



Tracking the Dairy Microbiota from Farm Bulk Tank to Skimmed Milk Powder

Aoife J. McHugh,^{a,b}  Conor Feehily,^{a,c} Mark A. Fenelon,^d David Gleeson,^e Colin Hill,^{b,c} Paul D. Cotter^{a,c}

^aFood Bioscience Department, Teagasc Food Research Centre, Cork, Ireland

^bSchool of Microbiology, University College Cork, Cork, Ireland

^cAPC Microbiome Ireland, Cork, Ireland

^dFood Chemistry and Technology Department, Teagasc Food Research Centre, Cork, Ireland

^eTeagasc Animal and Grassland Research and Innovation Centre, Cork, Ireland

ABSTRACT Microorganisms from the environment can enter the dairy supply chain at multiple stages, including production, milk collection, and processing, with potential implications for quality and safety. The ability to track these microorganisms can be greatly enhanced by the use of high-throughput DNA sequencing (HTS). Here HTS, both 16S rRNA gene amplicon and shotgun metagenomic sequencing were applied to investigate the microbiomes of fresh mid- and late-lactation milk collected from farm bulk tanks, collection tankers, milk silos, skimmed milk silos, a cream silo, and powder samples to investigate the microbial changes throughout a skim milk powder manufacturing process. 16S rRNA gene analysis established that the microbiota of raw milks from farm bulk tanks and in collection tankers were very diverse but that psychrotrophic genera associated with spoilage, *Pseudomonas* and *Acinetobacter*, were present in all samples. Upon storage within the whole-milk silo at the processing facility, the species *Pseudomonas fluorescens* and *Acinetobacter baumannii* became dominant. The skimmed milk powder generated during the mid-lactation period had a microbial composition that was very different from that of raw milk; specifically, two thermophilic genera, *Thermus* and *Geobacillus*, were enriched. In contrast, the microbiota of skimmed milk powder generated from late-lactation milk more closely resembled that of the raw milk and was dominated by spoilage-associated psychrotrophic bacteria. This study demonstrates that the dairy microbiota can differ significantly across different sampling days. More specifically, HTS can be used to trace microbial species from raw milks through processing to final powdered products.

IMPORTANCE Microorganisms can enter and persist in dairy at several stages of the processing chain. Detection of microorganisms within dairy food processing is currently a time-consuming and often inaccurate process. This study provides evidence that high-throughput sequencing can be used as an effective tool to accurately identify microorganisms along the processing chain. In addition, it demonstrates that the populations of microbes change from raw milk to the end product. Routine implementation of high-throughput sequencing would elucidate the factors that influence population dynamics. This will enable a manufacturer to adopt control measures specific to each stage of processing and respond in an effective manner, which would ultimately lead to increased food safety and quality.

KEYWORDS dairy, processing, metagenomics, skimmed milk powder, 16S rRNA gene amplicon sequencing, bulk tank milk, whole-milk silo, skimmed milk silo, collection tanker, microbiota

Citation McHugh AJ, Feehily C, Fenelon MA, Gleeson D, Hill C, Cotter PD. 2020. Tracking the dairy microbiota from farm bulk tank to skimmed milk powder. mSystems 5:e00226-20. <https://doi.org/10.1128/mSystems.00226-20>.

Editor Jack A. Gilbert, University of California, San Diego

Copyright © 2020 McHugh et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Paul D. Cotter, paul.cotter@teagasc.ie.

