

<https://doi.org/10.1038/s41522-024-00541-5>

# Microbiome mapping in dairy industry reveals new species and genes for probiotic and bioprotective activities

Check for updates

Francesca De Filippis<sup>1,2</sup>, Vincenzo Valentino<sup>1</sup>, Min Yap<sup>3</sup>, Raul Cabrera-Rubio<sup>3,4</sup>, Coral Barcenilla<sup>5</sup>, Niccolò Carlino<sup>6</sup>, José F. Cobo-Díaz<sup>5</sup>, Narciso Martín Quijada<sup>7,8,9</sup>, Inés Calvete-Torre<sup>10,11</sup>, Patricia Ruas-Madiedo<sup>10,11</sup>, Carlos Sabater<sup>10,11</sup>, Giuseppina Sequino<sup>1</sup>, Edoardo Pasolli<sup>1,2</sup>, Martin Wagner<sup>7,8</sup>, Abelardo Margolles<sup>10,11</sup>, Nicola Segata<sup>6</sup>, Avelino Álvarez-Ordóñez<sup>5</sup>, Paul D. Cotter<sup>3,12,13</sup> & Danilo Ercolini<sup>1,2,13</sup> ✉

The resident microbiome in food industries may impact on food quality and safety. In particular, microbes residing on surfaces in dairy industries may actively participate in cheese fermentation and ripening and contribute to the typical flavor and texture. In this work, we carried out an extensive microbiome mapping in 73 cheese-making industries producing different types of cheeses (fresh, medium and long ripened) and located in 4 European countries. We sequenced and analyzed metagenomes from cheese samples, raw materials and environmental swabs collected from both food contact and non-food contact surfaces, as well as operators' hands and aprons. Dairy plants were shown to harbor a very complex microbiome, characterized by high prevalence of genes potentially involved in flavor development, probiotic activities, and resistance to gastro-intestinal transit, suggesting that these microbes may potentially be transferred to the human gut microbiome. More than 6100 high-quality Metagenome Assembled Genomes (MAGs) were reconstructed, including MAGs from several Lactic Acid Bacteria species and putative new species. Although microbial pathogens were not prevalent, we found several MAGs harboring genes related to antibiotic resistance, highlighting that dairy industry surfaces represent a potential hotspot for antimicrobial resistance (AR) spreading along the food chain. Finally, we identified facility-specific strains that can represent clear microbial signatures of different cheesemaking facilities, suggesting an interesting potential of microbiome tracking for the traceability of cheese origin.

Cheese production is a powerful method to extend the shelf life of milk and retain its nutritional value<sup>1</sup>. Cheesemaking was probably first performed by chance more than 9000 years ago<sup>2</sup>, but since then a rich variety of cheeses have been developed, differing locally in the type of milk used (raw, pasteurized/thermized, and from different animals), production and ripening

technologies and microbial communities naturally selected or deliberately added to achieve the desired sensorial properties.

According to the USDA, > 6 and > 10 million metric tons of cheeses were produced during 2022–2023 in the United States and Europe, respectively (<https://fas.usda.gov/data/production/commodity/0240000>),

<sup>1</sup>Dept. of Agricultural Sciences, University of Naples Federico II, Portici, NA, Italy. <sup>2</sup>Task Force on Microbiome Studies, University of Naples Federico II, Napoli, NA, Italy. <sup>3</sup>Teagasc Food Research Centre, Moorepark, Fermoy, Cork, Ireland. <sup>4</sup>Institute of Agrochemistry and Food Technology, National Research Council (IATA-CSIC), Paterna, Spain. <sup>5</sup>Department of Food Hygiene and Technology and Institute of Food Science and Technology, Universidad de León, León, Spain. <sup>6</sup>Department CIBIO, University of Trento, Trento, Italy. <sup>7</sup>Austrian Competence Centre for Feed and Food Quality, Safety and Innovation, FFOQSI GmbH, Tulln an der Donau, Austria. <sup>8</sup>Department for Farm Animals and Veterinary Public Health, Unit of Food Microbiology, Institute of Food Safety, Food Technology and Veterinary Public Health, University of Veterinary Medicine Vienna, Vienna, Austria. <sup>9</sup>Department of Microbiology and Genetics, Institute for AgriBiotechnology Research (CIALE), University of Salamanca, Salamanca, Spain. <sup>10</sup>Department of Microbiology and Biochemistry of Dairy Products, Instituto de Productos Lácteos de Asturias - Consejo Superior de Investigaciones Científicas (IPLA-CSIC), Villaviciosa, Spain. <sup>11</sup>Microhealth Group, Instituto de Investigación Sanitaria del Principado de Asturias (ISPA), Oviedo, Spain. <sup>12</sup>APC Microbiome Ireland, University College Cork, Cork, Ireland. <sup>13</sup>These authors jointly supervised this work: Paul D. Cotter, Danilo Ercolini. ✉e-mail: [ercolini@unina.it](mailto:ercolini@unina.it)

and cheese consumption is projected to increase in the next decade<sup>3</sup>, showing that this sector continues to be deep-rooted in human cultural heritage. However, despite the success of the dairy industry, it is estimated that it is the third largest sector in terms of food loss, with ~25% of the dairy production discarded at industrial and household level because of the growth of undesirable microorganisms<sup>4</sup>. Indeed, the ecological properties of cheeses can support the growth of a wide range of microorganisms, including spoilers/pathogens. Notably, *Listeria monocytogenes* can grow well in surface-ripened and fresh cheeses, and several outbreaks involving this microorganism have been reported<sup>5</sup>. However, it is not the sole pathogen associated with cheeses, as others, such as *Staphylococcus aureus* and *Salmonella* spp., can also easily proliferate in the dairy environment<sup>6</sup>.

Cheeses can be spoiled by several microbial species, such as *Pseudomonas putida* and *Ps. fluorescens*. These taxa harbor a broad range of lipolytic and proteolytic enzymes often responsible for off-flavors and the discoloration of cheeses<sup>7–9</sup>. Indeed, most of the spoiling *Pseudomonas* spp. found in dairy products produce the extracellular, heat-stable metalloprotease AprX that targets many different sites in casein<sup>10</sup>. Also, blue pigmentation of fresh and soft cheeses caused by *Ps. fluorescens* has been widely described<sup>11,12</sup>. However, depending on the cheese type, other spoiling bacteria might prevail. For instance, the uncontrolled growth of *Clostridium tyrobutyricum* might lead to late blowing, with the formation of oversized 'eyes' (i.e., gas holes) during ripening, and to the development of rancid off-flavors in some Swiss- and Dutch-type cheeses<sup>13</sup> as well as in hard and long ripened Italian cheeses (e.g., Grana Padano and Parmigiano Reggiano<sup>14</sup>). Furthermore, a strain of *Serratia marcescens* was recently identified as the main responsible of red discoloration in Cabrales cheese<sup>15</sup>. Indeed, although this work is focused only on the bacterial community, also Fungi (yeasts and molds) can be important for the spoilage of cheese<sup>16</sup>, or can actively participate in the manufacturing some types of cheeses.

Although both pathogenic and spoilage microorganisms naturally occurring in milk are easily deactivated by heat treatments commonly performed as preliminary step in most cheese-making processes (e.g., pasteurization or thermization<sup>17</sup>), such microorganisms and their endospores can survive in the processing plant by forming a biofilm, a matrix made of extracellular polymeric substances strongly bound to the surface<sup>18</sup>, which may entrap microbial cells making the cleaning of the processing plants difficult<sup>6</sup>. However, biofilms might also act as a reservoir of protechnological strains conferring positive sensorial attributes<sup>19</sup>, particularly in cheesemaking. For example, the biofilm on wooden vats is considered as the main source of acidifying Lactic Acid Bacteria (LAB) in the production of the Protected Designation of Origin (PDO) Ragusano cheese<sup>20</sup> and several other cheeses<sup>21,22</sup>. Also, biofilms on dairy plant surfaces may represent a reservoir of non-starter LAB (NSLAB<sup>23</sup>), potentially contributing to shaping the sensorial profiles of cheeses.

Overall, the residential microbial communities inhabiting the cheese production environment may impact on the quality and safety of the final products, therefore merit an in-depth characterization in order to understand their contribution to shaping the cheese microbiome. To date, a validated procedure to detect and functionally characterize all the microorganisms from the processing environment is lacking, although public investment is currently focusing on this task. Indeed, the recently ended MASTER project (Microbiome applications for Sustainable Food Systems through Technologies and Enterprise, [www.master-h2020.eu](http://www.master-h2020.eu)), granted by the European Union within the Horizon2020 Programme, aimed at developing Standard Operating Procedures (SOPs) for microbiome mapping in food processing environments, to provide food industries with a rapid and effective tool to detect both potentially hazardous and beneficial microorganisms in their facilities. These SOPs will help cheesemakers to quickly detect target microorganisms, as well as to deeply describe the metabolic potential of the microbiome residing in their environment, supporting the overall quality and safety management plans. In the long run, this will lead to a reduction of food spoilage and waste and an improvement of food safety, paving the way to more sustainable food systems.

In the current study, we applied MASTER SOPs to characterize the resident microbiome in 73 facilities, and of 24 cheese types produced at these facilities, from across 4 European countries.

## Results

### Microbiome mapping in dairy industries shows substantially different microbiome compositions from surfaces through raw materials and end products

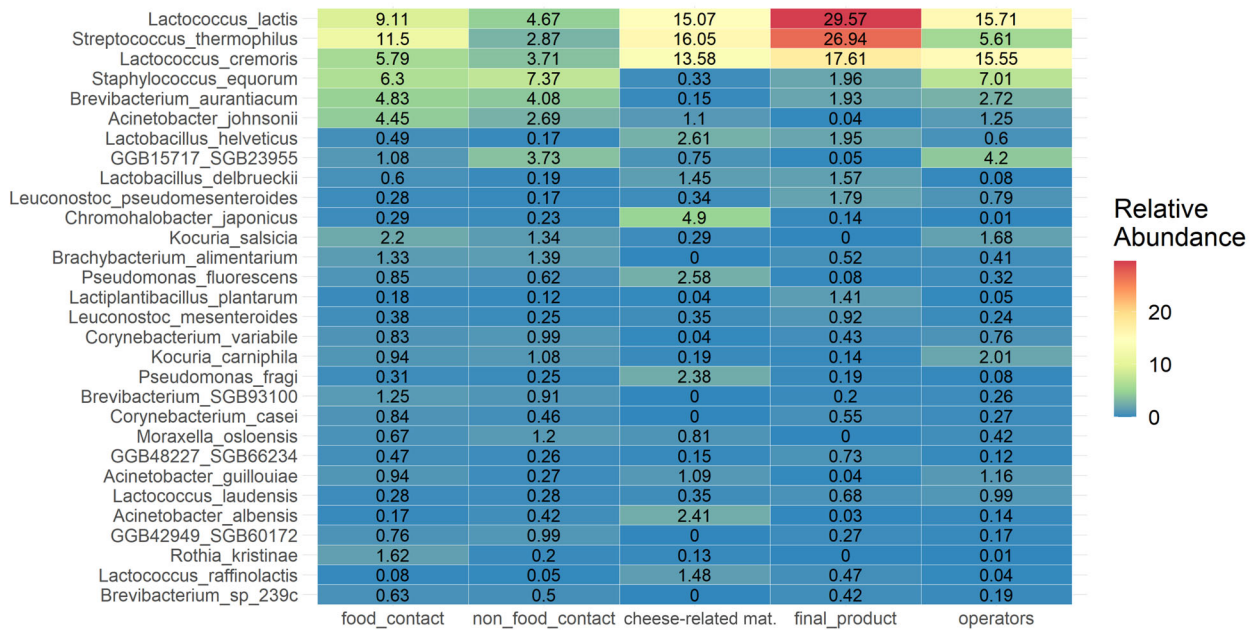
A total of 1250 samples were collected throughout the study, originating from Austria ( $n = 101$ ), Ireland ( $n = 274$ ), Italy ( $n = 216$ ), and Spain ( $n = 659$ ). Out of all samples, 544 samples were swabs from food contact surfaces (FC,  $n = 308$ ), non-food contact surfaces (NFC,  $n = 199$ ) and operators' hands/aprons (OP,  $n = 37$ ). Samples of raw milk ( $n = 67$ ), brine ( $n = 44$ ) and whey ( $n = 71$ ) made up the 182 cheese-related material (CRM) samples and the final products ( $n = 524$ ) consisted of samples from the core ( $n = 265$ ) and rind ( $n = 259$ ) of the cheese, which were also sampled before ( $n = 237$ ), during ( $n = 26$ ) and after ( $n = 261$ ) ripening.

