB-Tk version 1.3.0. Rings are color-coded according to the assigned species (outer ring) and cheese type (inner ring).

25%), although its average abundance significatively varied between the three cheese types (GP = 38.17%, PR = 19.44% and TG = 12.76%; Kruskal–Wallis test p value < 0.001).

As expected, 7 out of the top 10 species belonged to the *Lactobacillaceae* family. *Streptococcaceae* were represented by *Streptococcus thermophilus* and *Lactococcus lactis* (average relative abundance 7.01% and 0.64%, respectively; Fig. 1A).

Although the Shannon and Simpson's alpha diversity indices showed no significative differences between the cheese types (Supplementary Fig. 1A), we found two clear clusters, one comprising both PR and TG, and the other including GP cheeses (Supplementary Fig. 1B). Interestingly, although GP and TG belong to the same PDO and TG can be considered as a niche production within GP PDO, TG microbiome was more similar to PR. The presence of these clusters was further confirmed by the pairwise Permutational MANOVA (PERMANOVA; p value < 0.05), although the beta dispersion was significatively lower in GP cheeses compared to PR and TG (Fig. 1B, C).

Interestingly, the Random Forest (RF) supervised machine learning model performed on species-level abundance data correctly discriminated between GP and the PR/TG groups (AUC-ROC 0.92, 95% CI: 0.89–0.95; Supplementary Fig. 1C), and LEfSe LDA scores suggested *Lactobacillus helveticus*, *Streptococcus thermophilus*, and *Limosilactobacillus fermentum* as main markers of the cluster comprising PR/TG, and *Lc. rhamnosus*, *Lb. delbrueckii* and *Lc. zeae* as a marker in GP (Fig. 1D). Coherently, the Shapley Additive exPlanations (SHAP) values analysis showed *Lb. helveticus* and *Streptococcus thermophilus* as the main features able to classify the PR/TG, while *Lc. rhamnosus* and *Lb. delbrueckii* were found to be characteristic features of GP (Supplementary Fig. 1D).

A total of 452 metagenome-assembled genomes (MAGs) were reconstructed from the 120 samples, which clustered into 71 Species-level Genome Bins (SGBs) (Fig. 2). The SGB with the highest number of MAGs was *Lactobacillus helveticus* (SGB 2, $n = 70$), followed by *Streptococcus thermophilus* (SGB 10, $n = 57$), *Lactobacillus delbrueckii* (SGB 7, $n = 42$), and *Schleiferilactobacillus harbinensis* (SGB 4, $n = 32$; Table 1). Notably, the distribution of the top 10 SGBs in the different cheeses was not statistically significant (χ^2 test p value > 0.05), although proteins from *Lacticaseibacillus zeae* and *Lactobacillus delbrueckii* previously highlighted as markers of GP were found with significatively higher LFQ intensities in this group of cheeses compared with PR/TG (PERMANOVA p value < 0.05; Supplementary Fig. 2A). The most abundant SGBs (with >10 MAGs) are reported in Table 1. Interestingly, 29.6% ($n = 134$) of the MAGs belonged to SGBs with <10 genomes. In addition, we identified 33 SGBs represented by only 1 MAG (Supplementary Data 1).

Despite the high abundance of *Lacticaseibacillus rhamnosus* in all the samples, we were able to reconstruct only nine MAGs from this species, probably due to the high strain heterogeneity found in all the samples, as confirmed by StrainScan (Table 2).

The microbiome metabolic potential is correlated with flavor development

A total of 53 volatile organic compounds (VOCs) were identified in the 116 cheeses. Hexanoic acid was the most abundant overall (avg. 11,770.7 ppb),

Table 1 | Distribution of MAGs from SGBs with ≥10 genomes in the cheese types

SGB	Taxonomic assignment	Overall, *N* = 452	GP, *N* = 153	PR, *N* = 228	TG, *N* = 71
SGB_2	*Lactobacillus helveticus*	70 (15%)	29 (19%)	29 (13%)	12 (17%)
SGB_10	*St. thermophilus*	57 (13%)	18 (12%)	33 (14%)	6 (8.5%)
SGB_7	*Lb delbrueckii*	42 (9.3%)	12 (7.8%)	24 (11%)	6 (8.5%)
SGB_4	*Schleiferilactobacillus harbinensis*	32 (7.1%)	17 (11%)	7 (3.1%)	8 (11%)
SGB_19	*Limosilactobacillus fermentum*	27 (6.0%)	6 (3.9%)	16 (7.0%)	5 (7.0%)
SGB_33	*Lacticaseibacillus paracasei*	24 (5.3%)	8 (5.2%)	12 (5.3%)	4 (5.6%)
SGB_5	*Lc. casei/zeae*	16 (3.5%)	4 (2.6%)	10 (4.4%)	2 (2.8%)
SGB_55	*Propionibacterium freudenreichii*	14 (3.1%)	9 (5.9%)	3 (1.3%)	2 (2.8%)
SGB_57	*Lentilactobacillus parabuchneri*	14 (3.1%)	7 (4.6%)	5 (2.2%)	2 (2.8%)
SGB_51	*Pediococcus acidilactici*	12 (2.7%)	2 (1.3%)	7 (3.1%)	3 (4.2%)

For each SGB, the number of MAGs and the proportion (%) are reported.

Table 2 | Average number of *Lacticaseibacillus rhamnosus* strains (±std. dev; as estimated by StrainScan) in GP, PR, and TG cheeses

# of *Lacticaseibacillus rhamnosus* strains	Overall, *N* = 103	GP, *N* = 35	PR, *N* = 52	TG, *N* = 16	*p* value
	6.0 ± 5.0	9.6 ± 5.2	4.3 ± 3.7	4.1 ± 4.0	<0.001

Lc. rhamnosus strain heterogeneity was compared between the cheese types through the Kruskal–Wallis rank sum test.

followed by butanoic and octanoic acids (6,470.1 and 3,498.9 ppb, respectively). (Supplementary Data 2). The Ward's linkage hierarchical clustering method based on the Pearson's correlation distance computed on the parts per million abundances of VOCs highlighted the presence of two clusters of cheeses, one mainly dominated by PR cheeses, and the other showing a significantly higher proportion of GP (χ^2 test p value < 0.05; Fig. 3A). The distribution of TG cheeses in the two clusters was not significantly different. Interestingly, the LEfSe analysis performed on the VOCs abundance of GP and PR highlighted 6 compounds as markers of PR cheeses (Fig. 3B).

We further estimated correlations between the abundance of species and the occurrence of VOCs in cheeses. We found several positive correlations between VOCs and some minor LAB (Fig. 3C). Indeed, *Lentilactobacillus raoultii* positively correlated with 2-butanol and benzaldehyde (Spearman's rank correlation coefficient 0.221 and 0.225, respectively; FDR-adjusted p value < 0.01), while *Lentilactobacillus parafarraginis* was positively linked to 1-butanol and isovaleric acid (Spearman's rank correlation coefficient 0.241 and 0.272, respectively; FDR-adjusted p value < 0.01). Notably, significant correlations were found between some of the most abundant species and VOCs. *Streptococcus thermophilus* was positively associated with fruity, green, and baked aroma given by 2-pentanol (0.226), 2-furanmethanol (0.267), phenylacetaldehyde (0.298), and ethylcaprate (0.302; Supplementary Data 2). Similarly, *Lacticaseibacillus rhamnosus* was associated with 1-hexanol and 2-octanone (Spearman's correlation coefficient 0.280 and 0.219, respectively), whereas *Lacticaseibacillus paracasei* and *Lactobacillus delbrueckii* were significatively and positively correlated with isovaleric acid (0.218) and hexanal (0.219), respectively (Fig. 3C).

The diverse cheese types differ in the prevalence and relative abundance of proteolytic and lipolytic genes

Metagenomes were screened for genes potentially involved in proteolytic activities, and therefore in flavor formation. In particular, we focused on 25 MEROPS subfamilies reportedly encompassing enzymes active on caseins and on shorter casein-derived peptides (Supplementary Data 3;) Collectively, 5288 CDS matched with the 25 MEROPS gene families (44.4 ± 17 genes per sample on average), with dipeptidase A (MEROPS family C69.001) accounting for 1097 matches (20.7% of total matches), followed by a proline-specific peptidase (family M24.006) with 505 occurrences (9.55%) and peptidase T M20.003 (493 occurrences, 9.32%; Supplementary Data 3).

Interestingly, 404 out of 452 MAGs (89.4%) harbored 3069 peptidase-encoding CDS (58% of the total CDS matching to the 25 MEROPS families), and MAGs from *Lb. helveticus* (SGB 2), *Lb. delbrueckii* (SGB 7) and *Lc. paracasei* (SGB 33) showed the highest proteolytic potential, encompassing 13.1 ± 0.6, 11.5 ± 3.8, and 10.8 ± 2.4 genes on average, respectively. Although the per-MAG number of protease-encoding CDS was significatively correlated with the MAGs' completeness (Supplementary Fig. 2B), we observed that the two species harboring the highest number of protease-coding genes showed a similar completeness of 81% on average (Supplementary Fig. 2C).

Differences in proteolytic potential between the species were not only explained by the number of protease-encoding CDS, but also according to their presence-absence pattern (Fig. 4A). Peptidase G (PepG; MEROPS family C01.089), peptidase W (PepW; MEROPS family C01.091), and PrtB proteinase (MEROPS family S08.118) were among the most discriminating families between the SGBs (χ^2 test p value < 0.05), being exclusively present in *Lb. delbrueckii* (SGB 7), while peptidase E (PepE, MEROPS family C01.088) showed a higher prevalence in *Lb. helveticus* (SGB 2; χ^2 test p value < 0.05). Consistently, we observed that both the RPKM level of proteases specific to *Lb. delbrueckii* (namely, PepG, PepW and PrtB) and the species abundance were significatively higher in GP cheeses compared to the others (Wilcoxon rank sum test p value < 0.05; Supplementary Fig. 3A), whereas PepE and *Lb. helveticus* were enriched in TG and PR cheeses (Wilcoxon rank sum test p value < 0.05; Supplementary Fig. 3B).

About 42% of the protease-encoding CDS (n = 2219) were placed on contigs not binned into MAGs. Taxonomic assignment of these contigs with kraken2 revealed the contribution of minor taxa for which MAGs were not reconstructed (Fig. 4B).