Received 16 March 2020

Accepted 17 March 2020

Published 7 April 2020

Bovine milk is a nutritious natural food that can be processed into many different products, including dairy powders that can be used as a base for therapeutic, nutritional, and/or infant formulas. Processing is required to provide a safe and stable shelf life and has a considerable impact on the microbial communities of the resultant products (1). With increased global demand for dairy products, including milk powders that are incorporated into infant milk formula, there is an even greater need to understand the associated microbiota in order to optimize food safety and quality. Such an understanding should incorporate an appreciation of the impact of both raw ingredients and processing environments on the final product (2–7). However, more information is needed to build a comprehensive view of the dairy microbiota and the factors that contribute to its composition. Traditional microbiological detection techniques focus on culturable bacteria. However, these approaches will not capture viable but nonculturable bacteria (1, 8) or non-readily-culturable bacteria (9). They can also be susceptible to false positives/negatives (10, 11), may not differentiate between closely related species, rely on a specific test for each target microbe, and are often time-consuming. More recently, high-throughput DNA sequencing (HTS) has been used to study the influence of environmental factors on the dairy microbiota. Animal housing, cleaning, and milking practices (7); weather conditions (12); seasonal influences and on-farm storage conditions (13); as well as large-scale storage and seasonal transportation (14) have been shown to influence the raw-milk microbiota. Seasonality can be particularly important in pasture-based systems when milk quality is impacted by where the herd is located (indoor versus outdoor), among other factors. In addition, associations have been made between the raw-milk microbiota and somatic cell counts (a hygiene and herd health indicator) in bulk tank (BT) samples (12, 15). The impacts of processing, such as pasteurization, on the milk microbiota (1) have been studied. However, more studies are needed to provide a complete view of different types of processing as well as the impact of different processing runs and different processing days on the dairy microbiota. Previous studies are limited by the use of 16S rRNA gene amplicon sequencing, which can provide taxonomic resolution to the genus level only. Therefore, there is an insufficient understanding of the functional potential of the microbial communities and, indeed, characterization of nonbacterial microbial contamination along the dairy chain. Recently, shotgun metagenomic sequencing, which overcomes these issues, has been used to study dairy products (16–18). Here, 16S rRNA gene amplicon and shotgun metagenomic analyses are used together to facilitate an in-depth study of the dairy microbiome from the farm through transportation and processing to a skimmed milk powder (SMP) and, in the process, provide valuable information regarding the impacts of collection, storage, and processing on this.

RESULTS

Bulk tank milks contain a diverse microbiota that differs in samples collected from the mid- and late-lactation periods. DNA was extracted and 16S rRNA gene amplicons were sequenced for a total of 67 raw bulk tank (BT) milk samples collected on 1 day in the early- to mid-lactation period (May 2016), here referred to as mid-lactation (ML). This process was repeated in the late-lactation period (October 2016) (Fig. 1). The alpha diversity of the raw-milk microbiota in bulk tanks on farms was relatively high (Fig. 2A) compared to that in subsequent processing stages. However, the microbial alpha diversity in mid-lactation bulk tank samples was significantly lower ($P < 0.001$) than that in the corresponding late-lactation samples (Fig. 2A). Beta diversity showed that samples from bulk tank milks were dissimilar but broadly clustered together (Fig. 2B). Farm bulk tanks were composed of a high number of genera that were present at low relative abundances of less than 5%. In one sample, these low-abundance genera accounted for 74.9% of the total population, and an average of 46.4% was seen across all bulk tank samples (Fig. 3; see also Fig. S1 and Table S1 in the supplemental material). There were 42 genera present at high relative abundances ($>5\%$ relative abundance in at least one sample) (Table S1), including traditionally milk-associated taxa such as *Pseudomonas* (mean, 6.6%), *Acinetobacter* (mean, 5.2%),