From these samples, a total of 239 species were detected at  $> 0.01\%$  relative abundance. The top 3 species detected, with the highest average relative abundances across all the samples, were *Lactococcus lactis* ( $18.04\% \pm 26.8\%$ ), *Streptococcus thermophilus* ( $17.09\% \pm 28.8\%$ ) and *Lc. cremoris* ( $11.84\% \pm 22.6\%$ ) (Fig. 1A). When stratified into categories, raw materials and final products were dominated by these 3 species (Fig. 1A) and were found to have significantly lower alpha diversity compared to the FC and NFC surfaces and operators' swabs ( $p < 0.05$ ) (Fig. 1B). In addition, NFC had a significantly higher microbial diversity than FC ( $p < 0.05$ ; Fig. 1B). Apart from the three LAB species, other abundant taxa belonged to *Staphylococcus equorum*, *Brevibacterium aurantiacum* and *Acinetobacter johnsonii*, that were more abundant on surfaces than in the final products and raw materials (Fig. 1A). Significant differences in the taxonomic composition between all the categories suggested substantially different microbiome compositions across the various types of matrices sampled (PERMANOVA,  $p < 0.05$ ) (Fig. 1C). Considering potentially pathogenic/spoiling taxa, *A. johnsonii* was the most abundant, with average values of 4.5 and 2.7% in FC and NFC, respectively, but at  $< 0.1\%$  in the final products (Fig. 1A). *Staph. aureus* was detected in a total of 92 samples (about 58% of them being final products, while 20% and 10% corresponding to FC and NFC, respectively). However, the overall relative abundance was extremely low, being  $> 0.1\%$  in only 14 final products, 3 FC and 1 NFC surfaces. Moreover, *Listeria monocytogenes* and *Salmonella* spp. were not detected in any of the samples. *Pseudomonas fluorescens* and *Ps. fragi* were detected among the top 30 species across all categories but were only present in low relative abundances (Fig. 1A, average 0.73% and 0.55%, respectively). These two *Pseudomonas* spp. were generally found in higher relative abundances in the raw materials compared to the other samples, however, in 4 swab samples (3 FC and 1 NFC surface), they were the most abundant species detected (22–45%).

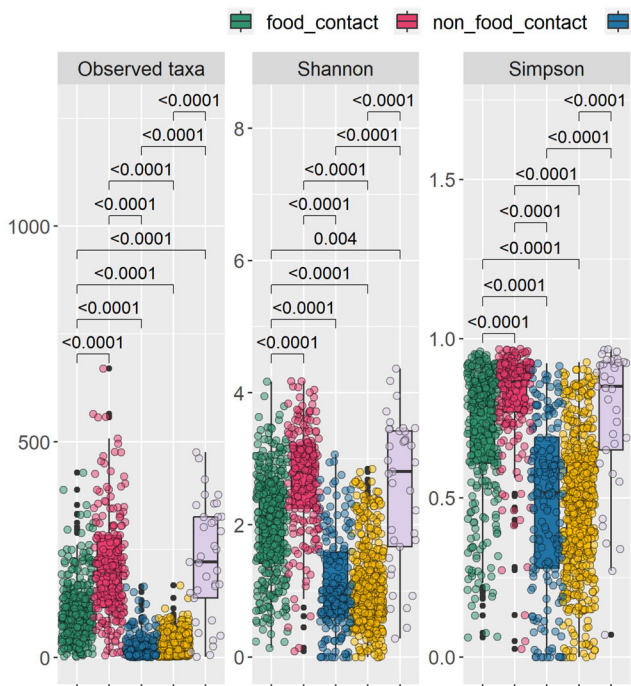
Within the samples representative of the raw materials used for cheese manufacture, a significantly higher alpha diversity was observed in brine, while whey had the lowest diversity (Supplementary Fig. 1). Consistently, whey samples clustered distinctly from samples of raw milk and brine (Supplementary Fig. 1) due to the dominance of 3 main species (*Lc. cremoris*, *Lc. lactis* and *Strep. thermophilus*) that account for an average relative abundance of 82% in all whey samples, and also due to the high abundance of *Chromohalobacter japonicus* (average relative abundance = 20.3%) and *Lc. raffinolactis* (average relative abundance = 3.6%) in brine and raw milk samples, respectively (Supplementary Table 1).

During production, as expected, the microbiome of the cheeses dynamically changed. The number of taxa detected was significantly different between the three time points (before, during and at the end of ripening; Supplementary Fig. 1), with more species observed during ripening, followed by after ripening. The diversity of the cheese microbiome significantly increased during ripening and stabilized towards the end of ripening (Supplementary Fig. 1). However, no clear clustering of the samples was observed (Supplementary Fig. 1). During ripening, different species

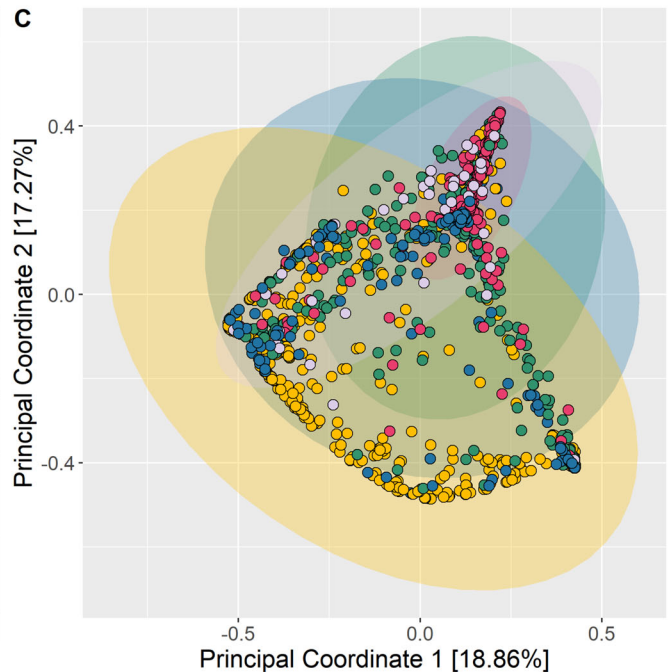
A



B



C



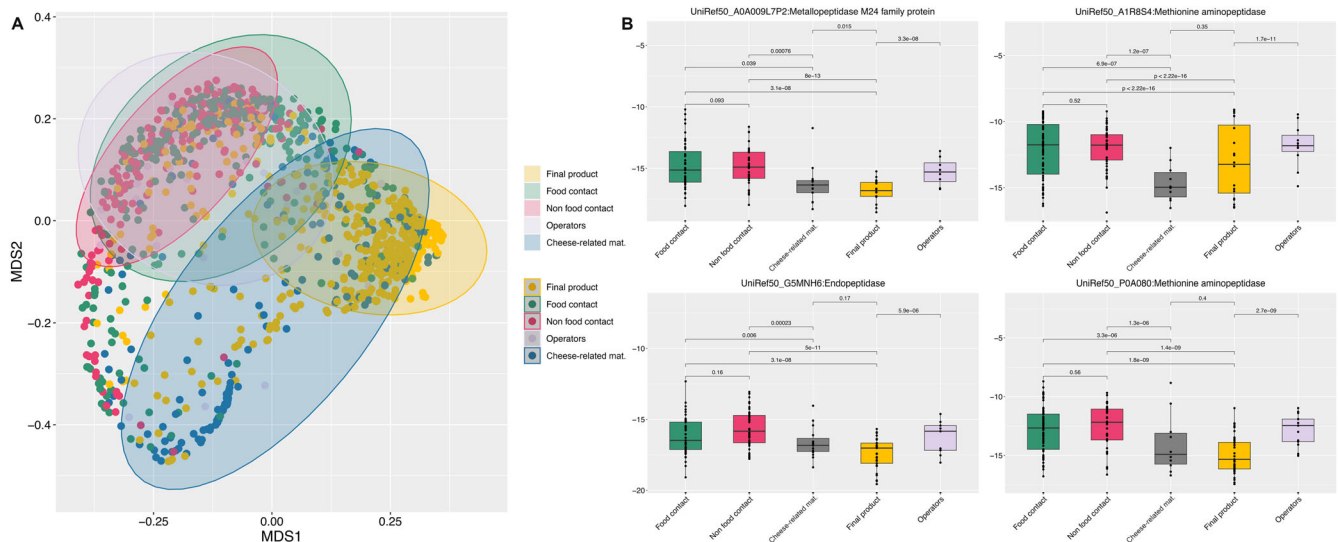
**Fig. 1 | Different taxonomic profiles are present in cheese and environmental swabs.**

**A.** Taxonomic composition of the 30 most abundant species in all samples, with the average relative abundances of each species in each category. **B.** Alpha diversity, in terms of observed species, Shannon and Simpson analysis of samples by category, and **(C)** beta diversity analysis represented by a Principal-Coordinate Analysis (PCoA) plot of Bray-Curtis distance, with ellipses representing clustering by category. Boxes represent the interquartile range (IQR) between the first and third

quartiles, and the line inside represents the median (2nd quartile). Whiskers denote the lowest and the highest values within 1.5 x IQR from the first and third quartiles, respectively. The significance was tested by applying pairwise Wilcoxon test. Average values are obtained from  $n = 308, 199, 37, 524,$  and  $182$  biologically independent samples from food contact, non food contact, operators' swabs, final products and cheese-related materials, respectively. The category "cheese-related materials" groups together milk, brine and whey culture.

were found to have significantly higher abundances (Supplementary Table 1), for example, the levels of *Lc. cremoris* and *Lc. lactis* were significantly higher before and after ripening when compared to during ripening, whereas the abundance of *Strep. thermophilus* was significantly higher during, compared to before and after, ripening.

We also compared the microbial community of cheese rind and core, observing a higher diversity on the rind compared to the core (Supplementary Fig. 1). Several species were found at significantly higher relative abundances in the core compared to the rind (Supplementary Table 1). Overall, the core samples had significantly higher abundances of the



**Fig. 2 | Different functional profiles are present in cheese and environmental swabs.** **A** Non-metric Multidimensional Scaling (NMDS) plot based on Jaccard's distance of functional profiles obtained by HUMAnN. Samples are colored according to the sample type. **B** Boxplots showing the abundance (log values) of UniRef50 genes detected in the different sample groups. Boxes represent the interquartile range (IQR) between the first and third quartiles, and the line inside represents the median (2nd quartile). Whiskers denote the lowest and the highest values within  $1.5 \times$  IQR

from the first and third quartiles, respectively. The significance was tested by applying pairwise Wilcoxon test. The category "cheese-related materials" groups together milk, brine and whey culture. Average values are obtained from  $n = 308, 199, 37, 524,$  and  $182$  biologically independent samples from food contact, non food contact, operators' swabs, final products and cheese-related materials, respectively.

fermenting LAB *Lactobacillus delbrueckii*, *Lb. helveticus* and *Strep. thermophilus* while significantly higher levels of aerobic bacteria were found in the rind samples (i.e. *B. alimentarium*, *B. aurantiacum*, other unidentified *Brevibacterium* species, *Chromohalobacter japonicus*, *Corynebacterium casei*, *C. variabile*, *Kocuria salsicia* and *Staph. equorum*). We further compared the cheeses grouped according to the ripening time (<10 days, not ripened; 10–30 days, medium ripened; > 30 days, long ripened) and observed that alpha diversity increased with ripening length (Supplementary Fig. 1). No significant differences were observed in beta diversity. However, cheeses that were not ripened had significantly higher abundances of *Lc. cremoris* and *Strep. thermophilus* compared to the medium- and long-ripened cheeses (Supplementary Table 1).

When grouping cheeses according to the production technology, we found that hard/semi-hard surface-ripened cheeses have the highest alpha diversity and that unripened pasta-filata cheeses were those with the lowest diversity (Supplementary Fig. 1). In addition, pasta-filata cheeses (both unripened and ripened) were dominated by *Strep. thermophilus* at significantly higher relative abundance than other cheeses (Supplementary Table 1). *Lactococcus* spp. were the most abundant taxa in other cheese types: *Lc. lactis* dominated in blue and hard/semi-hard not smear-ripened cheeses, while *Lc. cremoris* in soft cheese and hard/semi-hard, surface-ripened cheeses (Supplementary Table 1).

### The microbiome on dairy industry surfaces has an increased proteolytic activity and the potential to resist GIT passage and colonize the gut

We investigated functional profiles of microbiomes by mapping short-reads against functional databases through HUMAnN3. The different taxonomic compositions highlighted above (Fig. 1) were reflected in different functional profiles (Fig. 2A), as also highlighted by pairwise comparisons using MANOVA (corrected  $p$ -value < 0.01 for all the comparisons).