Although only 12 enzymatic features linked with proteolysis were found in cheeses metaproteome out of the 25 predicted in the metagenomes (Supplementary Data 4), both $\log_2$ LFQ intensities of proteins and RPKM abundance of genes consistently highlighted a separation in the proteolytic activity of the microbial communities according to the cheese type (PERMANOVA p value < 0.05; Fig. 4C and Supplementary Fig. 4A), with GP cheese samples clustering apart from PR and TG. This separation was driven by RPKM abundance of PepE (MEROPS subfamily C01.088) and PepN (M01.009), being more abundant in TG and PR cheeses, and by M20.003, M20.004, C01.089, and C01.091, identified as markers of GP cheeses (Supplementary Fig. 4B). Consistently. LFQ intensities of M20.003,

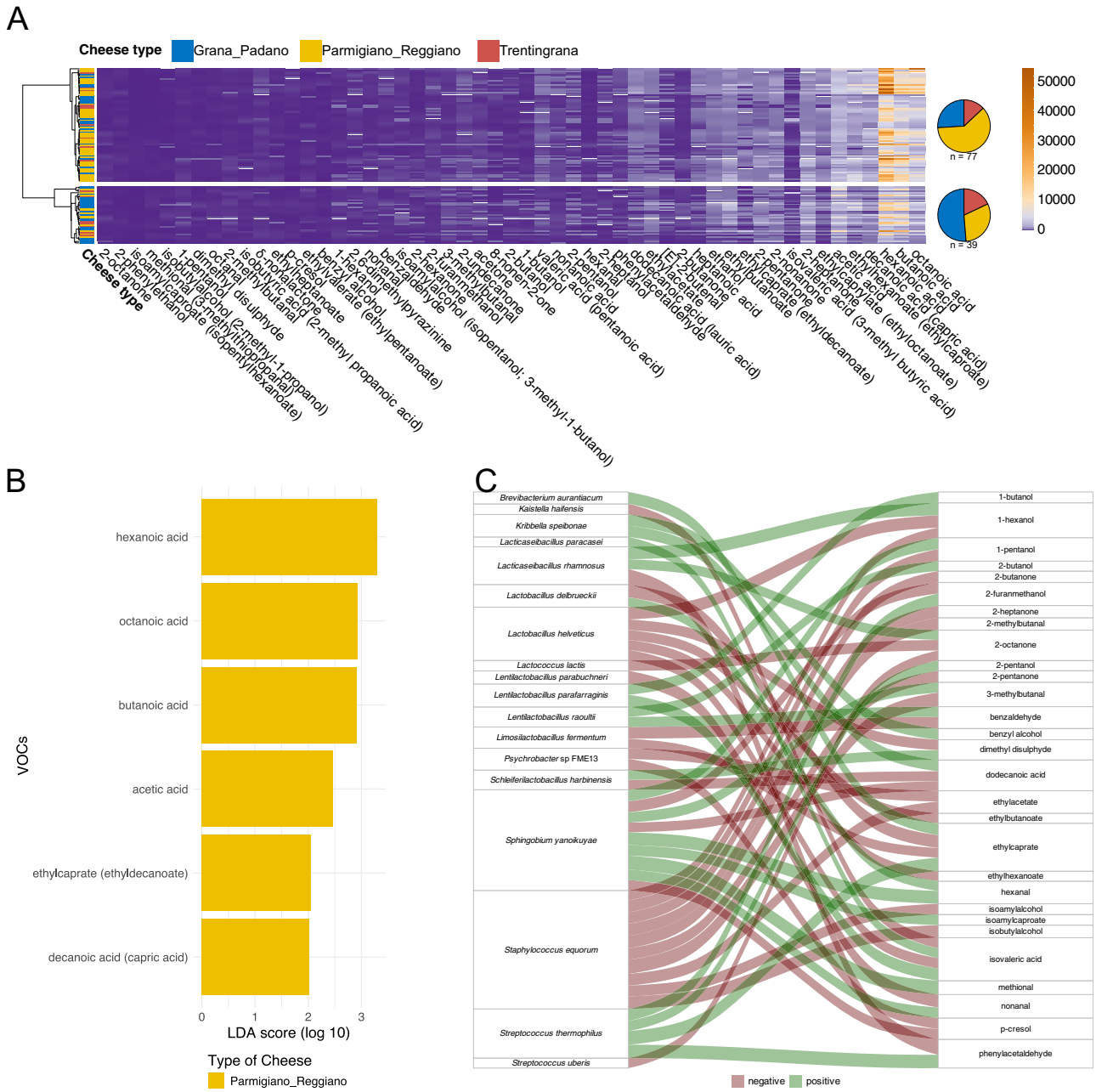

Fig. 3 | Volatile organic compounds (VOCs) distribution in GP, PR, and TG cheeses and their link with the microbiome. A Ward's hierarchical clustering based on Pearson's correlation index computed on the VOCs concentration (expressed as ppb = part per billion). Two main clusters are identified, and the right-side pie charts indicate the proportion of samples from each cheese type in the clusters. **B** Marker VOCs identified in the cheese types as detected by LEfSe analysis. **C** Alluvial plot showing correlations between the abundance of species and VOCs. Positive correlations are highlighted in green, whereas negative are in red. The thickness of each connector is proportional to the intensity of the correlation (Spearman's rank correlation coefficient).

M20.004, and M24.006 were markedly higher in GP cheeses (Supplementary Data 4).

We further screened the cheese metagenomes for the presence of genes related to lipolytic activities. Interestingly, lipase-coding genes further contributed to the separation between the cheese types. A total of 375 CDS matched with 16 genes previously involved in lipolysis (i.e., including homologs of the genes EstA, EstB, and EstC, coding for different esterases; Supplementary Data 5), with 3.4 CDS per sample on average. Similarly to what was observed with the proteolytic enzymes, the pairwise PERMANOVA highlighted a clear separation between the cheese types, with GP clustering apart (Fig. 5A). The main contributors to this separation were genes mapping against an EstA homolog found in *Lb. helveticus*, as well as to two homologs of EstA and EstB from *Lc. casei* and *Lc. paracasei*, respectively (Supplementary Fig. 4C). Notably, none of the lipolytic genes were expressed at the time of sampling, suggesting that their activity was exhausted early during ripening.

Furthermore, ~57% ($n = 214$) of the CDS mapping against lipase-coding genes were found in 178 MAGs (39% of the total MAGs reconstructed; Fig. 5B). However, the remaining 46% of CDS encoding for esterases belonged to contigs not binned within MAGs. These contigs were mainly assigned to *Lacticaseibacillus* spp. (70.8%), showing the potential contribution of this genus to lipolysis in long-ripened cheeses.

Microbial metabolism in cheeses potentially leads to the production of neuroactive compounds

Predicted genes were aligned to an in-house database including enzymes involved in the synthesis and release of neuroactive compounds

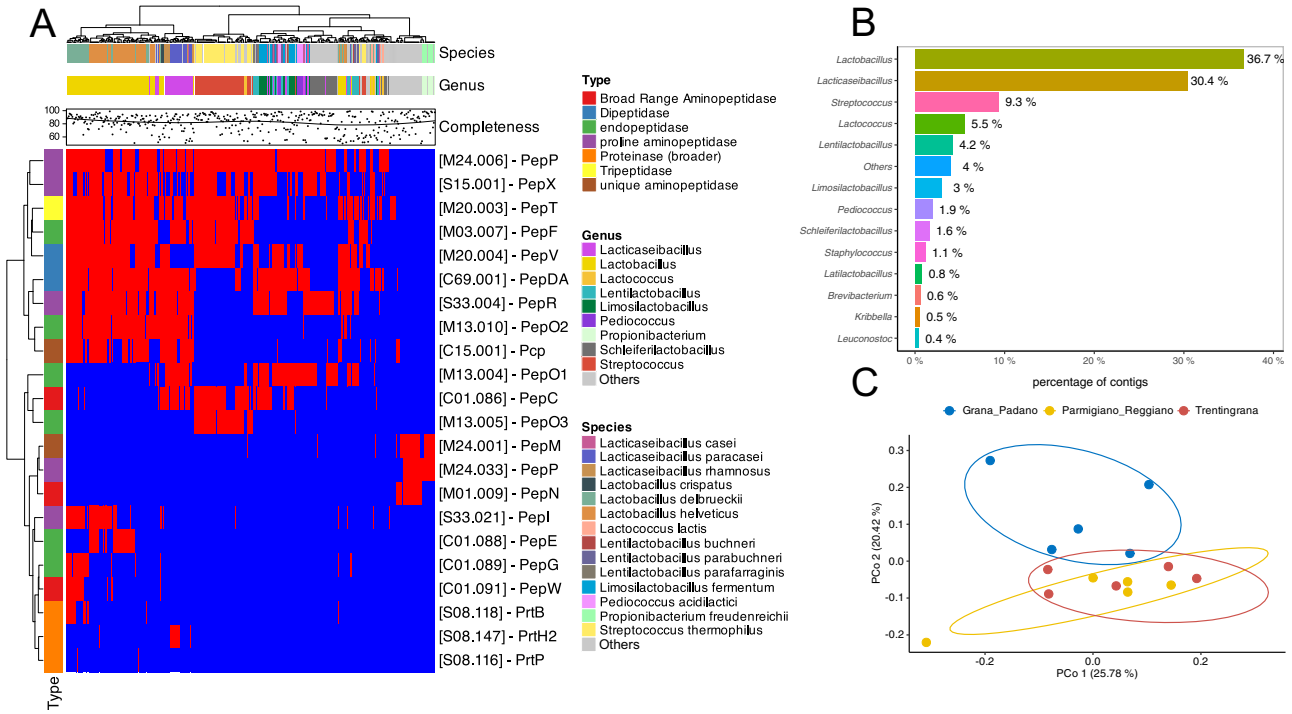

Fig. 4 | The proteolytic potential of GP, PR, and TG cheese microbiome. **A** Hierarchical clustering of the presence-absence profiles of proteolytic enzymes (on rows) active on caseins and their derivatives in MAGs from LAB species (on columns) based on the Jaccard distance metric. Top bars are colored according to the MAG genus and species, and the % completeness of each MAG estimated by CheckM is reported. **B** Bar chart showing the proportion of proteases-encoding contigs (%) not binned into MAGs. Contigs were taxonomically assigned to the genus level according to kraken2. **C** Principal Coordinate Analysis (PCoA) plot based on normalized Label-Free Quantification (LFQ) intensities of proteolytic enzymes. Samples are color-coded according to the cheese type. For each cheese type, 95% confidence ellipses are drawn assuming a multivariate t-distribution. The values in brackets indicate the percentage variance explained by each principal coordinate.