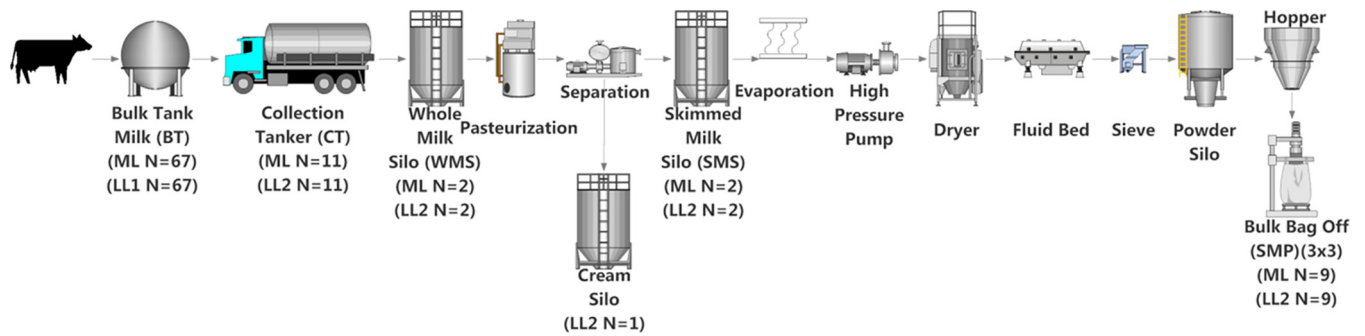


FIG 1 Sampling schematic. Sampling included mid-lactation (ML) samples (May 2016), where a whole processing run was sampled from farm bulk tank milk to skimmed milk powder. This included 67 farm bulk tanks, the 11 collection tankers used to collect this milk, and the whole-milk silo into which this milk was pooled (2 samples were obtained). Milk in the WMS was subjected to pasteurization and separation and stored in a skimmed milk silo from which 2 samples were obtained. Milk in the SMS was subjected to heating and drying to make a skimmed milk powder (SMP), from which 9 samples were obtained, 3 samples from each of 3 bags. The same 67 farms were resampled during late lactation (October 2016) (LL1, late lactation 1). On a separate day during this late-lactation period (December 2016) (LL2, late lactation 2), samples from tankers and the processing run, including an additional cream sample, were collected for analysis.

Lactococcus (mean, 4.7%), *Corynebacterium* (mean, 4.2%), and *Streptococcus* (mean, 2.5%) (Fig. 3 and Table S1). In general, the microbial diversity of the bulk tank milks was such that the taxonomic compositions differed across farms and the two sampling days. However, it was apparent that the bulk tank milk profiles from some farms remained relatively more stable across the two sampling points; e.g., farm 23 had high relative abundances of *Leuconostoc* and *Acinetobacter* in both the mid- and late-lactation samples (Fig. 3). Overall, 17 high-relative-abundance genera, which were genera present at a >5% relative abundance in at least one sample, differed significantly in their relative abundances between paired mid- and late-lactation bulk tank samples ($P < 0.05$). Nine genera showed higher relative abundances in mid-lactation bulk tanks, while eight were more abundant in late-lactation bulk tanks (Fig. 4). Among these, genera grouped as *Clostridium sensu stricto* subgroups 1 and 5 were both present at significantly higher relative abundances in the late-lactation bulk tank milks. Low-abundance genera, grouped as “others,” were also detected at significantly higher proportions in late-lactation bulk tanks (Fig. 4). One mid-lactation sample had a particularly high proportion of *Yersinia* (5.6%), leading to mid-lactation bulk tanks having, on average, a significantly higher relative abundance of this genus. There was also a larger number of genera (77) that were present at a lower relative abundance but at a >1% abundance in at least one sample, which significantly differed in abundance between mid- and late-lactation bulk tank samples (Fig. S2). It was again notable that taxa corresponding to the genus *Clostridium*, in this instance, *Clostridium sensu stricto* subgroups 15 and 18, were detected at significantly higher relative abundances in late-lactation samples.

Collection tanker milks retain relatively high microbial diversity but with some taxonomic convergence. Milk samples obtained from collection tankers (CTs) were, with the exception of bulk tank milks, among those containing the highest microbial alpha diversity (Fig. 2A). Milk from mid-lactation collection tankers had a significantly higher microbial alpha diversity ($P < 0.05$) than the corresponding late-lactation samples (Fig. 2A). From a beta diversity perspective, mid-lactation CT samples cluster closely with the bulk tanks from which they were filled (Fig. 2B). Late-lactation CT samples cluster further away from bulk tank samples, reflecting their collection on different days (as in the late-lactation period, the 67 farms from which bulk tank milk was collected did not yield a sufficiently large pool of milk to proceed with the powder manufacturing process) (Fig. 2B). Although the microbial composition of the tanker samples was diverse, and no two tankers had the same composition, a pattern of enrichment was apparent with respect to the taxa present at the highest relative abundances (Fig. 3 and Table S1), with *Acinetobacter*, *Pseudomonas*, *Lactococcus*, and *Corynebacterium* on average accounting for between 9% and 3% of the mid-lactation