In particular, the microbiome of raw materials and final products showed a higher relative abundance of pathways related to carbohydrate metabolism [e.g., (R,R)-butanediol biosynthesis, Supplementary Fig. 2], while surface-related microbiomes (FC, NFC, operators' swabs) were characterized by higher abundance of pathways associated with amino acid degradation (e.g., L-leucine, L-arginine and L-ornithine degradation and

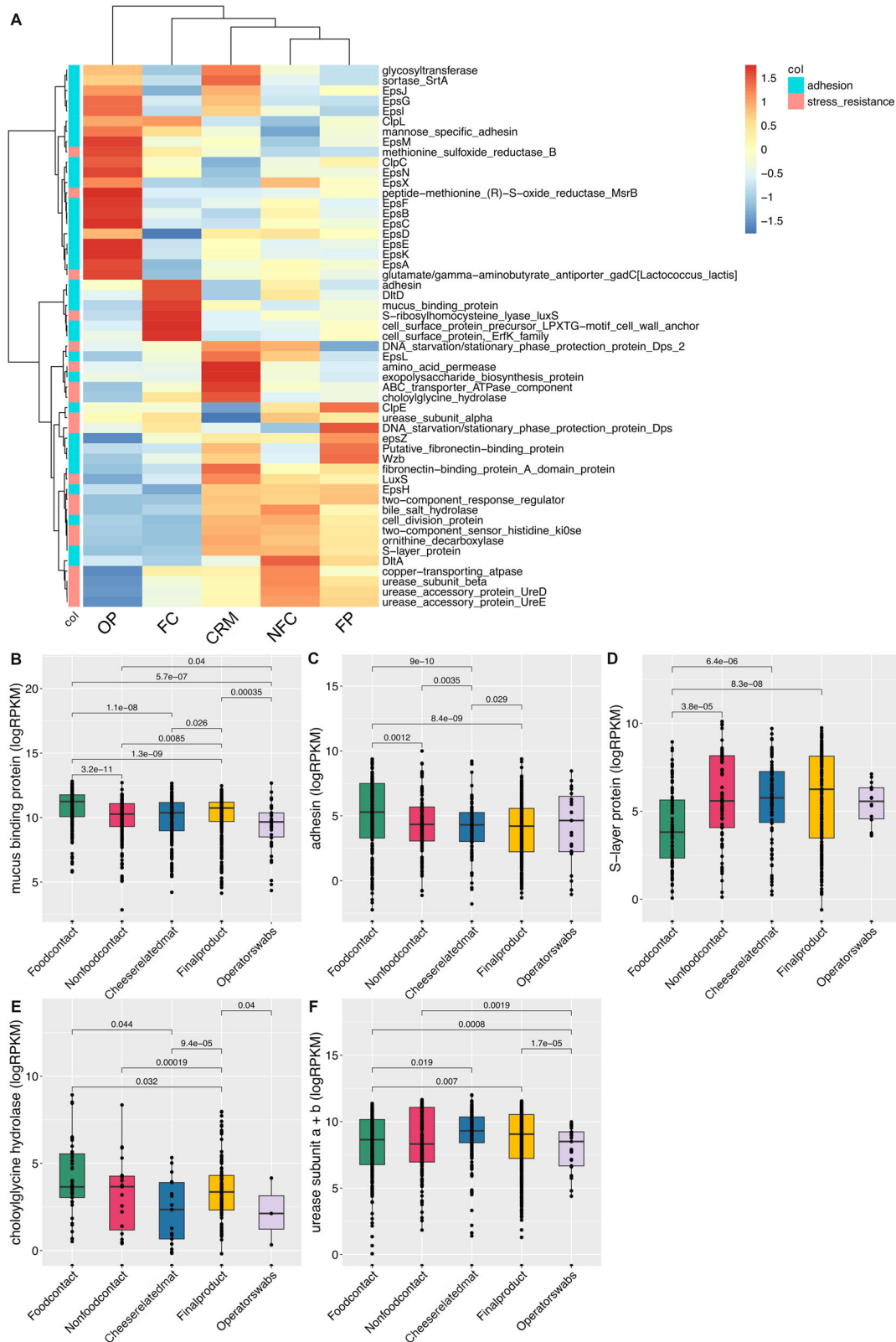
putrescine biosynthesis), as well as fatty acid oxidation and biotin biosynthesis (Supplementary Fig. 2).

In addition, we further screened gene-level profiles obtained using UniRef50 gene assignment. Consistently, the relative abundance of several genes coding for peptidases (e.g., UniRef50\_A0A009L7P2, metallopeptidase family M24; UniRef50\_A1R8S4 and UniRef50\_P0A0B0, methionine aminopeptidase; UniRef50\_G5MNH6 and UniRef50\_G5S167, endopeptidases) were higher in FC and NFC compared with both final products and raw materials (Fig. 2B).

In order to explore the functional potential of the microbiome in greater detail, we predicted microbial genes from assembled metagenomes and mapped them to a custom database including genes related to the resistance to stresses during the gastrointestinal transit, such as the resistance to acids and bile salts, and to the ability of strains to persist in the gut, such as the adhesion to the gut epithelium. Several genes related to stress resistance were enriched in NFC surfaces, final products, and raw materials, while genes related to adhesion prevailed on FC and operators' swabs (Fig. 3A). Notably, the choloylglycine hydrolase gene, coding for a hydrolase capable of degrading bile salts, urease-encoding genes, needed to counteract acid stress, and some genes involved in the adhesion to the gut epithelium (adhesin, mucus-binding protein) were higher in FC surfaces compared to the final product microbiome, but were also present in NFC surfaces. (Fig. 3B). These results suggest that FC and NFC surface microbiomes are well equipped with traits that may give potential ability to survive during the gastrointestinal transit and persist in the gut, when these microbes are transferred to the food products and ingested.

### The dairy plant microbiome shows a potential protective role through bacteriocin production

To understand if the microbiome within dairy industries hosts microorganisms able to produce bacteriocins, we mapped the contigs against the bacteriocin database distributed with the BAGEL4 tool. A total of 347 genes were detected across all the samples. A clustering between FC/NFC surfaces and ingredients/cheeses can be seen according to the presence-absence profiles of the bacteriocins (Supplementary Fig. 3). We then used the  $\chi^2$  test on the presence/absence matrix to understand which genes were significantly more prevalent in surfaces or food samples. Excluding the significant genes



**Fig. 3 | Genes related to gastrointestinal tract transit stress resistance and engraftment encoded in the metagenomes.** A Heatplot showing the abundance (Reads per Kilobase per Million, RPKM) of predicted genes related to gastrointestinal tract transit stress resistance and engraftment. **B–F** Boxplots showing the abundance (log RPKM) in the different sample groups of specific genes. Boxes represent the interquartile range (IQR) between the first and third quartiles, and the line inside represents the median (2nd quartile). Whiskers denote the lowest and the

highest values within  $1.5 \times$  IQR from the first and third quartiles, respectively. The significance was tested by applying pairwise Wilcoxon test. The category “cheese-related materials” groups together milk, brine and whey culture. Average values are obtained from  $n = 308, 199, 37, 524,$  and  $182$  biologically independent samples from food contact, non food contact, operators’ swabs, final products, and cheese-related materials, respectively.

that were present only in one sample ( $n = 5$ ), 115 genes showed a FDR-corrected  $p$ -value  $< 0.05$  (Supplementary Table 2). We observed that a significantly higher number of bacteriocin-encoding genes were detected on FC/NFC surfaces compared with ingredients/cheeses (Fig. 4A).

Interestingly, linocin M18-like bacteriocin gene clusters were more prevalent on FC/NFC surfaces (present in 51.3% and 62.4% of the samples, respectively). This pattern was also apparent for corynazolicin- and zoocin A-like clusters. Furthermore, surfaces showed a higher proportion of putative and potentially uncharacterized bacteriocins compared with ingredients and cheeses. On the contrary, lactococcin A-, lactococcin B- and streptide-like clusters were more frequently detected in ingredients and cheeses (Fig. 4B).

### Higher levels of antibiotic-resistance and virulence-associated genes are present on surfaces than cheeses in the dairy plants

We detected 32 different AMR genes (AMRG) families in the dairy plant microbiome. Among them, genes associated with resistance to aminoglycosides were the most abundant, with an average CPM of  $7.37 \pm 25.86$ , followed by those associated with resistance to tetracyclines ( $8.59 \pm 44.13$  CPM) and beta-lactams ( $5.30 \pm 18.44$  CPM). A considerable difference was observed between the types of AMRGs associated with surfaces and foods. Indeed, the average CPM abundance of aminoglycoside resistance genes was 2.91 and 2.50 for cheeses and ingredients, respectively, but reached 11.19 and 15.78 in FC and NFC surfaces, respectively (Fig. 5A). The same pattern was observed with respect to other AMRG families, except those encoding tetracycline resistance, which were less abundant on processing plant surfaces than in final product samples (Fig. 5A).

To better understand the distribution of each taxon - AMRG family pair in the cheese production environment, we focused on the taxonomic assignment of the contigs encoding AMRGs (Fig. 5B). Two hundred and thirty-seven out of 292 AMRG-encoding contigs (81.16%) from *A. johnsonii* contained genes associated with beta-lactam resistance. Of these, 48.10 and 34.60% were detected on FC and NFC surfaces, respectively, with the remaining part found in ingredients and products. In addition, 180 and 135 out of 333 AMRG-carrying contigs taxonomically assigned to *Staph. equorum* contained genes associated with resistance to macrolides and fosfomycins, respectively. Of these, the largest proportion of the contigs were detected on FC/NFC surfaces (59.26% for 'fosfomycin' resistance genes and 76.11% for 'macrolide' resistance genes; Fig. 5B). Then, we computed Spearman's correlations between microbiome taxonomic profiles and AMRG abundance. Consistently, the relative abundance of *B. aurantiacum*, *Staph. equorum* and *A. johnsonii* were significantly correlated with AMRG families: *Staph. equorum* and *B. aurantiacum* were positively correlated with Fosfomycin and Macrolide AMRG families (Spearman's  $\rho = 0.68$  and  $0.67$  for *Staph. equorum*,  $0.64$  and  $0.62$  for *B. aurantiacum*, respectively; corrected  $p$ -values  $< 0.001$ ; Supplementary Fig. 4). Moreover, *A. johnsonii* relative abundance was positively correlated with beta-lactams (Spearman's  $\rho = 0.60$ ,  $p$ -value  $< 0.001$ ). Interestingly, this AMRG family was exclusively linked to this taxon (Supplementary Fig. 4).

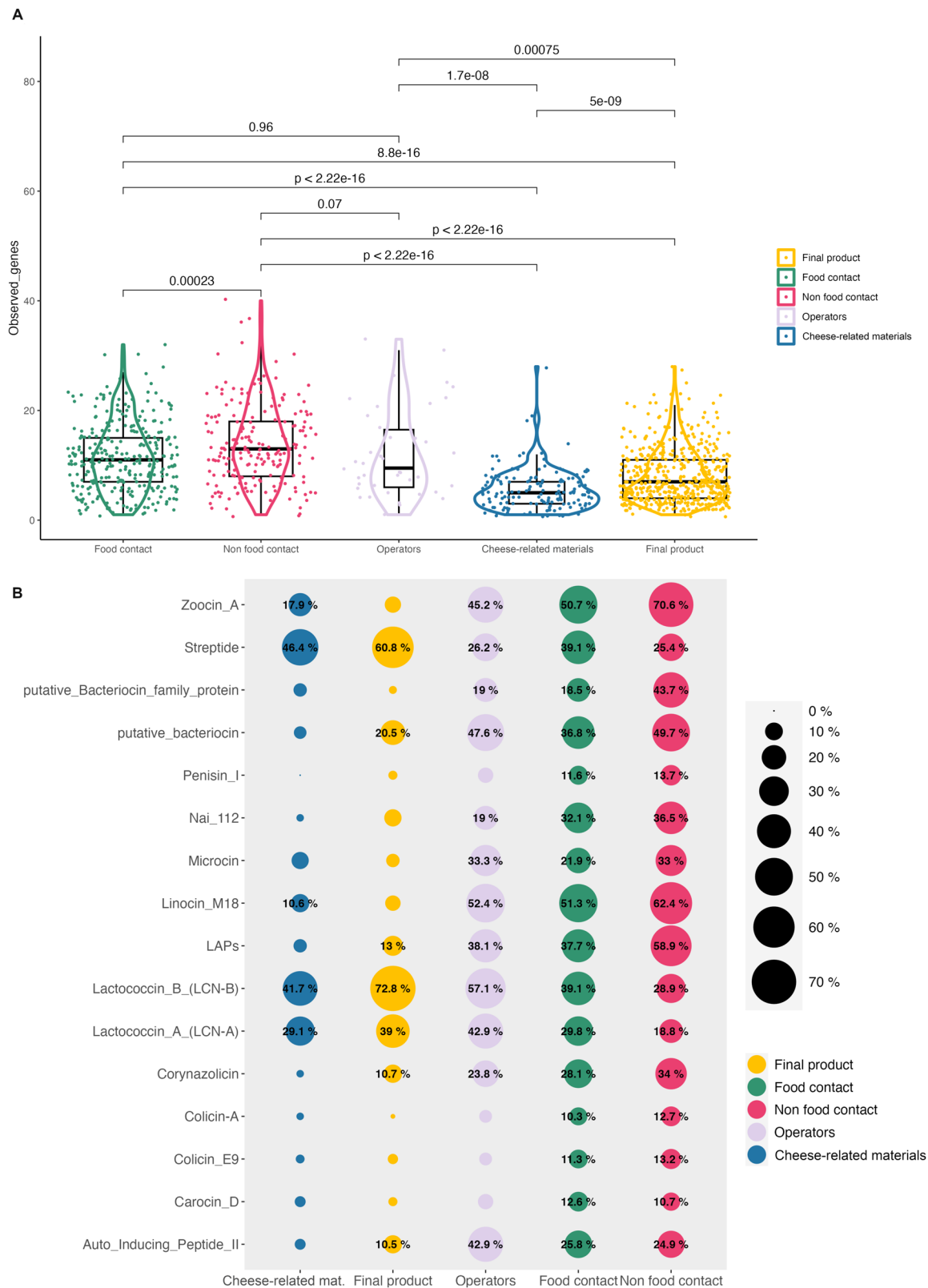
We also identified a total of 1056 genes predicted to belong to 13 virulence classes. The 'effector delivery system' virulence class was the most frequently detected, with 321 genes, followed by 'adherence' (167 genes) and 'nutritional/metabolic factor' (151 genes). Interestingly, the top 15 genes, with the highest average CPM abundance, were mostly predicted to be involved in biofilm formation and adherence (Supplementary Fig. 5) and were enriched on FC/NFC surfaces relative to products, suggesting a potential selection for adherence and biofilm formation traits in the cheese production environment. Indeed, the virulence gene showing the highest CPM was PA1464 (also reported as *cheW*), involved in flagellum motility. Consistently, the genes *flgG*, *flgI*, *fliP*, *fliG*, *fleN*, *fliM*, and *flgC*, encoding structural components of flagella or associated regulators, and the genes *pilG* and *pilH*, involved in type IV PilA synthesis, were enriched on FC/NFC surfaces compared with cheeses/ingredients.