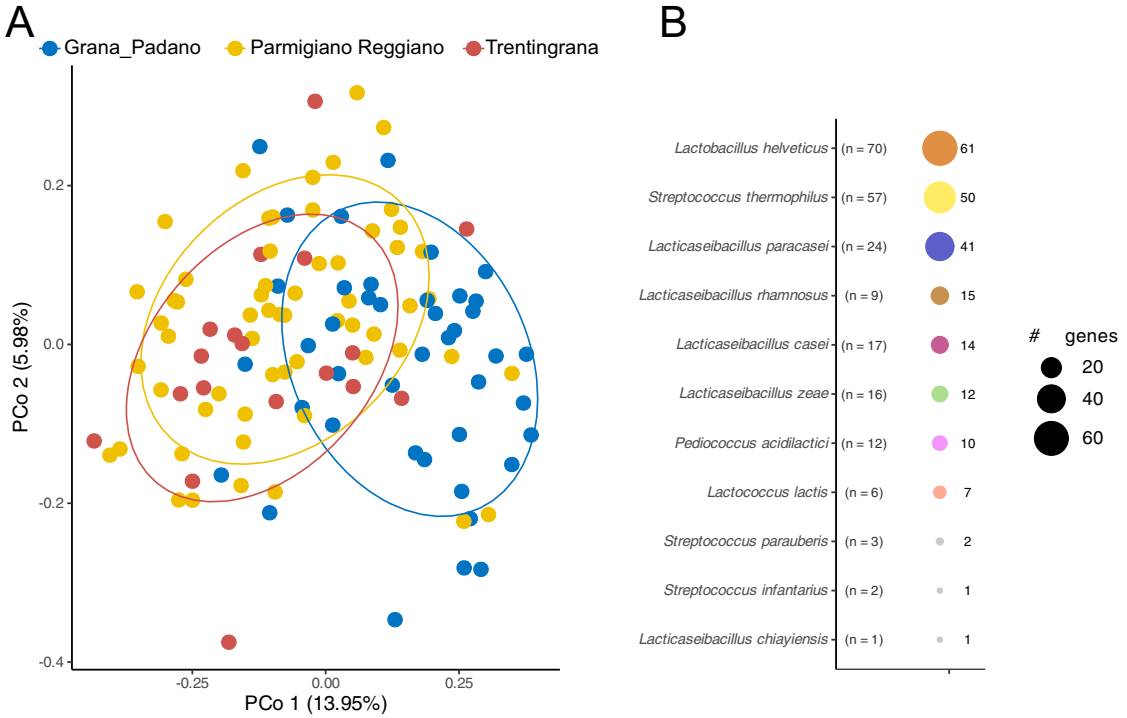

Fig. 5 | Lipolytic potential of GP, PR, and TG cheese microbiome. **A** Principal Coordinate Analysis (PCoA) based on the Aitchison distance computed on the RPKM abundance of 16 genes coding for lipolytic enzymes. The values in brackets indicate the percentage variance explained by each principal coordinate. **B** Bubble chart reporting the number of MAGs from each species encoding at least one of the genes involved in lipolysis. Bubbles size is proportional to the number of genes. Numbers in brackets indicate the total number of genomes reconstructed from each species.

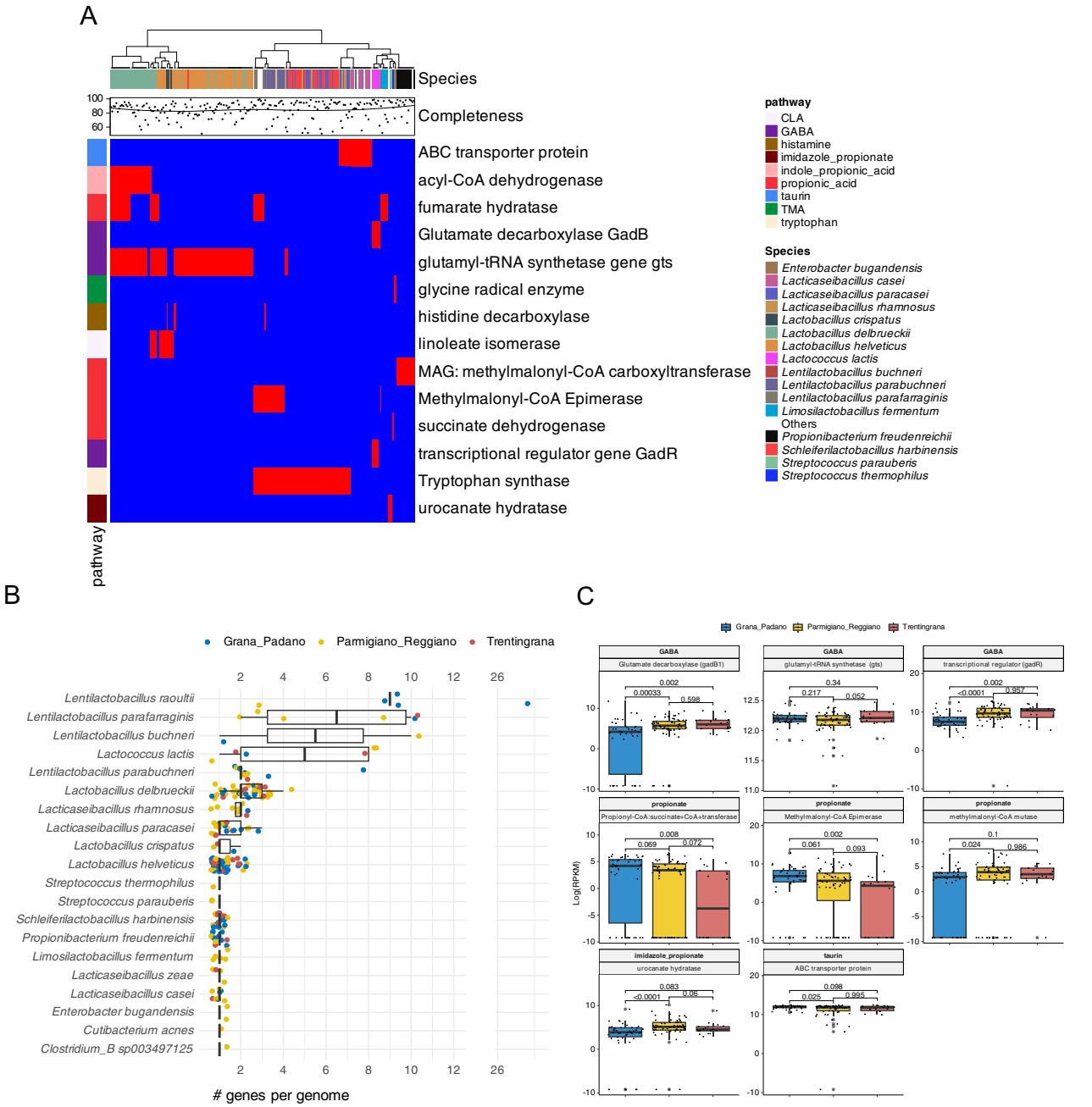

Fig. 6 | Presence of genes related to the biosynthesis of neuroactive molecules in the GP, PR, and TG microbiome. A Ward's linkage hierarchical clustering of presence-absence profiles of genes (on rows) linked with neuroactive compounds biosynthesis in MAGs (on columns). Red cells denote the presence of the gene. The bar on the left side shows the pathway where each gene is involved. On top, the scatter plot shows the % completeness estimated by CheckM for each MAG, while the bar denotes the taxonomic assignment of each MAG. **B** Boxplot showing the number of genes linked with neuroactive compounds production per genome. **C** Boxplots showing the Log(RPKM) abundance in the three cheese types of key genes linked with the production of GABA, propionate, imidazole propionate, and taurine. Boxes represent the interquartile range (IQR) between the first and third quartiles, and the line inside represents the median. Whiskers indicate the lowest and the highest values within 1.5 × IQR from the first and third quartiles, respectively. Medians are compared through the Wilcoxon rank sum test.

(Supplementary Data 6). Overall, 703 CDS mapped to 15 genes, with 169 of them (24%) matching with the tryptophan synthase, 141 (20.1%) with the glutamyl-tRNA synthetase gene, and 99 (14.1%) with the fumarate hydratase gene. Also, these CDS were included in 685 contigs, with 329 (~48%) binned into MAGs. In addition, the most contributing SGB in encoding genes linked with neuroactive compounds production was *Lb. delbrueckii* (SGB 7), with 94 genes and 2.2 genes/MAG on average, followed by *Lb. helveticus* (SGB 2), harboring 68 genes (~1 gene/genome involved in transportation of γ-aminobutyric acid (GABA)), although we did not find homologs of genes GadB and GadR, involved in synthesis of GABA and regulation of the operon, respectively. Interestingly, copies of GadB and GadR genes were exclusively found in genomes of *Lactococcus lactis* (SGB 37; Fig. 6A). In addition, despite their small SGB size (SGB 62, $n = 5$; SGB 164, $n = 6$; SGB 157, $n = 4$), we observed that three *Lentilactobacillus* species had the highest per-genome number of genes involved in the synthesis of neuroactive compounds (Fig. 6B).

We highlighted differences in the abundance of genes linked with the synthesis of neuroactive compounds between the cheese types. Consistent with our previous findings, we observed a grouping of TG and PR cheeses, which formed a cluster significantly different from GP

Table 3 | Proteins involved in pathways related to the biosynthesis of neuroactive molecules

Protein name	Final neurotransmitter	Type of cheese	ENA
Linoleate isomerase	Conjugated linoleic acids	Grana Padano	ABB43157.1
Glutamyl-tRNA synthetase	GABA	Grana Padano	KNE73827.1
Acyl-CoA dehydrogenase	Precursors	Parmigiano Reggiano	KRK23293.1
Fumarate hydratase	Precursors	Trentingrana	BBA81006.1
Tryptophan synthase	Serotonin	Trentingrana	PME03234.1

Table summarizing the matching proteins between data from metagenomic and metaproteomic datasets.