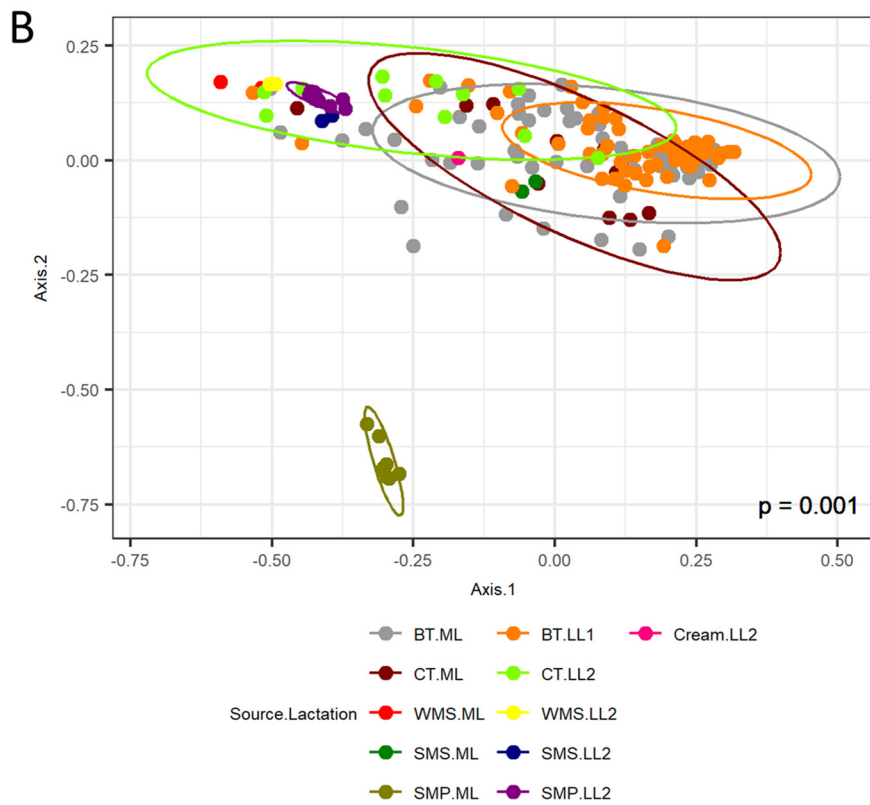
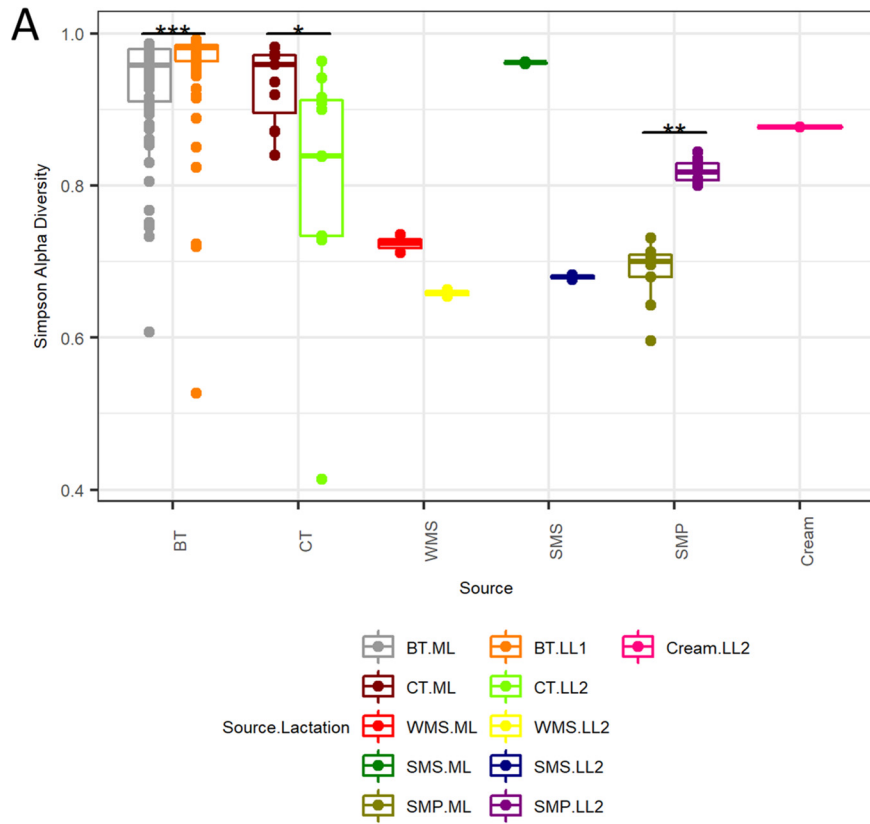


FIG 2 Microbial diversity indexes of sampling sites. (A) Simpson alpha diversity analysis of 16S rRNA gene amplicon sequence data for mid-lactation (ML) and late-lactation (LL1/LL2) samples from bulk tanks (BT), collection tankers (CT), whole-milk silos (WMS), skimmed milk silos (SMS), skimmed milk powder (SMP), and cream (LL2 only). Significant differences are highlighted (***, $P < 0.001$; **, $P < 0.01$;