The existence of different virulence traits harbored by communities from FC/NFC surfaces and ingredients/products was further confirmed by Non-Metric Multidimensional Scaling (NMDS) (Supplementary Fig. 6; PERMANOVA  $p$ -value  $< 0.001$ ). To better understand which virulence classes contributed to the separation of surfaces and foods, we compared the virulence family level CPM values between the sample categories. This analysis highlighted that the 'Adherence', 'Motility' and 'Biofilm' virulence classes were overrepresented on FC/NFC surfaces compared with ingredients/products (Wilcoxon's rank sum test  $p$ -value  $< 0.001$ ; Fig. 5C).

### The dairy plant microbiome includes previously unidentified species

We reconstructed a total of 9559 MAGs, which were clustered in 2111 SGBs (Supplementary Table 3). Approximately 63% of the MAGs (6043) corresponded to SGBs with  $\geq 10$  MAGs ( $n = 171$ ). These prevalent SGBs were considered for further analyses. Among the 171 prevalent SGBs, 130 showed  $< 0.05$  MASH genomic distance with one known genome from an isolate or MAG (known SGBs, kSGBs), while 41 were considered as potentially new species (unknown SGBs, uSGBs, with  $> 0.05$  distance from any MAG or isolate genome). Most of the MAGs (49%) were representative of the phylum Firmicutes (recently reclassified as Bacillota), followed by Actinobacteria (Actinomycetota; 39%) and Proteobacteria (Pseudomonadota; 11%). Bacillota MAGs dominated FC surfaces and the unripened cheeses, while Actinobacteriota genomes were mainly reconstructed from raw materials and operator's swabs (Fig. 6). Interestingly, although 3,028 MAGs were reconstructed from NFC surfaces, only 200 (about 7%) belonged to prevalent SGBs ( $> 10$  MAGs), indicating the presence of several low-abundance species on NFC (Fig. 6). Conversely, although a similar number of MAGs was reconstructed from FC ( $n = 3043$ ), about 58% ( $n = 1783$ ) belonged to prevalent SGBs. With respect to uSGBs, 26 (out of 41) were classified at phylum level as being representative of Actinomycetota, including approximately 71% of unidentified MAGs. In contrast, 7 and 6 uSGBs were identified as Pseudomonadota and Bacillota, respectively, while only one was representative of Thermus. SGBs representative of LAB were among the most abundant (Fig. 7A): *Lc. lactis* (two SGBs; SGB\_73,  $n = 463$  and SGB\_92,  $n = 324$ ), *Strep. thermophilus* (SGB\_12,  $n = 360$ ), *Leuconostoc mesenteroides* (SGB\_77,  $n = 138$ ), *Lactocaseibacillus paracasei* (SGB\_15,  $n = 129$ ), *Lactobacillus delbrueckii* (SGB\_0,  $n = 96$ ). In addition, several Actinomycetota SGBs were also frequently identified, such as *B. aurantiacum* (SGB\_62,  $n = 186$ ), *K. salsicia* and *K. palustris* (SGB\_156,  $n = 111$  and SGB\_11,  $n = 86$ ), *Brachybacterium alimentarium* (SGB\_64,  $n = 89$ ), and *C. casei* and *C. variabile* (SGB\_228,  $n = 75$  and SGB\_144,  $n = 74$ ). All these SGBs contained MAGs reconstructed from cheeses and surfaces, both FC and NFC (Fig. 7A). However, unsurprisingly, some differences in the prevalence in the different sample types were observed. *Kocuria* spp. (*K. salsicia*, *K. palustris* and *K. carniphila*) MAGs were typical of environmental samples, mainly being reconstructed from FC and NFC surfaces, as well as from operators' swabs (Fig. 7B). Conversely, *Staph. equorum* (SGB\_29,  $n = 249$ ), *Staph. saprophyticus* (SGB\_1059,  $n = 58$ ), *B. aurantiacum*, *C. casei*, *C. variabile* and *B. alimentarium* prevailed in FC, NFC and in ripened cheeses. With respect to the final products, SGBs identified as mesophilic LAB, such as *Lactocaseibacillus paracasei*, *Lactiplantibacillus plantarum* (SGB\_185,  $n = 75$ ), *Levilactobacillus brevis* (SGB\_1886,  $n = 56$ ), *Latilactobacillus curvatus* (SGB\_72,  $n = 46$ ) and *Lentilactobacillus osakiensis* (SGB\_545,  $n = 41$ ), were always more prevalent in ripened than unripened cheeses (Fig. 7B).

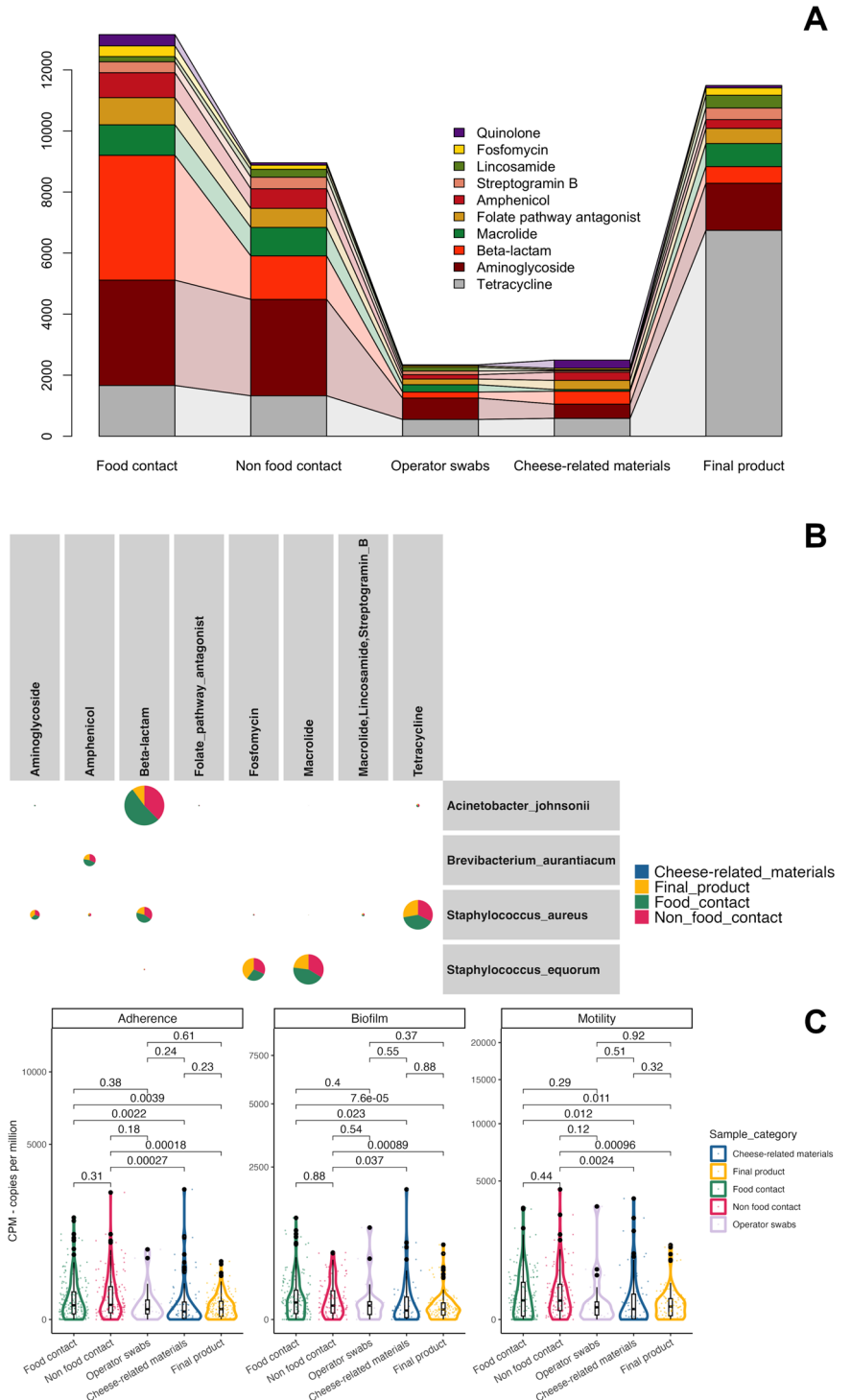
We further screened MAGs for the presence of AMR or bacteriocin-encoding genes (Fig. 8). A total of 554 MAGs contained putative AMRG. Among them, 174 (31.4%) encoded resistance to at least 2 classes of antimicrobial compounds. Most of the potentially antimicrobial resistant MAGs were reconstructed from surfaces (FC,  $n = 224$ , 40.4%; NFC,  $n = 158$ , 28.5%), and were identified as *Staph. equorum* (SGB\_29,  $n = 57$ ), *A. johnsonii* (SGB\_67,  $n = 23$ ) and *A. guillouiae* (SGB\_883,  $n = 18$ ). In particular, *Staph. equorum* represented the most abundant potentially antimicrobial resistant SGB on FC, NFC and final cheese samples, accounting for 20.7%,



**Fig. 4 | Analysis of the bacteriocins encoded in the metagenomes.** The category “cheese-related materials” groups together milk, brine and whey culture. **A** Boxplot reporting the number of observed bacteriocin genes for each sample. **B** Genes observed with a significantly higher frequency in one of the sample groups. Percentages report the proportion of samples from each category testing positive for the selected genes. Boxes represent the interquartile range (IQR) between the first and

third quartiles, and the line inside represents the median (2nd quartile). Whiskers denote the lowest and the highest values within  $1.5 \times$  IQR from the first and third quartiles, respectively. The significance was tested by applying pairwise Wilcoxon test. Average values are obtained from  $n = 308, 199, 37, 524,$  and  $182$  biologically independent samples from food contact, non food contact, operators’ swabs, final products and cheese-related materials, respectively.