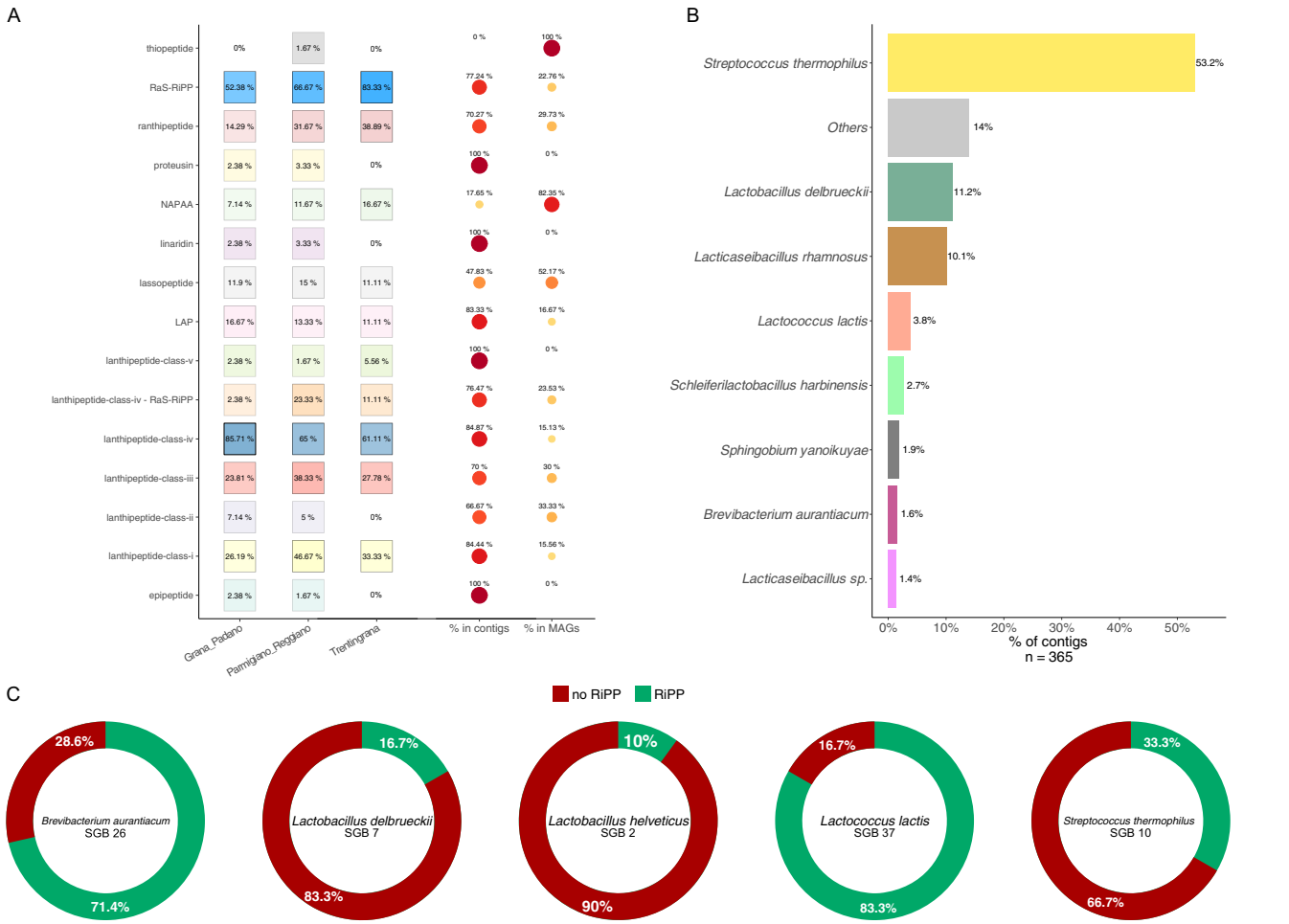

Fig. 7 | LAB species from each metagenome are the main encoders of RiPPs. **A** Heat plot showing the percentage of metagenomes from each cheese type where at least 1 gene cluster linked with the synthesis of RiPPs was found. For each RiPP, % of genes encoded in MAGs or contigs is reported. **B** The percentage of contigs encoding for RiPPs linked to each species is reported. **C** The percentage of MAGs encoding for RiPPs is reported for selected SGBs.

according to the RF algorithm (Pairwise PERMANOVA p value < 0.05; AUC = 0.82; 95% CI = 0.79–0.87, Supplementary Fig. 5A). The separation was driven by 6 genes with a different RPKM abundance between the cheese types showing an importance value above 25% (Supplementary Fig. 5B). In particular, gene GadB1 and GadR were significatively underrepresented in GP cheeses, while Propionyl-CoA:succinate-CoA transferase and Methylmalonyl-CoA Epimerase (both involved in the synthesis of propionate) were higher in GP cheeses (Fig. 6C) and identified as markers of this group of cheeses (Supplementary Fig. 5C).

Also, six proteins involved in the synthesis of neuroactive molecules, their precursors or intermediates, were found in both the metagenomic predictions and the metaproteome (Table 3), suggesting their activity in ripened cheeses.

Cheese metagenomes are enriched in genes linked with bio-protective activities

We detected 485 gene clusters linked with the biosynthesis of Ribosomally Synthesized and Post-Translationally modified Peptides (RiPPs) in 105 samples. The most frequently detected type was the radical S-adenosyl-L-methionine (RaS)-RiPP, with 145 gene clusters, followed by class IV, class I, and class III lanthipeptides (n = 136, 45, and 40, respectively). The χ^2 test did not detect different prevalence of RiPP types between the cheese types (p value > 0.05; Fig. 7A). More than 75% of the gene clusters were present on contigs not included within MAGs. About 53% of these were taxonomically assigned to *Str. thermophilus*, while 11.2 and 10.1% were linked to *Lb. delbrueckii* and *Lc. rhamnosus*, respectively. Notably, 51 gene clusters (~14% of the total) were linked to contigs from minor species (Fig. 7B). As expected, most of the contigs assigned to *Str. thermophilus* (n = 94) encoded for RaS-

RiPP, followed by class I and class IV lanthipeptides (37 and 32 contigs, respectively). On the contrary, *Lb. delbrueckii* and *Lc. rhamnosus* were exclusively linked to class III and class IV lanthipeptides.

Furthermore, 120 gene clusters (24.7% of the total) were embedded within 77 MAGs from several species, including *Str. thermophilus* (19 MAGs), *Lb. delbrueckii* (7 MAGs), *Lb. helveticus* (7 MAGs), *Brevibacterium aurantiacum* (5 MAGs), and *Lactococcus lactis* (5 MAGs). The percentage of MAGs from each species encoding for RiPP is reported in Fig. 7C.

Discussion

Cheeses marked with the PDO label have their unique sensory attributes, which result from the link of the traditional processing technology with the use of ingredients and raw materials from a strictly delimited geographical territory. Consequently, cheese-specific microbial consortia can be identified, leading to peculiar flavor characteristics[20,21]. In this work, we identified taxonomic, functional, and metabolic markers discriminating between GP, TG, and PR cheeses, three of the most famous Italian PDO long-ripened cheeses. According to PDO regulations, GP, TG, and PR share similar production techniques, including the use of cow milk, NWC for fermentation, calf rennet for coagulation, fine curd cutting, and curd cooking at approximately 53–56 °C. However, key differences distinguish these cheeses. Indeed, GP and TG utilize partially skimmed milk from two milkings, whereas PR combines skimmed evening milk with whole morning milk, resulting in a slightly higher fat content, possibly influencing the raw milk LAB community assembly, as well as the microbiome composition of cheeses over the ripening, as previously pointed out[10]. Feeding practices also vary: PR and TG cows are exclusively grass-fed, while GP allows silage in the diet. Also, the addition of up to 2.5 mg of lysozyme per 100 g of raw milk is allowed in GP to prevent *Clostridium tyrobutyricum* growth, but not in PR and TG. Ripening durations differ significantly, with PR aging a minimum of 12 months (often exceeding 24 months), TG typically 18–24 months, and GP at least 9 months, influencing texture and flavor complexity.

All the cheeses were dominated by mesophilic NSLAB (e.g., *Lacticaseibacillus rhamnosus*, *Lc. paracasei*). The proteolytic and lipolytic potential of these microorganisms favor their adaptation during the cheese ripening, when lactose is exhausted, thus explaining their dominance in long-ripened cheeses[7,13,22]. Unexpectedly, we observed that the microbiome composition of TG was more similar to PR than GP, probably reflecting the analogies in PR and TG production. Both high- and low-abundant species drove this separation. Indeed, *Lc. rhamnosus* and *Lc. zeae* (a species frequently isolated from dairy environment showing strong phylogenetic similarity to *Lacticaseibacillus casei*[23]) were highlighted as taxonomic markers of GP, whereas two thermophilic LAB, namely *St. thermophilus* and *Lb. helveticus*, were significatively more abundant in PR and TG. Thermophilic LAB are active during the first days of production of long-ripened, hard cheeses[13], and their lower abundance in ripened GP cheese might reflect their depletion during the first stages of cheese production, possibly due to the lysozyme addition. The use of this enzyme in GP is a solid strategy to prevent late blowing defects due to clostridia[24,25], but it might inhibit some LAB groups or strains. Although lactobacilli are generally considered tolerant to lysozyme[26], several studies showed a strain-dependent sensitivity[27,28]. Indeed, Bassi et al.[29] highlighted that GP cheeses processed with lysozyme had a higher proportion of *Lb. delbrueckii* but lower levels of *Lb. helveticus*, consistently with our results. Therefore, the technological differences in PDO cheesemaking might be at the basis of the unexpected similarity of TG with PR instead of GP. Importantly, we did not find spoilage or pathogenic bacteria, possibly thanks to the adverse conditions of the cheese itself (e.g., low water activity and concentration of free sugars) and to signals of bacteriocins synthesis by *St. thermophilus* and other SLAB probably occurring in the first production stages of cheeses.