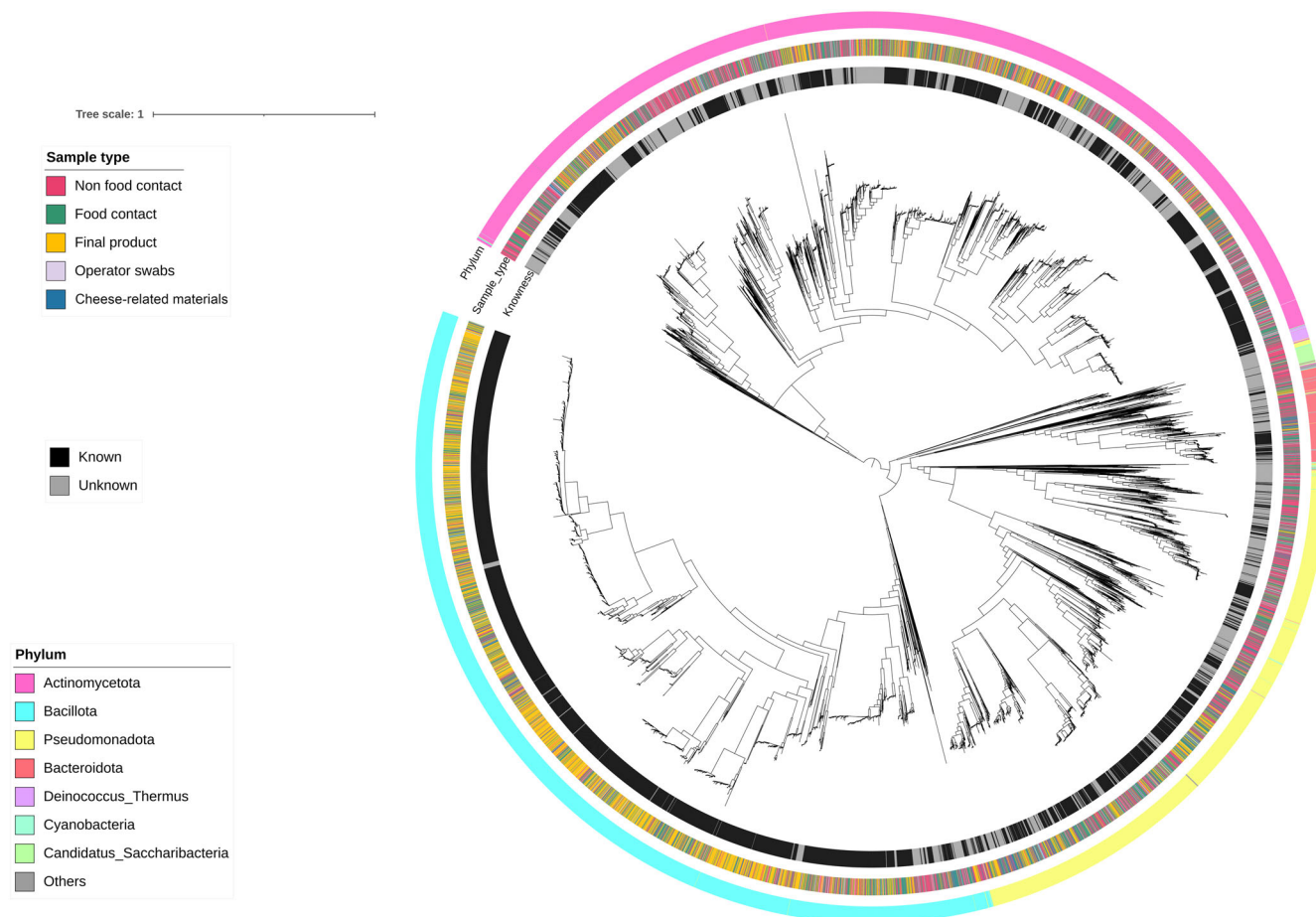
**Fig. 5 | Analysis of antimicrobial-resistance and virulence-associated genes.** The category “cheese-related materials” groups together milk, brine and whey culture. **A** Connected barplot showing the variation in the average Copies Per Million (CPM) abundance of each Antimicrobial Resistance Gene (AMRG) family between the classes of samples. **B** Taxonomic assignment of contigs encoding for AMRG. For each AMRG - taxon pair, each slice is proportional to the percentage of contigs reconstructed from a category of samples. The size of each pie is proportional to the number of contigs linked to an AMRG - taxon pair. **C** Sum of the CPM abundance of genes linked with Adherence, Biofilm formation and Motility virulence traits. Statistical differences between the groups were calculated through the Wilcoxon’s rank-sum test. Boxes represent the interquartile range (IQR) between the first and third quartiles, and the line inside represents the median (2nd quartile). Whiskers denote the lowest and the highest values within 1.5 x IQR from the first and third quartiles, respectively. The category “cheese-related materials” groups together milk, brine and whey culture. Average values are obtained from  $n = 308, 199, 37, 524,$  and  $182$  biologically independent samples from food contact, non food contact, operators’ swabs, final products and cheese-related materials, respectively.



32.8% and 12.3% of the total resistant MAGs from each group of samples, respectively (Fig. 8A). Although antimicrobial resistant *Staph. equorum* MAGs were most abundant overall, only 57 out of 257 *Staph. equorum* MAGs (22.2%) harbored at least one AMRG. Among these MAGs, genes predicted to encode resistance to macrolides (8.56%), fosfomicin (5.4%), tetracyclines (5.4%) and beta-lactams (5.1%; Fig. 8A) were most common. Conversely, a high proportion of *A. johnsonii* and *A. guillouiae* MAGs contained putative antimicrobial resistance genes (47.9% and 66.7%, respectively). As expected, almost all of the MAGs from both species contained beta-lactam resistance genes, and 33.3% of the *A. guillouiae* MAGs also contained genes associated with aminoglycoside resistance (Fig. 8A).

3415 MAGs harbored at least 1 gene linked with the production of bacteriocins (as classified in BAGEL4 database): 1098 were reconstructed from FC (32.2%) and 959 from NFC (28.1%), while 1006 and 203 were detected in MAGs from final products and ingredients, respectively. These MAGs mainly belonged to *Strep. thermophilus* (SGB\_12,  $n = 234$  genomes), the two *Lc. lactis* SGBs (SGB\_92,  $n = 234$ ; SGB\_73,  $n = 154$ ), *Staph. equorum* (SGB\_29,  $n = 193$ ), *B. aurantiacum* (SGB\_62,  $n = 125$ ) and *Lactocaseibacillus paracasei* (SGB\_15,  $n = 99$ ). More than 50% of the genomes containing bacteriocin-associated genes found in final cheeses belonged to the SGBs reported above. However, the same SGBs accounted for ~36% of the total MAGs harboring potential bacteriocin-producing genes found on surfaces,





**Fig. 6 | Phylogenetic tree of Metagenome-Assembled Genomes (MAGs) reconstructed from cheeses and dairy environment.** Phylogenetic tree of all the MAGs reconstructed in this study, spanning 2111 Species-level Genome Bins (SGBs). From outer to inner, rings are colored according to phylum-level taxonomic assignment, sample type and identification of SGBs at species level (as reported in the Methods).

The category “cheese-related materials” groups together milk, brine and whey culture. MAGs have been reconstructed from  $n = 308, 199, 37, 524,$  and  $182$  biologically independent samples from food contact, non food contact, operators’ swabs, final products, and cheese-related materials, respectively.

highlighting that some minor and less prevalent species might contribute to bacteriocin secretion on FC surfaces (Fig. 8B).

Bacteriocin production was a prevalent trait in all the prevalent SGBs. Indeed, 44% and 20% of the *Lactocaseibacillus paracasei* MAGs harbored genes linked with carnocin and enterolysin A biosynthesis, respectively, while 30% of *Strep. thermophilus* (SGB\_12) MAGs harbored genes linked with the synthesis of streptide (Fig. 8B).

#### Different strains are selected in dairy facilities producing the same cheese type

Finally, we looked for genomics variations within the most abundant SGBs. Considering those cheese types for which we sampled several different facilities, we identified in several SGBs the presence of facility-specific strains. For *Lactobacillus delbrueckii* (SGB\_0), we identified the presence of two putative strains, but MAGs did not cluster according to the cheese or the sample type (Fig. 9A). However, when considering only MAGs from Caciocavallo dairies (an Italian, ripened pasta-filata cheese), we found that MAGs reconstructed from samples collected in the same factory showed a lower ANI distance and were phylogenetically closer compared with those from different facilities, suggesting the presence of different strains (Fig. 9B-C). This facility-driven clustering of MAGs was observed also for other prevalent SGBs: *Strep. thermophilus* (SGB\_12) in Afuega'l Pitu (a Spanish soft-ripened cheese made from acid-coagulated curd) (Supplementary Fig. 7), *Lc. lactis* (SGB\_73) in Casin (a Spanish not smear-hard ripened “pasta amasada” – i.e., kneaded curd – cheese) and Caciocavallo (Supplementary

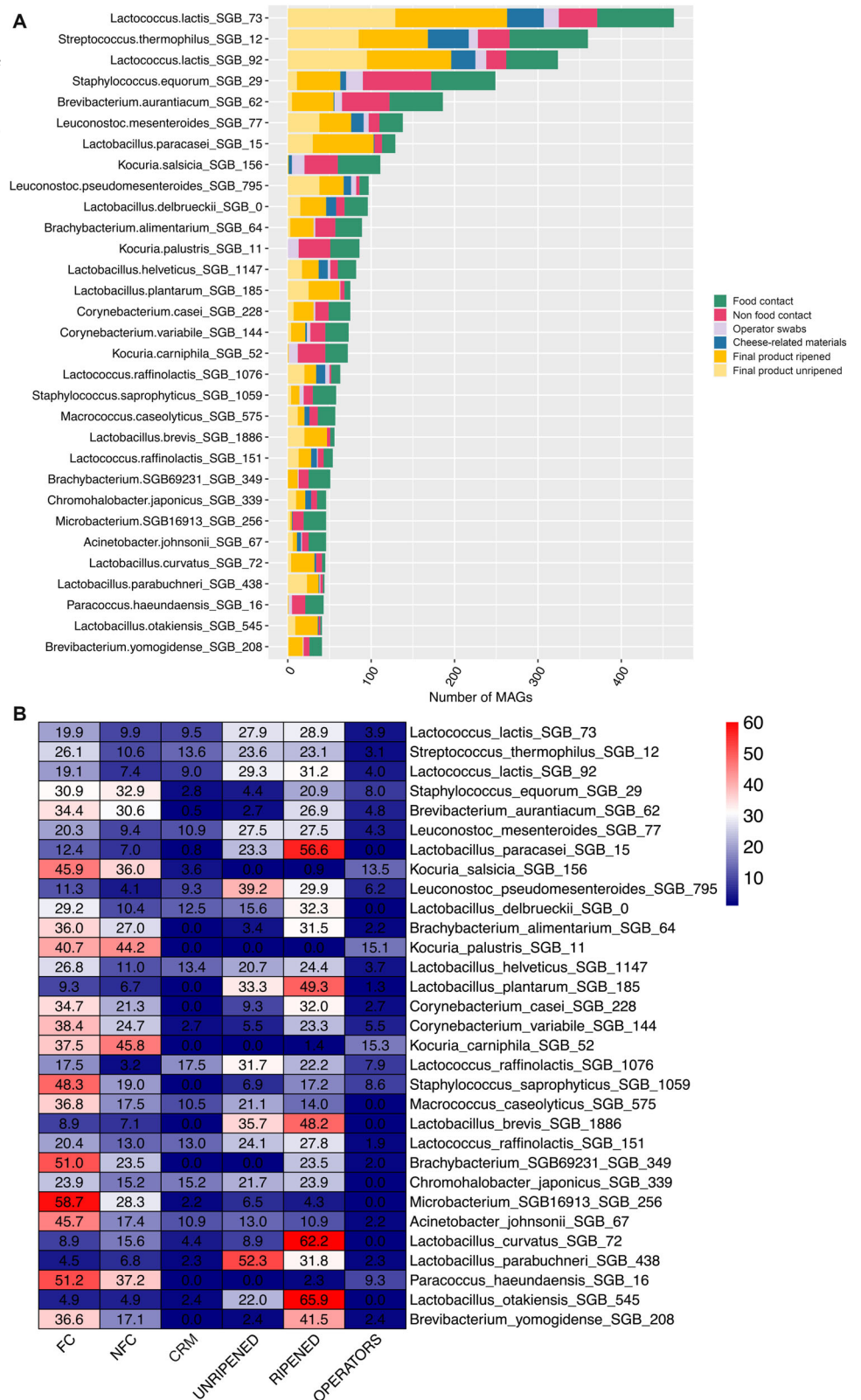
Fig. 8) and *Leuc. mesenteroides* (SGB\_77) in Casin (Supplementary Fig. 9) dairies.

#### Discussion

The surfaces of food industries are inhabited by a complex microbiome, well adapted to the specific processing conditions and microenvironments, that can be considered as residential<sup>24</sup>. Several of these microbes can form biofilms on the surfaces present in facilities, where microbes are embedded in the self-produced extracellular polymeric substances<sup>19,25</sup>. Throughout the food chain, equipment and food contact surfaces provide suitable reservoirs for biofilm formation and accumulation. From these surfaces microbes can be easily transferred to the food product during handling, manufacturing, and storage. In addition, surfaces that are not in direct contact with foods are reservoirs of microbes, which may reach the product. Therefore, the microbiome of the food industry can be regarded as a primary source of microbes for the product. In light of this, mapping its variability across different industries, as well as understanding which microbes, and potential activities encoded within their genomes, prevail in them is of utmost importance in order to improve food quality and safety and reduce food loss in a sustainability perspective.

Several other studies have focused on the microbiome in dairy industries, although the number of companies sampled has always been limited<sup>26–28</sup>. Here we discuss the largest dairy microbiome mapping ever performed, obtained by spanning 73 cheese-making facilities located in 4

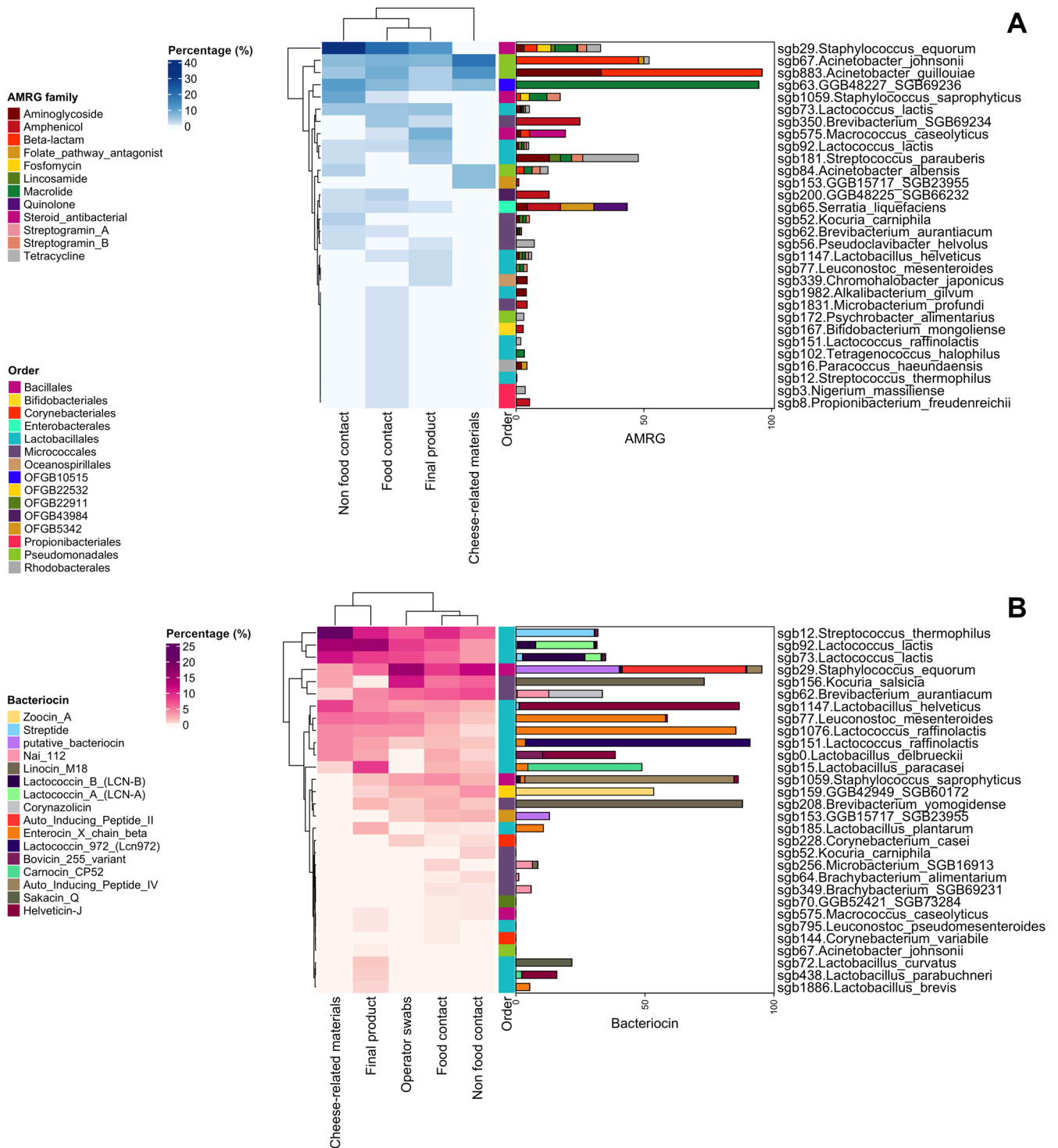
**Fig. 7 | MAGs distribution in cheeses and dairy environment.** **A** Bar chart showing the number of Metagenome-Assembled Genomes (MAGs) for the top 30 Species-level Genome Bins (SGBs) reconstructed for each sample group. **B** Heatplot reporting, for each of the top 30 SGBs, the proportion (%) of MAGs reconstructed from each sample group. FC, food-contact surfaces; NFC, non food-contact surfaces; Ripened, ripened cheeses (>30 days); Unripened, unripened cheeses (<30 days); CRM (cheese-related materials), that groups together milk, brine, and whey culture.



different EU countries, producing different types of cheeses and leading to microbiome assessment in 1250 samples. Moreover, the study was performed using a common, standardized and validated protocol for sample collection, processing, sequencing, and data analysis<sup>29</sup>.

We demonstrated that the surfaces of dairy industries harbor a complex microbial community, with > 3000 different species, several of

them never identified previously. Consistently with previous reports<sup>27,30,31</sup>, Bacillota, and LAB in particular, can be regarded as part of a core microbiome present in all the dairies and all the different surfaces sampled, with higher relative abundance on food contact surfaces. However, several Actinomycetota (e.g., *Brevibacterium*, *Corynebacterium*, *Kocuria*), Pseudomonadota (e.g. *Acinetobacter*) and non-LAB

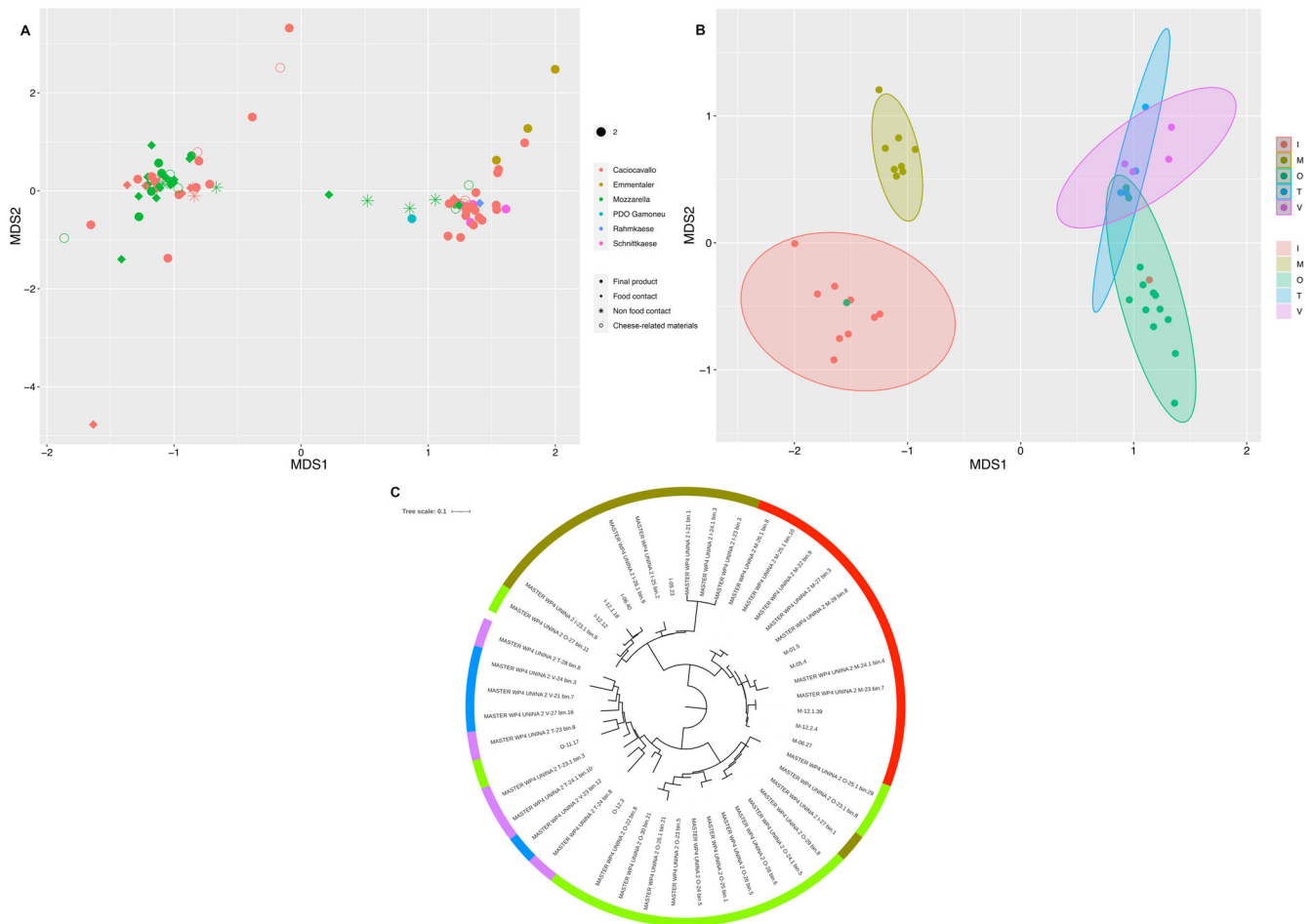


**Fig. 8 | MAGs harbor Antimicrobial Resistance and bacteriocin coding genes.** For each group of samples, the heatmaps show the proportion (%) of Metagenome-Assembled Genomes (MAGs) within a Species-level Genome Bin (SGB) containing at least 1 Antimicrobial Resistance (AMR; A) or bacteriocin (B) coding gene. The

side color bar is colored according to order-level taxonomic assignment of the SGB. The bar charts represent the proportion (%) of each AMR or bacteriocin class within the SGB. The category “cheese-related materials” groups together milk, brine, and whey culture.

Bacillota (e.g. *Staph. equorum*) were also identified both in FC and NFC surfaces, as well as in the final products. *Brevibacterium*, *Corynebacterium*, and *Staph. equorum* are well-known inhabitants of the microbiome of surface-ripened cheeses. They actively participate in their ripening<sup>32,33</sup> and have been previously reported as inhabitants of ripening rooms in traditional cheese production plants<sup>34,35</sup>. *Kocuria* genus includes potential pathogenic species causing bacteremia,

previously isolated from cheeses and cheese brine<sup>36,37</sup>, showing the ability to form biofilm<sup>38</sup> and sometimes reported as carriers of AR genes<sup>38</sup>. In addition, some possible pathogens (e.g., *Staph. aureus*, *A. johnsonii*) were also identified at low abundance, highlighting their ability to persist in the environment and their possible transfer from the surfaces to the cheeses. This also shows the potential of our approach for epidemiological investigations aimed to track the origin of pathogenic strains



**Fig. 9 | MAGs of *Lb. delbrueckii* can be clustered in different putative strains.** Non-metric Multidimensional Scaling (NMDS) based on ANI distance matrix of MAGs belonging to SGB\_0 (*Lb. delbrueckii*). **A** All MAGs in SGB\_0 are shown. Points are colored according to the cheese type, while different shapes indicate the

sample types. **B** Only MAGs reconstructed from Caciocavallo cheese facilities are included and points are colored according to the facility code. **C** Phylogenetic tree of MAGs belonging to SGB\_0 reconstructed from Caciocavallo cheese facilities. The category “cheese-related materials” groups together milk, brine and whey culture.

involved in foodborne outbreaks. However, we also highlighted that pathogens co-exist in the environment with other taxa (e.g., LAB) suggesting the presence of an equilibrium that limits their proliferation. Indeed, a wide range of bacteriocin-producing microbes was identified on cheese industry surfaces (e.g., LAB, *Staph. equorum*, *B. aurantiacum*), harboring genes coding for the biosynthesis of different classes of bacteriocins, including linocin, corynazolicin, zoocin, and lactococcins that were the most common. Genes coding for linocin biosynthesis were particularly abundant on FC and NFC surfaces. Linocin M18 was first identified in *B. linens*, and several in vitro tests highlighted its ability to inhibit the growth of *Listeria monocytogenes*<sup>39</sup>. Notably, this pathogen was not detected in any of the samples analyzed. The production of bacteriocins, besides other mechanisms (e.g., competition for resources, production of other antagonistic substances, development of cheese flavor<sup>40,41</sup>), may exert a protective role, limiting the proliferation of spoiling and pathogenic species within the food industry microbial community. Indeed, the selection of beneficial microbes in food industry surfaces may represent an ecologic and sustainable hurdle against pathogens and spoilers potentially residing in the processing plants, helping in reducing the use of chemicals for disinfection.

Our results also highlighted that the microbiome on dairy surfaces may be a source of technologically relevant activities. Indeed, several genes coding for peptidases, as well as related to aminoacid degradation pathways were enriched on the surfaces, suggesting that surfaces can be an important source of microbes and metabolic activities that

are important for cheese ripening and flavor production, as often reported<sup>42</sup>.

We also showed that the resident microbiome in dairy plants harbors genes related to the ability to potentially survive the stress conditions encountered during the gastrointestinal tract transit (e.g., acids, bile salts) and to engraft in the gut, becoming part of the gut microbiome. Previous reports suggested that LAB strains from fermented foods show similarity with those present in the human gut<sup>43</sup> and that they have the genomic repertoire that allow them to persist in the gut<sup>44</sup>. Our study firstly highlighted that this can be translated also to the dairy plant microbiome, that, once transferred to the products, may represent a source of microbes for the human gut. This can be particularly of concern since we also found that the microbiome in dairy surfaces also harbors a wide pattern of AMR genes, where tetracycline, beta-lactam, and amphenicol resistance genes are the most abundant. Interestingly, these are the most common classes of AMR genes found in pathogenic strains isolated from milk and in milk microbiome<sup>45–47</sup>, as well as the residual antibiotics commonly found in retail milk from different countries<sup>48</sup>. Recently, fermented foods, particularly cheeses, were highlighted as a hotspot for AMR transfer<sup>49</sup>. Our study confirms this evidence and highlights that the food processing environment may be involved in this process. Indeed, our results suggest that the FC and NFC surfaces in dairies have a complex resistome that may reach the final product, and, upon ingestion, AMR genes may be transferred to the gut microbiome or to pathogens through horizontal gene transfer events.