The lipolytic and proteolytic potential was reflected in the taxonomic composition. For example, the higher abundance of PepE in TG and PR was consistent with the distribution pattern of *Lb. helveticus*, that was the sole species harboring this gene. Similarly, GP cheese metagenomes were dominated by *Lb. delbrueckii* and by its related proteases such as PepW, which showed the highest efficiency in releasing free amino acids in a cheese model[30]. The different distribution and abundance of proteinases and peptidases in the cheese types, and their uninterrupted and even enhanced activity after the SLAB cell autolysis[31] might lead to different peptidic and aminoacidic profiles connected with peculiar sensorial traits, as well as with the presence of different bioactive compounds in the cheese. Indeed, the proteome and volatilome analyses corroborated this idea, since the VOC profiles mirrored the sample clustering obtained using the microbiome data, while about half of the metagenome-predicted proteolytic enzymes were found in the proteome. It is widely reported that NSLAB are the main contributors to flavor development in ripened cheeses, with strain-specific variations: Bancalari and colleagues[32] highlighted the distinct contribution of *Lc. casei* and *Lc. paracasei* strains in releasing a wide range of aromatic compounds, while Sgarbi et al. reported the different releases of 1-hexanol (associated with a pleasant fruity/floral note[32,33]) by strains of *Lc. casei* and *Lc. rhamnosus* isolated from PR. Furthermore, non-starter strains of *Lc. casei* and *Lc. rhamnosus* were shown to increase the production of compounds derived from amino acid catabolism that play an important role in the aroma of aged hard cheeses, such as 3-methylbutanal and 3-methylbutanol in model cheeses[34]. Consistently, we observed that most of the odorant compounds were positively correlated with species belonging to *Lacticaseibacillus* and *Lentilactobacillus*, although we were not able to better describe the strain-level potential in determining the aromatic bouquet of cheeses. However, the final flavor composition of ripened cheeses might also be influenced by other factors, such as the activity of enzymes released by lysed cells, as previously shown[35]. The activity of enzymes from both alive and lysed cells cannot be estimated by a pure metagenomic approach, suggesting the relevance of integrated multi-omics analyses to decipher complex interactions and mechanisms leading to a specific aroma.

Besides providing specific properties that contribute to award products with the PDO marks, food fermentation further contributes to the release of bioactive compounds, improving human health. Some studies have displayed the immunomodulatory role of molecules detected in ripened cheeses[18,36], while others have reported the presence of phosphopeptides resulting from proteolysis, which enhance the absorption of minerals[15], and of microbial-derived vitamins[37]. Furthermore, some microbial metabolites produced using amino acids as precursors may explicate an activity on the nervous system. Indeed, an intervention study using a psychobiotic diet rich in prebiotic and fermented foods significantly influenced the participants' metabolome, increasing the concentration of tryptophan-derived metabolites including L-tryptophan, tryptamine, quinolinic acid, indole-3-acetic acid, 5-hydroxyindole-3-acetic acid, and 3-hydroxykynurenine, all of which are closely linked to neuroactive functions, suggesting an influence on excitatory neurotransmission and neuroinflammatory processes[38]. Additionally, Moreira et al.[39] found several neuroactive compounds, such as agmatine and serotonin, accumulating over cheese ripening. Our results showed a high abundance of genes linked with the biosynthesis and transport of neuroactive compounds, including GABA, propionate, and conjugated linoleic acid (CLA), suggesting that the microbial communities inhabiting the long-ripened GP, PR, and TG cheeses might release these compounds in the cheese matrix over the ripening, making them bioavailable for consumers. In cheese, proteolysis of caseins can release L-glutamate, which may then serve as a substrate for GABA production by LAB[40,41]. The main actors potentially involved in the biosynthetic pathways of CLA, GABA, taurine, and propionate in our samples were NSLAB, including *Lentilactobacillus* spp., that showed the highest number of genes per genome, although a minor fraction of the genes were also found in *Lb. delbrueckii*, *Lb. helveticus* and *Lactococcus lactis*, with the latter encoding for enzymes linked with both regulation and synthesis of GABA (Fig. 6A). Our results are consistent with previous works, since Park and Oh[42] found a high ability to produce GABA in *Lentilactobacillus buchneri* strains from naturally ripened cheeses. Moreover, a *Lentilactobacillus buchneri* strain had a neuroprotective effect on PC12 cells when exposed to agents that can induce neuronal cell death[43]. Similarly, the GABA biosynthetic potential of strains of *Lactococcus lactis* during fermentation in dairy environments (traditional

cheeses and kefir) has been described[44], which relies on the expression of the *gadCB* operon[45]. GABA has an important function in the enteric nervous system (ENS) in terms of influencing gut functions such as acid secretion, gastric emptying, intestinal motility, and pain perception[46]. Interestingly, GABA produced by a *Lactococcus lactis* strain in a mouse model was associated with reduced visceral pain, helping to manage gastrointestinal disorders (i.e., IBS) correlated to stressful/anxious profiles[47].

Previous works suggested that SLAB may also play an active role in the synthesis of neuroactive compounds. Indeed, a *Lb. helveticus* strain isolated from a fermented dairy product ameliorated stress tolerance and brain performances in an animal study[48]. Although ripened cheeses consumption has already been linked to improved mental health[18], our metagenomic and metaproteomic data underline that PR, GP, and TG might be an under-explored source of potential psychobiotic strains with the potential to enhance human mental health, possibly linking Italian PDO cheese consumption to the prevention of mental health issues. Furthermore, the abundance of genes linked with the production of neuroactive compounds or their precursors effectively discriminated between the cheese types, probably reflecting the species-level composition of the cheeses, thus representing a possible additional trademark of PDO cheeses.

Overall, more than half of the enzymes predicted from metagenomics reads and involved in neuroactive precursor biosynthesis were identified through proteomic analysis, suggesting active expression of these functional proteins in the ripened cheese microbiome, albeit with varying intensities. Particularly, the presence of linoleate isomerase is noteworthy. This enzyme catalyzes the conversion of linoleic acid to CLA, which has demonstrated anti-inflammatory and neuroprotective properties, potentially mitigating neuroinflammatory conditions and supporting overall brain health[49], while dietary intake of cis-9, trans-11 CLA reduced amyloid β-protein accumulation and suppressed inflammation in Alzheimer's disease models, suggesting cognitive benefits[50]. GP samples exhibited notable levels of fumarate hydratase and tryptophan synthase. Fumarate hydratase has been implicated in cellular processes related to energy production, metabolic regulation, and oncosuppression activities[51], while tryptophan synthase is essential for the biosynthesis of tryptophan, a precursor to serotonin, which plays a crucial role in mental well-being[52]. The acyl-CoA dehydrogenase initiates the β-oxidation of fatty acids, leading to the production of acetyl-CoA, a vital substrate for the synthesis of several neurotransmitters, including acetylcholine, which is essential for memory and learning. Therefore, the presence of acyl-CoA dehydrogenase suggests active lipid metabolism, which may facilitate the production of lipid–derived precursors necessary for neurotransmitter biosynthesis in addition to flavor compounds.

Moreover, as previously suggested[53–55], our results highlight the impact of cheesemaking parameters, such as the use of lysozyme in GP and the different skimming procedures, on the starting assembly of the microbial community in GP/PR, through the selection of either *L. helveticus* or *L. delbrueckii*. Also, our results show the crucial role that the microbiome plays during cheese ripening in determining cheese-specific flavors and characteristics. This contribution may be explicated by live microorganisms, but also by microbial enzymes released by microorganisms no longer viable[10,31,56]. At the same time, our results support previous works highlighting the potential role of microbial metabolism in providing fermented foods with health benefits, and suggest that Italian PDO cheeses might represent a source of potential psychobiotic strains to be exploited in the production of functional foods beneficial for mental health. In addition, these data fire up the questioning about the interaction patterns that occur in the cheeses, underlying the need to focus more on this task to better understand the interactions between microorganisms and how to exploit them to obtain products with desired properties and free of spoilage or pathogenic microorganisms.

However, the main limitation of this study is that we analyzed products from the market, so we did not have access to the specific information about the processing technology and parameters, which were assumed from the PDO disciplinary. This makes it difficult to assess the link between the cheesemaking procedure and the specific microbiome dynamics observed. Future longitudinal studies on PDO cheese microbiome might better elucidate the role of cheesemaking parameters on microbiome evolution and activities.

Methods

Samples collection, DNA extraction, and purification

PDO GP ($n = 42$), TG ($n = 18$), and PR ($n = 60$) cheese samples were collected either from local markets and shops. Samples' metadata are reported in Supplementary Data 7. Samples were stored at 4 °C until their arrival at the laboratory. There, about 10 g of cheese were suspended 1:10 in Phosphate Buffered Saline solution (PBS, pH 7.4) and homogenized using a peristaltic homogenizer (IUL, Barcelona, Spain). The liquid phase was collected in 50 mL centrifuge tubes (Sarstedt, Nümbrecht, Germany) and centrifuged at 6500 × *g* for 15 min. The cellular pellet was washed twice with fresh PBS and stored at −80 °C. DNA was extracted by suspending the cellular pellet with 250 µL of 4 M guanidium thiocyanate and 500 µL of 5% sodium lauroyl sarcosinate solution. The solution was transferred to 2 mL centrifuge tubes (Sarstedt, Nümbrecht, Germany) containing sterile 0.1 mm silica beads and incubated at 70 °C for 1 h at 700 rpm in a ThermoMixer C (Eppendorf, Hamburg, Germany). Subsequently, the samples were mixed for 20 min using a Vortex Genie 2 (Scientific Industries Inc., New York, United States) and RNA was degraded by incubation of the supernatant with 10 µL of 20 mg/mL RNase A (Thermo Fisher Scientific, Waltham, Massachusetts, United States) at 37 °C for 20 min.

The lysate was purified using the NucleoSpin gDNA Clean-up Kit (Macherey-Nagel, Dueren, Germany) according to the manufacturer's instructions. The purified DNA was eluted in 50 µL of DE Buffer (5 mM Tris/HCl, pH 8.5) and quantified using the Qubit dsDNA High Sensitivity quantification kit (Thermo Fisher Scientific, Waltham, Massachusetts, US). Finally, Whole Metagenome Sequencing was performed using an Illumina NovaSeq sequencer (Illumina, San Diego, California, US), leading to 2 × 150 bp reads.

Bioinformatic analysis

Low-quality metagenomic reads were filtered out using Prinseq lite[57] (version 0.20.4; parameters "-trim_qual_right 5" and "-min_len 60"). The taxonomic composition of metagenomes was estimated with MetaPhlAn[58] (version 4.0.6; database version mpa_vOct22_CHOCOPhlAnSGB_202212), and all the taxa were retained for further analysis. High-quality reads were assembled into contigs using MEGAHIT[59] (version 1.2.2), and contigs < 1000 bp were discarded. Contigs were taxonomically labeled using kraken2[60] and the "PlusPF" database (01/02/2024 release; available at https://benlangmead.github.io/aws-indexes/k2). Presence of Ribosomally Synthesized and Post-translationally modified Peptides (RiPPs) was predicted with antiSMASH[61] (version 7.1.0, option "–hmmdetection-strictness strict").

To reconstruct MAGs from the samples, a previously developed pipeline was used[62]. For each sample, reads were aligned against the contigs using Bowtie2[63] (version 2.2.9; options "–very-sensitive-local" and "–no-unal"), and the resulting alignment files were provided to MetaBAT[64] (version 2.12.2) to reconstruct metagenomic bins. Completeness and contamination of each bin were estimated through CheckM[65] (version 1.0.13; pipeline "lineage_wf"). MAGs showing completeness ≥50% and contamination <5% (i.e., medium/high quality bins) were clustered into SGBs as previously reported[62]. The taxonomy of each MAG was assessed through GTDB-Tk[66] (version 1.3.0; "classify_wf" pipeline) using the release89 database. Furthermore, the alignment of genes predicted from each MAG to a set of 120 bacterial marker genes was used to construct a phylogenetic tree of all the genomes. The resulting phylogenetic tree was visualized in iTol[67].

In spite of the high abundance of *Lacticaseibacillus rhamnosus* in the metagenomes, only a few MAGs from this species were obtained, possibly due to a high strain heterogeneity that led to filter out the reconstructed genomes. To estimate the *Lc. rhamnosus* strain heterogeneity in each metagenome, we fed the StrainScan algorithm[68] (version 1.0.14) with all the NCBI RefSeq genomes from the species (downloaded in July 2024; Supplementary Data 8) through "strainscan_build" (options -k 31 -t 20 -u

100000 -e 1), then we estimated the number of strains present in each metagenome through the main command, with default options.

Contigs from each metagenome were processed with MetaGeneMark[69] (version 3.26) in order to detect coding sequences (CDS). Predicted genes were mapped against the MEROPS protease database[70] (release 12.1) with diamond blastx[71] (version 2.0.4; options "–evalue 0.00001" and "–very-sensitive"), and only matches showing >50% identity over 50% of the query sequence were retained. A non-redundant set of protease-encoding genes was obtained by clustering the positive hits at 95% of similarity with CD-HIT[72] (version 4.6.8), then the number of reads mapping against the set of core genes with BowTie2 was normalized using the Read per Kilobase per Million mapped reads (RPKM) method[73]. To estimate the presence and the abundance of lipolytic enzymes and proteins involved in the biosynthesis of neuroactive compounds, we applied the above approach, mapping genes against a custom database (Supplementary Data 4 and 5).

Proteomic analysis

Metaproteomic analysis was performed on a subset of 15 cheeses (GP = 5; PR = 5; TG = 5). Eleven g of cheese samples were diluted in 99 mL of sodium citrate water (2% w/V) preheated at 40 °C. The diluted samples were then homogenized for 2 min at speed 4 using the Omni Mixer Homogenizer (Omni International, Waterbury, CT, USA). The homogenized mixture was left to settle for 2 min, then 10 mL of supernatant was collected into 15 mL tubes and centrifuged at 8000 × *g* for 10 min at 4 °C. Then, the supernatant was discarded, and the pellet resuspended with 1 mL of citrate water into a new 1.5 mL tube. Successively, the suspension was centrifuged at 13,200 × *g* for 3 min at room temperature. The supernatant was then discarded, while the cell pellet was washed twice with 1 mL of citrate water and centrifuged at 13,200 × *g* for 3 min at room temperature. The cell pellet was then prepared for LC-MS/MS analysis using the protocol described previously[74]. Briefly, the cell pellets were resuspended in 150 µL lysis buffer (100 mM Tris HCl, 4% SDS, 10 mM DTT), vortexed, and subjected to cell lysis using 0.2 g of glass beads (425–600 µm—30–40 U.S. sieve—SIGMA, Life Science) in a bead beating system (FastPrep-24™, MP Biomedicals, Irvine, California, USA), three rounds, each at 1600 rpm for 50 s. After that, the samples were sonicated at 240 W for 3 min (Bransonic Ultrasonic cleaner, Model 2510E-MTH, Danbury, CT, USA). Protein extraction was performed according to the methodology reported in ref. 74. The peptides were resuspended in a loading buffer (2% ACN and 0.05% TFA), sonicated for 10 min, and the protein concentration was adjusted to 0.2 mg/mL before instrumental analysis. A nano UPLC (nanoElute, Bruker) coupled with a trapped ion mobility spectrometry/quadrupole time of flight mass spectrometer (timsTOF Pro, Bruker) was used according to the methodology described previously[74].

The raw LC-MS/MS data were processed using MaxQuant software 2.0.3.1. Label-free quantification (LFQ) algorithm (false discovery rates were set to 0.01 at peptide and protein levels), and additional results were filtered (minimally one peptide was necessary for protein identification). The protein database used was generated from all the MAGs obtained from the metagenome sequencing (as described above). Missing data were inputted using random draws from a Gaussian distribution centered around a minimal value to control for proteins missing not at random. Filtered proteinGroups normalized (LFQ) intensities from MaxQuant were further used for statistical testing.

Volatilome analysis

Volatile compounds (VOCs) were extracted using solid phase microextraction. Cheese samples, stored at −20 °C, were finely grated, and 25 g of the sample was placed in a 100 mL glass bottle with a magnetic stirring bar. To this, 6.25 g of sodium phosphate (NaH_2PO_4; Sigma-Aldrich), 25 mL of distilled water, and 50 µL of 2-methyl-3-heptanone (purity 99%, Sigma-Aldrich, St. Louis, MO, USA), used as an internal standard (560 ppm in water solution), were added. The sample was conditioned at 50 °C for 10 min without stirring, followed by magnetic stirring (150 rpm) for 20 min at the same temperature[75].

VOCs adsorption was performed by introducing a 50/30 µm divinylbenzene/carboxen/polydimethylsiloxane (DVB/CAR/PDMS) 2 cm fiber (Supelco, Bellefonte, PA) into the headspace of the bottle and exposing the polymer for 30 min at 40 °C while stirring. The VOCs were then desorbed directly into the GC inlet port, maintained at 250 °C, with a 4:1 split ratio for 10 min.

The GC/MS analysis was performed using an Agilent 7890A GC System gas chromatograph coupled to an Agilent 5975C VL MSD with Triple-Axis-Detector mass spectrometer (Agilent Technologies, Inc., Palo Alto, CA, USA). The GC was equipped with a Zebron ZB-WAX capillary column (60 m × 0.25 mm i.d. × 0.25 µm; Phenomenex, USA). Helium was used as the carrier gas at a flow rate of 1 mL/min. The temperature program started at 40 °C for 10 min, followed by a 5 °C/min increase until 240 °C, where it was held constant for 11 min[75]. Mass spectra were recorded at 70 eV, and the source, quadrupole, and interface temperatures were set at 230 °C, 150 °C, and 250 °C, respectively.

VOCs were identified by comparing retention times and mass spectra obtained from pure reference compounds analyzed under the same conditions. All chemical standards were purchased from Sigma-Aldrich (St. Louis, MO, USA). Identification was confirmed by comparing mass spectra with those in the National Institute of Standards and Technology (NIST) database. The fiber was conditioned at 270 °C for 1.5 h before analysis, and a blank test was performed before each analysis. Quantitative data for the volatile compounds in the sample were obtained by normalizing the peak areas of each compound to the peak area of the internal standard, using MSD ChemStation 5975 TAD Data Analysis software (Agilent Technologies, Palo Alto, CA, USA).

Data analysis, multi-omics integration, and data visualization

Statistical analysis was performed in the R environment (https://www.r-project.org). The normality of variables was tested with the Shapiro-Wilk test through the "shapiro.test" function from the "base" R package. The Wilcoxon's rank sum test ("wilcox.test" function from the "base" R package) was used to compare the abundance of variables not normally distributed (i.e., abundance of species and RPKM values of genes) between the cheese types. Significance of results was set at p value < 0.05. P values were adjusted to account for multiple comparisons with the Bonferroni method when needed.

The "vegdist" function from the "vegan" R package was used to compute the distance matrices between the samples with different metrics, whereas the function "diversity" estimated the Shannon and Simpson's diversity indices. The distance matrices resulting from the "vegdist" were provided to the "prcomp" function from the "base" R package to perform Principal Coordinate Analyses (PCoAs). Statistically significant differences in the distance metrics between the groups were highlighted through the Permutational Multivariate ANOVA ("adonis2" function from the "vegan" R package).

The chi-squared test ("chisq.test" function from the "base" R package) highlighted differences in the prevalence of genes between the cheese types.

Correlations between the RPKM values of genes and the abundance of microbial taxa were computed with Spearman's rank correlation coefficient through the functions included in the "corrr" R package.

Metagenome- and MAG-driven LFQ intensities were normalized and log2-scaled through the function "normalize_vsn" from the "DEP" R package. Normalized intensities of protein groups were compared between groups of samples or MAGs through the Wilcoxon rank sum test, as reported above.

All the heatmaps were produced with the "ComplexHeatmap" R package, whereas "ggplot2" was used for scatter plots, barplots, pie charts, and violin plots. The Linear Discriminant Analysis Effect Size (LEfSe) analysis was obtained through the *lefser* R package on microbial taxa and VOCs datasets using the type of cheese as a grouping variable (https://github.com/waldronlab/lefser).