Finally, although most of the dairy plants we sampled produce PDO cheeses and therefore apply the same standardized production process, we demonstrated that a strain-level selection may occur in different facilities producing the same cheese (e.g., the Italian PDO Caciocavallo Silano; the Spanish PDO “Afuega'l Pitu” and PDO “Casin”). Different strains may come from raw materials (e.g., milk) or starter cultures used during manufacturing (commercial or natural starters), but their presence in environmental swabs suggests that these strains are able to persist in the dairy plant and take over during the manufacture and ripening process. Strain diversity in cheeses may lead to different organoleptic attributes<sup>50,51</sup>. Therefore, these results support the idea that cheeses from different producers may have peculiar traits. Moreover, mapping strain diversity across PDO cheese producing plants may be useful to design novel strategies for origin tracking based on microbiome fingerprinting. However, we have to point out some limitations of our study that might have partially influenced the results. Firstly, microbial biomass was not quantified, so the comparison of results between samples with high (e.g., cheeses) and low (e.g., surfaces) biomass might be biased. In addition, our study is based on DNA analysis. Therefore, we cannot exclude that part of this DNA might come from dead or damaged cells.

We provided a comprehensive mapping of the microbiome across dairy plants in different EU countries. Our study demonstrated that the cheese industry harbors a highly complex microbiome, selected by the specific technological process and that can be transferred to the product. We highlighted that it may exert a positive role, actively participating in cheese flavor production and inhibiting pathogen development. However, the results also suggested that dairy industry surfaces can be an important reservoir of AR strains that may contaminate the product and survive during the gastrointestinal passage, reaching the gut microbiome.

The integration of procedures for microbiome mapping in the dairy industry can support the overall quality management strategy. This may help to improve cheese quality and safety, without compromising the specificity of each product, the link with traditional practices and with the geographical area of production.

## Methods

### Sampling, DNA extraction, and metagenome sequencing

Seventy-three facilities producing different types of cheeses and located in 4 EU countries (Austria,  $n = 6$ ; Ireland,  $n = 15$ ; Italy,  $n = 16$ ; Spain,  $n = 36$ ) were visited (February–October 2020) after the completion of routine cleaning procedures (Supplementary Table 4). Surfaces, cheese-related materials (milk, whey culture, brine) and final products (fresh, medium and long-ripened cheeses) were collected according to the MASTER SOP-01 (<https://www.master-h2020.eu/SOPS.ht ml>, ref. 29). For all cheeses except Buffalo Mozzarella, samples of cheese rind and core were analyzed separately. Food contact (FC) and non-food contact (NFC) surfaces from the facilities were sampled using Whirl-Pak Hydrated PolyProbe swabs (Whirl-Pak, Madison, Wisconsin, US), covering an area of about 1 m<sup>2</sup>, or a sampling unit (e.g., one table, one sink). FC included (according to the type of cheese and the industry set up) draining tables, curd vats, curd shredders, molding machines and ripening shelves, while NFC included walls, floors, sinks and drains. When available, different areas in the facility were sampled: processing, ripening, and packing rooms. In addition, swabs were collected from the hands/aprons of the employees working on the production line. Five swabs from each sampling point were collected. A list of the samples collected in each country is provided as Supplementary Table 5. In the laboratory, the 5 swabs from each surface were pooled together and 10 mL of Phosphate Buffered Saline (PBS) 1X were added to each pool. After homogenization in a stomacher, the cell suspension was collected and centrifuged at 5000 x g for 15 minutes. DNA extraction was performed from the cell pellets using the PowerSoil Pro Kit, adopting a previously validated modified version of the standard protocol optimized for food processing environments and low biomass samples<sup>29</sup>. Metagenomic libraries were prepared using the Nextera XT Index Kit v2 (Illumina, San Diego, California, United States), and metagenome

sequencing was performed on an Illumina NovaSeq platform, leading to 2 × 150 bp reads. Negative controls were collected both at the industry and in the laboratory, leaving a pool of 5 swabs exposed to the air for 1 minute. Negative controls were then processed as reported for the other samples. However, DNA yield was really low (< 1 ng/μl), and library preparation failed for all the negative controls.

### Reads pre-processing, taxonomic, and functional profiling

The raw reads were filtered using a validated pipeline (available at <https://github.com/SegataLab/preprocessing>) composed of three sequential steps: i) discarding of low-quality (quality < 20), short (L < 75 bp) and with-ambiguous-nucleotides (n < = 2) reads using Trim Galore v0.6.6 ([https://www.bioinformatics.babraham.ac.uk/projects/trim\\_galore/](https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/)); ii) identification and removal of human (HG19 human genome release) and bacteriophage phiX174 DNA (Illumina spike-in) contamination by mapping reads against target reference genomes through BowTie2 v2.2.9<sup>52</sup> (with parameter `-sensitive-local`) and iii) splitting of remaining reads into forward, reverse and unpaired files).

Taxonomic and functional profiles were obtained using short reads with MetaPhlan<sup>53</sup> (version 4.0.2 (qstat = 0.2) with the marker database ChocoPhlan v202204) and HUMAnN<sup>54</sup> (version 3.0, options `--metaphlan-options "-t rel_ab --index v30_CHOCOPPhlan_201901"`), respectively.

### Assembly, Metagenome-Assembled Genomes (MAGs) reconstruction and analysis

High-quality reads were assembled independently using MEGAHIT<sup>55</sup> v. 1.2.1 (with default options) and contigs > 1000 bp were used to predict genes by using MetaGeneMark<sup>56</sup> v. 3.26. Predicted genes were aligned (using DIAMOND<sup>57</sup> v. 2.0.4 and the option `-very-sensitive`), against a custom database including known microbial genes potentially involved in stress resistance and adhesion to the gut epithelium<sup>58</sup>. Accession numbers and gene identification are reported in Supplementary Table 6. An e-value cutoff of 1e<sup>-5</sup> was applied, and a hit was required to display > 95% of identity over at least 50% of the query length.

Metagenome assemblies were processed through TORMES<sup>59</sup> v. 1.3.0 (options `--gene_min_id 80 --gene_min_cov 80`) to detect antimicrobial resistance (AMR) and virulence genes. For AMR identification, TORMES relies on three databases, i.e., ResFinder<sup>60</sup>, CARD<sup>61</sup> and ARG-ANNOT<sup>62</sup>, whereas virulence genes were detected through the VFDB<sup>63</sup>. Only hits showing >80% identity over >80% of the query length were retained.

To estimate the AMR/virulence genes abundance, raw reads were filtered against the ResFinder database through BowTie2<sup>52</sup> v2.2.9 using the parameters reported above. The resulting sam files were filtered through an in-house script ([https://github.com/SegataLab/MASTER-WP5-pipelines/blob/master/07-AMR\\_virulence\\_genes/count\\_reads.rb](https://github.com/SegataLab/MASTER-WP5-pipelines/blob/master/07-AMR_virulence_genes/count_reads.rb)), then Copies Per Million (CPM) were computed by normalizing the n. of mapped bacterial reads (as estimated by viromeQC<sup>64</sup>) for the total number of reads in the metagenome. In addition, BAGEL4<sup>65</sup> was used to identify the genes predicted to encode bacteriocins in the metagenomes.

To obtain the gene abundance, short reads were mapped to the genes using BowTie2 (options: `-very-sensitive-local -no-unal`) and the number of mapped reads was normalized using the RPKM method (reads per kilobase per million mapped reads), considering the formula (number of hits for each gene/gene length)/total number of mapped reads per sample<sup>66</sup>.

Contigs shorter than 1,000 bp were discarded. The coverage for the remaining contigs was calculated by aligning them against the original reads with BowTie2 v2.2.9 and used for binning through MetaBAT 2<sup>67</sup> v. 2.12.1. The quality of Metagenome-Assembled Genomes (MAGs) was estimated with CheckM<sup>68</sup> v. 1.1.3: only MAGs with > 50% completeness and < 5% contamination were retained for further analyses. MAGs included in this study were included in a > 1 M genomes database, MetaRefSGB vMar22, where they were clustered according to genetic distances. These MAGs were part of a wider catalog recently developed within the MASTER project. Pairwise genetic distances between all genomes were calculated using MASH<sup>69</sup> (version 2.0, option `"-s 10,000"` for sketching). A MASH distance <

5% from any of the database genomes was considered to place the MAG within the relative species-level genome bin (SGB). When a MAG showed > 5% distance from any of the reference genomes, it was considered a novel species (unknown SGB, uSGB). In this case, the taxonomic assignment was made at genus (>5 and <15% distance), family (> 15 and < 25% distance), or phylum (> 25% distance) level. Thresholds previously reported were used<sup>70</sup>. To define subspecies within each SGB, pairwise average nucleotide identities (ANI) distances between MAGs were calculated through FastANI<sup>71</sup> (using default options), while phylogenetic analyses were performed using PhyloPhlAn<sup>72</sup> (options --diversity low and --accurate; version 3.0.67). Phylogenetic trees were visualized using iTol<sup>73</sup>.

### Statistical analyses

All the statistical analyses were performed in a R environment (<https://www.R-project.org/>; version 4.2.2). The relative abundances of taxonomic/functional profiles and the genes' CPM/RPKM were compared through the Wilcoxon-Mann-Whitney test using the R function 'wilcox.test'. Jaccard and Bray-Curtis distance matrices were computed with the function 'vegdist' ('vegan' package), then Principal Coordinates Analyses and Nonmetric Multidimensional Scaling (NDMS) were performed through 'cmdscale' and 'metaMDS', respectively (both from the 'base' R package). Permutational MANOVA (function 'adonis2' from the 'vegan' R package) highlighted if the groupings of samples according to the distance matrices were statistically significant. Pairwise comparisons between permutational MANOVA results were done with the function 'pairwise.perm.manova' from the 'RVAideMemoire' package. The  $\chi^2$  test was applied on contingency tables built from presence-absence matrices to detect genetic features significantly more prevalent in specific groups of samples. All the *p*-values were FDR-adjusted when needed. Also, Spearman's  $\rho$  between AMR-taxa and AMR-virulence abundances were computed with the function 'correlate' from the 'corr' R package and plotted using 'cor.graph' (for AMR-taxa correlations) and 'corrplot' (for AMR-virulence correlations). Bar/pie charts and box/scatter/violin/correlation plots were generated using the 'ggplot2' and 'ggpubr' R packages, whereas heatmaps were plotted through 'pheatmap' (using the euclidean distance and the complete-linkage clustering method).

### Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

### Data availability

The raw sequence reads generated from samples collected in Italy, Spain and Austria have been deposited in the Sequence Read Archive (SRA) of the NCBI under accession number PRJNA997821, whereas those from Irish facilities are available in the European Nucleotide Archive database under accession number PRJEB63604. All software used are freely available for download.