To classify cheese types based on microbial profiles and gene abundances, a RF supervised machine learning model was implemented using the "caret" R package. The model training employed repeated cross-

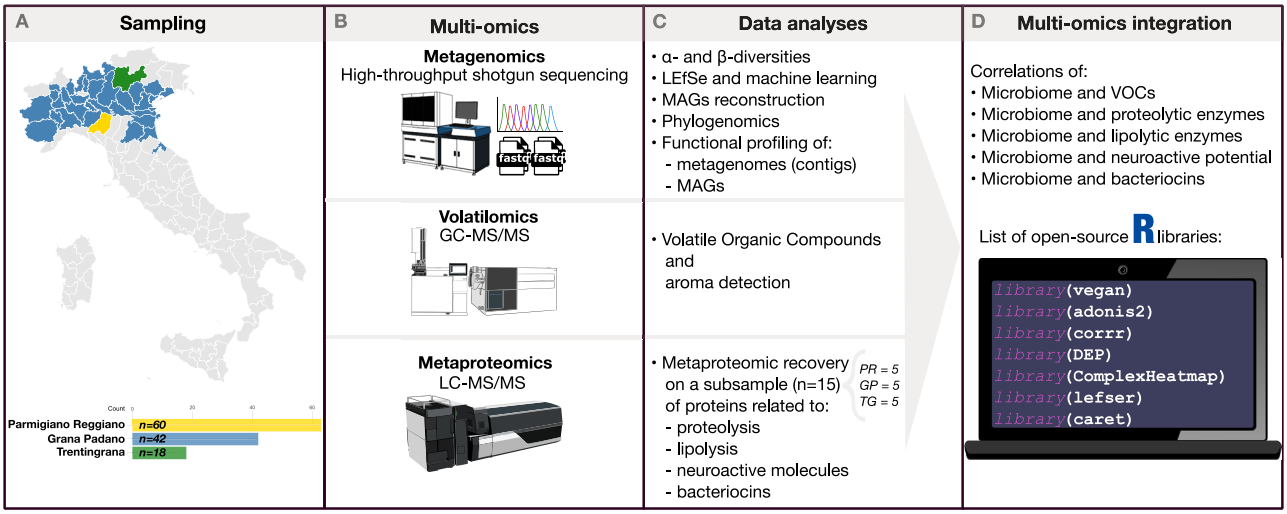

Fig. 8 | Multi–omics workflow used for characterizing Italian PDO hard cheeses. The workflow used in the study combined metagenomics, volatilomics, and metaproteomics to explore the microbiome and functional potential of three Italian PDO cheese types: Parmigiano Reggiano (PR, $n = 60$), Grana Padano (GP, $n = 42$), and Trentingrana (TG, $n = 18$). **A** All collected cheese samples were PDO labeled and were purchased in their respective production area (according to PDO disciplinaries) in Northern Italy, with sample distribution shown on the map and counts summarized in the bar chart. **B** The multi-omics data generation included shotgun metagenomic sequencing, volatilomics analysis via gas chromatography-mass spectrometry (GC-MS/MS), and metaproteomics analysis using liquid chromatography-mass spectrometry (LC-MS/MS) for protein identification. **C** Data analyses including computation of α- and β-diversities, application of Linear discriminant analysis Effect Size (LEfSe) and machine learning algorithms, metagenome-assembled genomes (MAGs) reconstruction, phylogenomics, and functional profiling of both metagenomes (contigs) and MAGs. Then, volatile organic compound detection was computed to detect the aroma of cheeses. Lastly, metaproteomics was performed on a subsample ($n = 15$) and proteins related to proteolysis, lipolysis, neuroactive molecules, and bacteriocins across three cheese types (PR = 5, GP = 5, TG = 5). **D** The multi-omics integration was performed through correlations across microbiome, volatilome, and metaproteome datasets using R open–source packages (e.g., *vegan, adonis2, corrr, DEP, ComplexHeatmap, lefser, caret*), supporting the integration of microbial, functional, and biochemical features. Icons from panel B by Maxime-U-Garcia (https://maxulysse.github.io/) and DBCLS (https://togotv.dbcls.jp/en/pics.html) are licensed under CC-BY 4.0 Unported https://creativecommons.org/licenses/by/4.0/.

validation (10-fold, 3 repeats) via "trainControl" on all cheese samples, so each sample was used in both training and validation sets. Hyperparameters were tuned by grid search to ensure robustness and prevent overfitting. The performance of the model was evaluated using the Receiver Operating Characteristic (ROC) curves, using the Area Under the Curve (AUC) as the primary metric. Furthermore, supervised machine learning model explainability was performed both employing variable importance scores and SHAP analysis, computed respectively using "varImp" and "kernelshap" R packages, to investigate those features that had the strongest impact on the classification model.

The comprehensive multi-omics approach integrating metagenomics, volatilomics, and metaproteomics data is summarized in Fig. 8, which details the sampling strategy, analytical methodologies, data processing pipeline, and correlation analyses performed to examine relationships between microbial communities and cheese metabolic profiles.

Data availability

Raw reads generated from the samples have been deposited in the Sequence Read Archive (SRA) of NCBI under accession numbers PRJNA1233909 and PRJNA1206071 (Supplementary Data 7).

Received: 14 April 2025; Accepted: 14 August 2025;
Published online: 27 August 2025

References

1. Galli, B. D. Sustainability implications and relevance of using omics sciences to investigate cheeses with protected designation of origin. *J. Sci. Food Agric.* **104**, 6388–6396 (2024).
2. Masotti, F., Hogenboom, J. A., Rosi, V., De Noni, I. & Pellegrino, L. Proteolysis indices related to cheese ripening and typicalness in PDO Grana Padano cheese. *Int. Dairy J.* **20**, 352–359 (2010).
3. Ercolini, D. High-throughput sequencing and metagenomics: moving forward in the culture-independent analysis of food microbial ecology. *Appl. Environ. Microbiol.* **79**, 3148–3155 (2013).
4. Giraffa, G. The Microbiota of Grana Padano cheese. A review. *Foods* **10**, 2632 (2021).
5. De Filippis, F. et al. Microbiome mapping in dairy industry reveals new species and genes for probiotic and bioprotective activities. *npj Biofilms Microbiomes* **10**, 67 (2024).
6. Afshari, R., Pillidge, C. J., Dias, D. A., Osborn, A. M. & Gill, H. Cheesomics: the future pathway to understanding cheese flavour and quality. *Crit. Rev. Food Sci. Nutr.* **60**, 33–47 (2020).
7. De Filippis, F., Genovese, A., Ferranti, P., Gilbert, J. A. & Ercolini, D. Metatranscriptomics reveals temperature-driven functional changes in microbiome impacting cheese maturation rate. *Sci. Rep.* **6**, 21871 (2016).
8. 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. *Compr. Rev. Food Sci. Food Saf.* **17**, 371–390 (2018).
9. Anastasiou, R. et al. Omics approaches to assess flavor development in cheese. *Foods* **11**, 188 (2022).
10. Neviani, E., Bottari, B., Lazzi, C. & Gatti, M. New developments in the study of the microbiota of raw-milk, long-ripened cheeses by molecular methods: the case of Grana Padano and Parmigiano Reggiano. *Front. Microbiol.* **4**, 36 (2013).
11. Qian, M. & Reineccius, G. Potent aroma compounds in Parmigiano Reggiano cheese studied using a dynamic headspace (purge-trap) method. *Flavour Fragr. J.* **18**, 252–259 (2003).
12. Hayaloglu, A. A. Volatile composition and proteolysis in traditionally produced mature Kashar cheese. *Int. J. Food Sci. Technol.* **44**, 1388–1394 (2009).

13. Bottari, B. et al. The interrelationship between microbiota and peptides during ripening as a driver for Parmigiano Reggiano cheese quality. *Front. Microbiol.* **11**, 581658 (2020).
14. Sforza, S., Ferroni, L., Galaverna, G., Dossena, A. & Marchelli, R. Extraction, semi-quantification, and fast on-line identification of oligopeptides in Grana Padano cheese by HPLC-MS. *J. Agric. Food Chem.* **51**, 2130–2135 (2003).
15. Summer, A. et al. Cheese as functional food: the example of Parmigiano Reggiano and Grana Padano. *Food Technol. Biotechnol.* **55**, 277–289 (2017).
16. Rangel, A. H. D. N. et al. An overview of the occurrence of bioactive peptides in different types of cheeses. *Foods* **12**, 4261 (2023).
17. Dinan, T. G., Stanton, C. & Cryan, J. F. Psychobiotics: a novel class of psychotropic. *Biol. Psychiatry* **74**, 720–726 (2013).
18. Balasubramanian, R., Schneider, E., Gunnigle, E., Cotter, P. D. & Cryan, J. F. Fermented foods: Harnessing their potential to modulate the microbiota-gut-brain axis for mental health. *Neurosci. Biobehav. Rev.* **158**, 105562 (2024).
19. Ijaz, M. U. et al. Microbiome and neurological disorders. in *Human Microbiome* 273–301. https://doi.org/10.1007/978-981-97-3790-1_9 (Springer Nature Singapore, 2024).
20. Magliulo, R. et al. Microbiome signatures associated with flavor development differentiate Protected Designation of origin water Buffalo Mozzarella cheese from different production areas. *Food Res. Int.* **192**, 114798 (2024).
21. Afshari, R. et al. New insights into cheddar cheese microbiota-metabolome relationships revealed by integrative analysis of multi-omics data. *Sci. Rep.* **10**, 3164 (2020).
22. Bottari, B., Levante, A., Neviani, E. & Gatti, M. How the fewest become the greatest. L. casei's impact on long ripened cheeses. *Front. Microbiol.* **9**, 2866 (2018).
23. Kim, E., Yang, S.-M. & Kim, H.-Y. Differentiation of Lacticaseibacillus zeae using pan-genome analysis and real-time PCR method targeting a unique gene. *Foods* **10**, 2112 (2021).
24. da Silva Duarte, V., Lombardi, A., Corich, V. & Giacomini, A. Assessment of the microbiological origin of blowing defects in Grana Padano Protected Designation of Origin cheese. *J. Dairy Sci.* **105**, 2858–2867 (2022).
25. Avila, M., Gómez-Torres, N., Hernández, M. & Garde, S. Inhibitory activity of reuterin, nisin, lysozyme and nitrite against vegetative cells and spores of dairy-related Clostridium species. *Int. J. Food Microbiol.* **172**, 70–75 (2014).
26. Kleerebezem, M. et al. The extracellular biology of the lactobacilli. *FEMS Microbiol. Rev.* **34**, 199–230 (2010).
27. D'Incecco, P. et al. Lysozyme affects the microbial catabolism of free arginine in raw-milk hard cheeses. *Food Microbiol.* **57**, 16–22 (2016).
28. Dias, R., Vilas-Boas, E., Campos, F. M., Hogg, T. & Couto, J. A. Activity of lysozyme on Lactobacillus hilgardii strains isolated from Port wine. *Food Microbiol.* **49**, 6–11 (2015).
29. Bassi, D., Puglisi, E. & Cocconcelli, P. S. Understanding the bacterial communities of hard cheese with blowing defect. *Food Microbiol.* **52**, 106–118 (2015).
30. Courtin, P. et al. Accelerating cheese proteolysis by enriching Lactococcus lactis proteolytic system with lactobacilli peptidases. *Int. Dairy J.* **12**, 447–454 (2002).
31. Gatti, M. et al. A model to assess lactic acid bacteria aminopeptidase activities in Parmigiano Reggiano cheese during ripening. *J. Dairy Sci.* **91**, 4129–4137 (2008).
32. Bancalari, E. et al. An integrated strategy to discover Lactobacillus casei group strains for their potential use as aromatic starters. *Food Res. Int.* **100**, 682–690 (2017).
33. Sgarbi, E. et al. Nonstarter lactic acid bacteria volatilomes produced using cheese components. *J. Dairy Sci.* **96**, 4223–4234 (2013).
34. Randazzo, C. L. et al. Preliminary characterization of wild lactic acid bacteria and their abilities to produce flavour compounds in ripened model cheese system. *J. Appl. Microbiol.* **103**, 427–435 (2007).
35. Lazzi, C. et al. Can the development and autolysis of lactic acid bacteria influence the cheese volatile fraction? The case of Grana Padano. *Int. J. Food Microbiol.* **233**, 20–28 (2016).
36. Illikoud, N., Mantel, M., Rolli-Derkinderen, M., Gagnaire, V. & Jan, G. Dairy starters and fermented dairy products modulate gut mucosal immunity. *Immunol. Lett.* **251–252**, 91–102 (2022).
37. Walther, B. et al. Quantitative analysis of menaquinones (vitamin K2) in various types of cheese from Switzerland. *Int. Dairy J.* **112**, 104853 (2021).
38. Berding, K. et al. Feed your microbes to deal with stress: a psychobiotic diet impacts microbial stability and perceived stress in a healthy adult population. *Mol. Psychiatry* **28**, 601–610 (2023).
39. Moreira, G. M. M. et al. Effect of ripening time on proteolysis, free amino acids, bioactive amines and texture profile of Gorgonzola-type cheese. *LWT* **98**, 583–590 (2018).
40. Lacroix, N., St-Gelais, D., Champagne, C. P. & Vuillemard, J. C. Gamma-aminobutyric acid-producing abilities of lactococcal strains isolated from old-style cheese starters. *Dairy Sci. Technol.* **93**, 315–327 (2013).
41. Sousa, R. J. M., Ribeiro, S. C., Baptista, J. A. B. & Silva, C. C. G. Evaluation of gamma-aminobutyric acid content in Portuguese cheeses with protected designation of origin status. *J. Dairy Res.* **90**, 1–4 (2023).
42. Park, K.-B. & Oh, S.-H. Isolation and characterization of Lactobacillus buchneri strains with high γ-aminobutyric acid producing capacity from naturally aged cheese. *Food Sci. Biotechnol.* **15**, 86–90 (2006).
43. Cho, Y. R., Chang, J. Y. & Chang, H. C. Production of gamma-aminobutyric acid (GABA) by Lactobacillus buchneri isolated from kimchi and its neuroprotective effect on neuronal cells. *J. Microbiol. Biotechnol.* **17**, 104–109 (2007).
44. Valenzuela, J. A., Flórez, A. B., Vázquez, L., Vasek, O. M. & Mayo, B. Production of γ-aminobutyric acid (GABA) by lactic acid bacteria strains isolated from traditional, starter-free dairy products made of raw milk. *Benef. Microbes* **10**, 579–587 (2019).
45. Sanders, J. W. et al. A chloride-inducible acid resistance mechanism in Lactococcus lactis and its regulation. *Mol. Microbiol.* **27**, 299–310 (1998).
46. Chen, M. et al. Neurotransmitter and intestinal interactions: focus on the Microbiota-gut-brain axis in irritable bowel syndrome. *Front. Endocrinol.***13**, 817100 (2022).
47. Laroute, V. et al. Lactococcus lactis NCDO2118 exerts visceral antinociceptive properties in rat via GABA production in the gastro-intestinal tract. *eLife* **11**, e77100 (2022).
48. Liang, S. et al. Administration of Lactobacillus helveticus NS8 improves behavioral, cognitive, and biochemical aberrations caused by chronic restraint stress. *Neuroscience* **310**, 561–577 (2015).
49. Murru, E. et al. Conjugated linoleic acid and brain metabolism: a possible anti-neuroinflammatory role mediated by PPARα activation. *Front. Pharmacol.* **11**, 587140 (2020).
50. Fujita, Y. et al. Dietary cis-9, trans-11-conjugated linoleic acid reduces amyloid β-protein accumulation and upregulates anti-inflammatory cytokines in an Alzheimer's disease mouse model. *Sci. Rep.* **11**, 9749 (2021).
51. Valcarcel-Jimenez, L. & Frezza, C. Fumarate hydratase (FH) and cancer: a paradigm of oncometabolism. *Br. J. Cancer* **129**, 1546–1557 (2023).
52. Wikoff, W. R. et al. Metabolomics analysis reveals large effects of gut microflora on mammalian blood metabolites. *Proc. Natl. Acad. Sci. USA* **106**, 3698–3703 (2009).
53. Bertani, G. et al. Dynamics of a natural bacterial community under technological and environmental pressures: the case of natural whey

starter for Parmigiano Reggiano cheese. *Food Res. Int.* **129**, 108860 (2020).
54. Neviani, E., Levante, A. & Gatti, M. The microbial community of natural whey starter: Why is it a driver for the production of the most famous Italian long-ripened cheeses?. *Fermentation* **10**, 186 (2024).
55. Lazzi, C., Rossetti, L., Zago, M., Neviani, E. & Giraffa, G. Evaluation of bacterial communities belonging to natural whey starters for Grana Padano cheese by length heterogeneity-PCR. *J. Appl. Microbiol.* **96**, 481–490 (2004).
56. De Dea Lindner, J. et al. Parmigiano Reggiano cheese: evolution of cultivable and total lactic microflora and peptidase activities during manufacture and ripening. *Dairy Sci. Technol.* **88**, 511–523 (2008).
57. Schmieder, R. & Edwards, R. Quality control and preprocessing of metagenomic datasets. *Bioinformatics* **27**, 863–864 (2011).
58. Blanco-Míguez, A. et al. Extending and improving metagenomic taxonomic profiling with uncharacterized species using MetaPhlAn 4. *Nat. Biotechnol.* **41**, 1633–1644 (2023).
59. Li, D., Liu, C.-M., Luo, R., Sadakane, K. & Lam, T.-W. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* **31**, 1674–1676 (2015).
60. Wood, D. E. & Salzberg, S. L. Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biol.* **15**, R46 (2014).
61. Blin, K. et al. antiSMASH 7.0: new and improved predictions for detection, regulation, chemical structures and visualisation. *Nucleic Acids Res.* **51**, W46–W50 (2023).
62. Pasolli, E. et al. Extensive unexplored human microbiome diversity revealed by over 150,000 genomes from metagenomes spanning age, geography, and lifestyle. *Cell* **176**, 649–662.e20 (2019).
63. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
64. Kang, D. D., Froula, J., Egan, R. & Wang, Z. MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. *PeerJ* **3**, e1165 (2015).
65. 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 Res.* **25**, 1043–1055 (2015).
66. Chaumeil, P.-A., Mussig, A. J., Hugenholtz, P. & Parks, D. H. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics* **36**, 1925–1927 (2019).
67. Letunic, I. & Bork, P. Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. *Nucleic Acids Res.* **52**, W78–W82 (2024).
68. Liao, H., Ji, Y. & Sun, Y. High-resolution strain-level microbiome composition analysis from short reads. *Microbiome* **11**, 183 (2023).
69. Zhu, W., Lomsadze, A. & Borodovsky, M. Ab initio gene identification in metagenomic sequences. *Nucleic Acids Res.* **38**, e132 (2010).
70. Rawlings, N. D., Waller, M., Barrett, A. J. & Bateman, A. MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Res.* **42**, D503–D509 (2014).
71. Buchfink, B., Reuter, K. & Drost, H.-G. Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nat. Methods* **18**, 366–368 (2021).
72. Fu, L., Niu, B., Zhu, Z., Wu, S. & Li, W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28**, 3150–3152 (2012).
73. Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* **5**, 621–628 (2008).
74. da Silva Duarte, V. et al. Database selection for shotgun metaproteomic of low-diversity dairy microbiomes. *Int. J. Food Microbiol.* **418**, 110706 (2024).
75. Balivo, A. et al. Can hydroponic forage affect the chemical and sensory properties of PDO buffalo Mozzarella cheese? *Int. J. Dairy Technol*. **78**, e13147 (2025).

Acknowledgements

This work was partially funded by the project HOLOGRAM—Exploiting autochthonous microbial resources from traditional Italian fermented foods for gut-brain axis modulation, funded by the European Union—NextGenerationEU, NRRP Missione 4 "Istruzione e Ricerca"—Componente C2, Investimento 1.1, "Fondo per il Programma Nazionale di Ricerca e Progetti di Rilevante Interesse Nazionale (PRIN)" (P2022X8A9M) and by the European Union—NextGenerationEU, NRRP—Mission 4, Component 2, Investment 1.4—National Biodiversity Future Center—CN_00000033 (CUP E63C22000990007; D.M. Prot. 1034 of 17/06/2022). The activities were developed within the infrastructure funded by the European Union—NextGenerationEU, NRRP—Mission 4 "Education and Research" Component 2: from—research to business, Investment 3.1: Fund for the realization of an integrated system of research and innovation infrastructures (IR0000033, CUP I83C22001040006, Project METRO-FOOD-IT, D.M. Prot. n.120 del 21/06/2022).

Author contributions

V.V., R.M., and F.D.F. wrote the first manuscript draft; F.D.F. carried out the sampling; V.V. and A.E. performed DNA extraction; A.G. and A.B. performed the volatilome analysis; D.P. and A.M.K. performed the proteome analysis; C.M.C. and G.S. carried out the search for genes linked with neuroactive compounds production; V.V. and R.M. performed bioinformatic and statistical analyses and prepared figures and tables; F.D.F. and D.E. designed and supervised the study and acquired funding. All authors reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41522-025-00815-6.

Correspondence and requests for materials should be addressed to Francesca De Filippis.