Received: 19 March 2024; Accepted: 23 July 2024;

Published online: 02 August 2024

### References

- Fox, P. F. & McSweeney, P. L. H. *Cheese*. p. 5–21 (Elsevier, 2017).
- Wei, W. & Yano, H. Development of “new” bread and cheese. *Processes* **8**, 1541 (2020).
- OECD/FAO. Per capita consumption of processed and fresh dairy products in milk solids, in *OECD-FAO Agricultural Outlook 2021–2030*. <https://doi.org/10.1787/42d5ba73-en> (OECD Publishing, Paris, 2021).
- Martin, N. H., Torres-Frenzel, P. & Wiedmann, M. Invited review: controlling dairy product spoilage to reduce food loss and waste. *J. Dairy Sci.* **104**, 1251–1261 (2021).
- Martinez-Rios, V. & Dalgaard, P. Prevalence of *Listeria* monocytogenes in European cheeses: a systematic review and meta-analysis. *Food Control* **84**, 205–214 (2018).
- de Oliveira, C. A. F., Corassin, C. H., Lee, S. H.I., Gonçalves, B. L. & Barancelli, G. V. *Nutrients in Dairy and their Implications on Health and Disease*. p. 61–75 (Elsevier, 2017).
- Rajmohan, S., Dodd, C. E. R. & Waites, W. M. Enzymes from isolates of *Pseudomonas fluorescens* involved in food spoilage. *J. Appl. Microbiol.* **93**, 205–213 (2002).
- Marchand, S. et al. Heterogeneity of heat-resistant proteases from milk *Pseudomonas* species. *Int. J. Food Microbiol.* **133**, 68–77 (2009).
- Longhi, R. D. et al. *Pseudomonas fluorescens* and *Pseudomonas putida* from refrigerated raw milk: genetic diversity and lipoproteolytic activity. *J. Dairy Res.* **89**, 86–89 (2022).
- Baglinière, F. et al. Proteolysis of ultra high temperature-treated casein micelles by AprX enzyme from *Pseudomonas fluorescens* F induces their destabilisation. *Int. Dairy J.* **31**, 55–61 (2013).
- Martin, N. H., Murphy, S. C., Ralyea, R. D., Wiedmann, M. & Boor, K. J. When cheese gets the blues: *Pseudomonas fluorescens* as the causative agent of cheese spoilage. *J. Dairy Sci.* **94**, 3176–3183 (2011).
- del Olmo, A., Calzada, J. & Nuñez, M. The blue discoloration of fresh cheeses: a worldwide defect associated to specific contamination by *Pseudomonas fluorescens*. *Food Control* **86**, 359–366 (2018).
- Dousset, X., Jaffrès, E. & Zagorec, M. *Encyclopedia of Food and Health*. p. 106–112 (Elsevier, 2016).
- D'Incecco, P., Faoro, F., Silvetti, T., Schrader, K. & Pellegrino, L. Mechanisms of *Clostridium tyrobutyricum* removal through natural creaming of milk: a microscopy study. *J. Dairy Sci.* **98**, 5164–5172 (2015).
- Rodríguez, J., Lobato, C., Vázquez, L., Mayo, B. & Flórez, A. B. Prodigiosin-producing *Serratia marcescens* as the causal agent of a red colour defect in a blue cheese. *Foods* **12**, 2388 (2023).
- Garnier, L., Valence, F. & Mounier, J. Diversity and control of spoilage fungi in dairy products: an update. *Microorganisms* **5**, 42, (2017).
- Rukke, E. O., Sørhaug, T. & Stepaniak, L. *Reference Module in Food Science* (Elsevier, 2016).
- Galié, S., García-Gutiérrez, C., Miguélez, E. M., Villar, C. J. & Lombó, F. Biofilms in the food industry: health aspects and control methods. *Front. Microbiol.* **9**, 898 (2018).
- Alvarez-Ordóñez, A., Coughlan, L. M., Briandet, R. & Cotter, P. D. Biofilms in food processing environments: challenges and opportunities. *Annu. Rev. Food Sci. Technol.* **10**, 173–195 (2019).
- Lortal, S. et al. Tina wooden vat biofilm: a safe and highly efficient lactic acid bacteria delivering system in PDO Ragusano cheese making. *Int. J. Food Microbiol.* **132**, 1–8 (2009).
- Cruciata, M. et al. Formation and characterization of early bacterial biofilms on different wood typologies applied in dairy production. *Appl. Environ. Microbiol.* **84**, e02107–17 (2018).
- Sun, L. & D'Amico, D. J. The impact of environmental conditions and milk type on microbial communities of wooden vats and cheeses produced therein. *Food Microbiol.* **115**, 104319 (2023).
- Somers, E. B., Johnson, M. E. & Wong, A. C. L. Biofilm formation and contamination of cheese by nonstarter lactic acid bacteria in the dairy environment. *J. Dairy Sci.* **84**, 1926–1936 (2001).
- De Filippis, F., Valentino, V., Alvarez-Ordóñez, A., Cotter, P. D. & Ercolini, D. Environmental microbiome mapping as a strategy to improve quality and safety in the food industry. *Curr. Opin. Food Sci.* **38**, 168–176 (2021).
- Møretrø, T. & Langsrud, S. Residential bacteria on surfaces in the food industry and their implications for food safety and quality. *Compr. Rev. Food Sci. Food Saf.* **16**, 1022–1041 (2017).
- Bokulich, N. A. & Mills, D. A. Facility-specific “house” microbiome drives microbial landscapes of artisan cheesemaking plants. *Appl. Environ. Microbiol.* **79**, 5214–5223 (2013).
- Stellato, G., De Filippis, F., La Storia, A. & Ercolini, D. Coexistence of lactic acid bacteria and potential spoilage Microbiota in a dairy

- processing environment. *Appl. Environ. Microbiol.* **81**, 7893–7904 (2015).
28. McHugh, A. J. et al. Microbiome-based environmental monitoring of a dairy processing facility highlights the challenges associated with low microbial-load samples. *Npj Sci. Food* **5**, 1–13 (2021).
  29. Barcenilla, C. et al. Improved sampling and DNA extraction procedures for microbiome analysis in food-processing environments. *Nat. Protoc.* **19**, 1291–1310 (2024).
  30. Calasso, M. et al. Relationships among house, rind and core microbiotas during manufacture of traditional Italian cheeses at the same dairy plant. *Food Microbiol.* **54**, 115–126 (2016).
  31. Reuben, R. C., Langer, D., Eisenhauer, N. & Jurburg, S. D. Universal drivers of cheese microbiomes. *iScience* **26**, 105744 (2023).
  32. Mounier, J., Rea, M. C., O'Connor, P. M., Fitzgerald, G. F. & Cogan, T. M. Growth characteristics of *Brevibacterium*, *Corynebacterium*, *Microbacterium*, and *Staphylococcus* spp. isolated from surface-ripened cheese. *Appl. Environ. Microbiol.* **73**, 7732–7739 (2007).
  33. Giuffrida, D. et al. Carotenoids from the ripening bacterium *Brevibacterium linens* impart color to the rind of the French cheese, Fourme de Montbrison (PDO). *Nat. Prod. Res.* **34**, 10–15 (2020).
  34. Settanni, L. et al. In-depth investigation of the safety of wooden shelves used for traditional cheese ripening. *Appl. Environ. Microbiol.* **87**, e0152421 (2021).
  35. Quijada, N. M. et al. Autochthonous facility-specific microbiota dominates washed-rind Austrian hard cheese surfaces and its production environment. *Int. J. Food Microbiol.* **267**, 54–61 (2018).
  36. Yang, C. et al. Strain-level multiomics analysis reveals significant variation in cheeses from different regions. *Lebenson. Wiss. Technol.* **151**, 112043 (2021).
  37. Sun, L. & D'Amico, D. J. Composition, succession, and source tracking of microbial communities throughout the traditional production of a farmstead cheese. *mSystems* **6**, e0083021 (2021).
  38. Youn, H.-Y. & Seo, K.-H. Isolation and characterization of halophilic *Kocuria* salsicia strains from cheese brine. *Food Sci. Anim. Resour.* **42**, 252–265 (2022).
  39. Valdés-Stauber, N. & Scherer, S. Isolation and characterization of Linocin M18, a bacteriocin produced by *Brevibacterium linens*. *Appl. Environ. Microbiol.* **60**, 3809–3814 (1994).
  40. Peterson, S. B., Bertolli, S. K. & Mougous, J. D. The central role of interbacterial antagonism in bacterial life. *Curr. Biol.* **30**, R1203–R1214 (2020).
  41. Silva, C. C. G., Silva, S. P. M. & Ribeiro, S. C. Application of bacteriocins and protective cultures in dairy food preservation. *Front. Microbiol.* **9**, 594 (2018).
  42. Montel, M.-C. et al. Traditional cheeses: rich and diverse microbiota with associated benefits. *Int. J. Food Microbiol.* **177**, 136–154 (2014).
  43. Pasolli, E. et al. Large-scale genome-wide analysis links lactic acid bacteria from food with the gut microbiome. *Nat. Commun.* **11**, 2610 (2020).
  44. De Filippis, F., Pasolli, E. & Ercolini, D. The food-gut axis: lactic acid bacteria and their link to food, the gut microbiome and human health. *FEMS Microbiol. Rev.* **44**, 454–489 (2020).
  45. Hassani, S. et al. High prevalence of antibiotic resistance in pathogenic foodborne bacteria isolated from bovine milk. *Sci. Rep.* **12**, 1–10 (2022).
  46. Liu, J., Zhu, Y., Jay-Russell, M., Lemay, D. G. & Mills, D. A. Reservoirs of antimicrobial resistance genes in retail raw milk. *Microbiome* **8**, 99 (2020).
  47. Tóth, A. G. et al. Antimicrobial resistance genes in raw milk for human consumption. *Sci. Rep.* **10**, 1–7 (2020).
  48. Sachi, S., Ferdous, J., Sikder, M. & Hussani, S. Antibiotic residues in milk: past, present, and future. *J. Adv. Vet. Anim. Res.* **6**, 315 (2019).
  49. Wolfe, B. E. Are fermented foods an overlooked reservoir of antimicrobial resistance? *Curr. Opin. Food Sci.* **51**, 101018 (2023).
  50. Stefanovic, E., Kilcawley, K. N., Rea, M. C., Fitzgerald, G. F. & McAuliffe, O. Genetic, enzymatic and metabolite profiling of the *Lactobacillus casei* group reveals strain biodiversity and potential applications for flavour diversification. *J. Appl. Microbiol.* **122**, 1245–1261 (2017).
  51. Niccum, B. A., et al. Strain-level diversity impacts cheese rind microbiome assembly and function. *mSystems* **5**, e00149–20 (2020).
  52. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
  53. Blanco-Míguez, A. et al. Extending and improving metagenomic taxonomic profiling with uncharacterized species using MetaPhlan 4. *Nat. Biotechnol.* **41**, 1633–1644 (2023).
  54. Beghini, F. et al. Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with bioBakery 3. *Elife* **10**, e65088 (2021).
  55. 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).
  56. Zhu, W., Lomsadze, A. & Borodovsky, M. Ab initio gene identification in metagenomic sequences. *Nucleic Acids Res.* **38**, e132–e132 (2010).
  57. Buchfink, B., Reuter, K. & Drost, H.-G. Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nat. Methods* **18**, 366–368 (2021).
  58. Lebeer, S., Vanderleyden, J. & De Keersmaecker, S. C. J. Genes and molecules of lactobacilli supporting probiotic action. *Microbiol. Mol. Biol. Rev.* **72**, 728–764 (2008).
  59. Quijada, N. M., Rodríguez-Lázaro, D., Eiros, J. M. & Hernández, M. TORMES: an automated pipeline for whole bacterial genome analysis. *Bioinformatics* **35**, 4207–4212 (2019).
  60. Florensa, A. F., Kaas, R. S., Clausen, P. T. L. C., Aytan-Aktug, D. & Aarestrup, F. M. ResFinder – an open online resource for identification of antimicrobial resistance genes in next-generation sequencing data and prediction of phenotypes from genotypes. *Microb. Genom.* **8**, 000748 (2022).
  61. Alcock, B. P. et al. CARD 2023: expanded curation, support for machine learning, and resistome prediction at the Comprehensive Antibiotic Resistance Database. *Nucleic Acids Res.* **51**, D690–D699 (2023).
  62. Gupta, S. K. et al. ARG-ANNOT, a new bioinformatic tool to discover antibiotic resistance genes in bacterial genomes. *Antimicrob. Agents Chemother.* **58**, 212–220 (2014).
  63. Liu, B., Zheng, D., Zhou, S., Chen, L. & Yang, J. VFDB 2022: a general classification scheme for bacterial virulence factors. *Nucleic Acids Res.* **50**, D912–D917 (2022).
  64. Zolfo, M. et al. Detecting contamination in viromes using ViromeQC. *Nat. Biotechnol.* **37**, 1408–1412 (2019).
  65. van Heel, A. J. et al. BAGEL4: a user-friendly web server to thoroughly mine RiPPs and bacteriocins. *Nucleic Acids Res.* **46**, W278–W281 (2018).
  66. 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).
  67. Kang, D. D. et al. MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ* **7**, e7359 (2019).
  68. 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).
  69. Ondov, B. D. et al. Mash: fast genome and metagenome distance estimation using MinHash. *Genome Biol.* **17**, 132 (2016).
  70. 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).
  71. Jain, C., Rodriguez-R, L. M., Phillippy, A. M., Konstantinidis, K. T. & Aluru, S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat. Commun.* **9**, 5114 (2018).

72. Asnicar, F. et al. Precise phylogenetic analysis of microbial isolates and genomes from metagenomes using PhyloPhlAn 3.0. *Nat. Commun.* **11**, 2500 (2020).
73. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* **49**, W293–W296 (2021).

### Acknowledgements

This work was supported by the project MASTER-Microbiome Applications for Sustainable food systems through Technologies and Enterprise, receiving funding from the European Union's Horizon 2020 research and innovation programme (GA 818368). This manuscript reflects only the authors' views and the European Commission is not responsible for any use that may be made of the information it contains. N.M.Q is currently funded by the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie (GA 10103437).

### Author contributions

F.D.F. wrote the first manuscript draft; F.D.F., V.V., R.C.R., C.B., J.F.C.D., N.M.Q., I.C.T., P.R.M., C.S., and G.S. carried out sampling, DNA extraction, and metagenome sequencing; N.C. and E.P. carried out metagenomics reads pre-processing; F.D.F., V.V., and M.Y. analyzed data and prepared figures and tables; F.D.F., A.A.O., P.D.C., and D.E. designed and supervised the study; M.W., A.M., N.S., A.A.O., P.D.C., and D.E. 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/s41522-024-00541-5>.

**Correspondence** and requests for materials should be addressed to Danilo Ercolini.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024