(Continued on next page)

tanker milk microbiota composition (Fig. 3). These taxa were also present at high relative abundances in the previous bulk tank samples. Analysis of the corresponding samples collected during late lactation showed that although the tanker's milk microbiota composition was diverse, *Pseudomonas* (mean, 20.3%) and *Acinetobacter* (mean, 25.4%) became the most dominant genera in each sample (Fig. 3 and Table S1). These were the genera with the highest relative abundances in bulk tank samples as well. Mid-lactation tanker milk samples had a higher proportion of low-abundance (<5%) and very-low-abundance (<1%) genera than the late-lactation tanker samples (Fig. 3 and Fig. S1). *Pseudomonas*, *Bacillus*, and *Ruminococcaceae* UCG.005 differed significantly between the mid- and late-lactation tanker samples, with *Pseudomonas*, or, more specifically, *Pseudomonas fluorescens* as determined by shotgun metagenomic sequencing, being present at a higher relative abundance in late-lactation samples. *Bacillus* and *Ruminococcaceae* UCG.005 were noted to be present at higher relative abundances in mid-lactation samples (Fig. 5). With respect to *Bacillus*, the pattern was driven by a higher relative abundance in mid-lactation tankers 4, 8, and 11 (Fig. 3 and Fig. S1). Shotgun metagenomic data established that at the species level, this taxon corresponded to *Bacillus coagulans* (Fig. 6). When SUPER-FOCUS was used to assign a functional classification to shotgun metagenomic reads (Fig. S3), higher relative abundances of virulence functions were noted in raw milks in pooled tanker samples than in heat-processed samples of skimmed milk silos (SMSs) or skimmed milk powders (Fig. 7). A high proportion of the shotgun metagenomic reads sequenced were not of microbial origin and were assigned to *Bos taurus* (Fig. S4), and so the resulting low number of microbe-associated reads did not allow strain-level classification or in-depth functional classification (Fig. S3). For this reason, the functional classifications with the most notable differences between groups are the only ones discussed (Fig. 7).

Whole-milk silo storage results in an increased dominance of psychrotrophic spoilage-associated bacteria. A decrease in microbial alpha diversity was observed in whole-milk silo (WMS) samples, regardless of the period of collection, compared to the corresponding samples from collection tankers (Fig. 2A). Beta diversity analysis displayed tight clustering of whole-milk silo samples, regardless of the lactation stage. Whole-milk silo samples cluster together, away from mid-lactation collection tankers and at the edge of the ellipse for late-lactation collection tankers (Fig. 2B). A high relative abundance of spoilage-associated psychrotrophic bacteria was observed in both mid- and late-lactation samples (Fig. 3), with *Pseudomonas*, *Acinetobacter*, as well as the nonpsychrotrophic *Lactococcus* genus dominating. These genera had begun to dominate the previous CT samples, particularly the late-lactation CT samples. Taxonomic compositions from mid- and late-lactation samples were similar (Fig. 3), *Leuconostoc* was the only genus present at a higher abundance in mid-lactation (mean, 6.3%) than in late-lactation (mean, 0.1%) samples (Table S1). In contrast, low-abundance genera, denoted "others," were present at higher relative abundances in late-lactation (mean, 5.4%) than in mid-lactation (mean, 3.1%) samples (Table S1). However, all WMS samples had a smaller proportion of low-abundance genera than the collection tankers and bulk tank milks (Fig. 3). This reflected the lower alpha diversity of these samples than those collected earlier in the milk processing chain (Fig. 2A). WMS samples contained no unique genera at a >5% relative abundance that had not been noted in previous BT and CT samples (Table S1 and Fig. S5). Shotgun metagenomic analysis revealed species-level classifications of the most abundant genera across all WMS samples as *Pseudomonas fluorescens*, *Acinetobacter baumannii*, and *Lactococcus lactis* (Fig. 6); this was in agreement with 16S genus-level classifications. Due to the high proportion of shotgun metagenomic reads assigned to *Bos taurus* (Fig. S4), the resulting low number of bacterium-associated reads did not allow strain-level classification.

FIG 2 Legend (Continued)

*, $P < 0.05$). (B) Bray-Curtis multidimensional scaling analysis of 16S rRNA gene amplicon sequence data for mid- and late-lactation samples from bulk tanks, collection tankers, whole-milk silos, skimmed milk silos, skimmed milk powder, and cream (late lactation only).

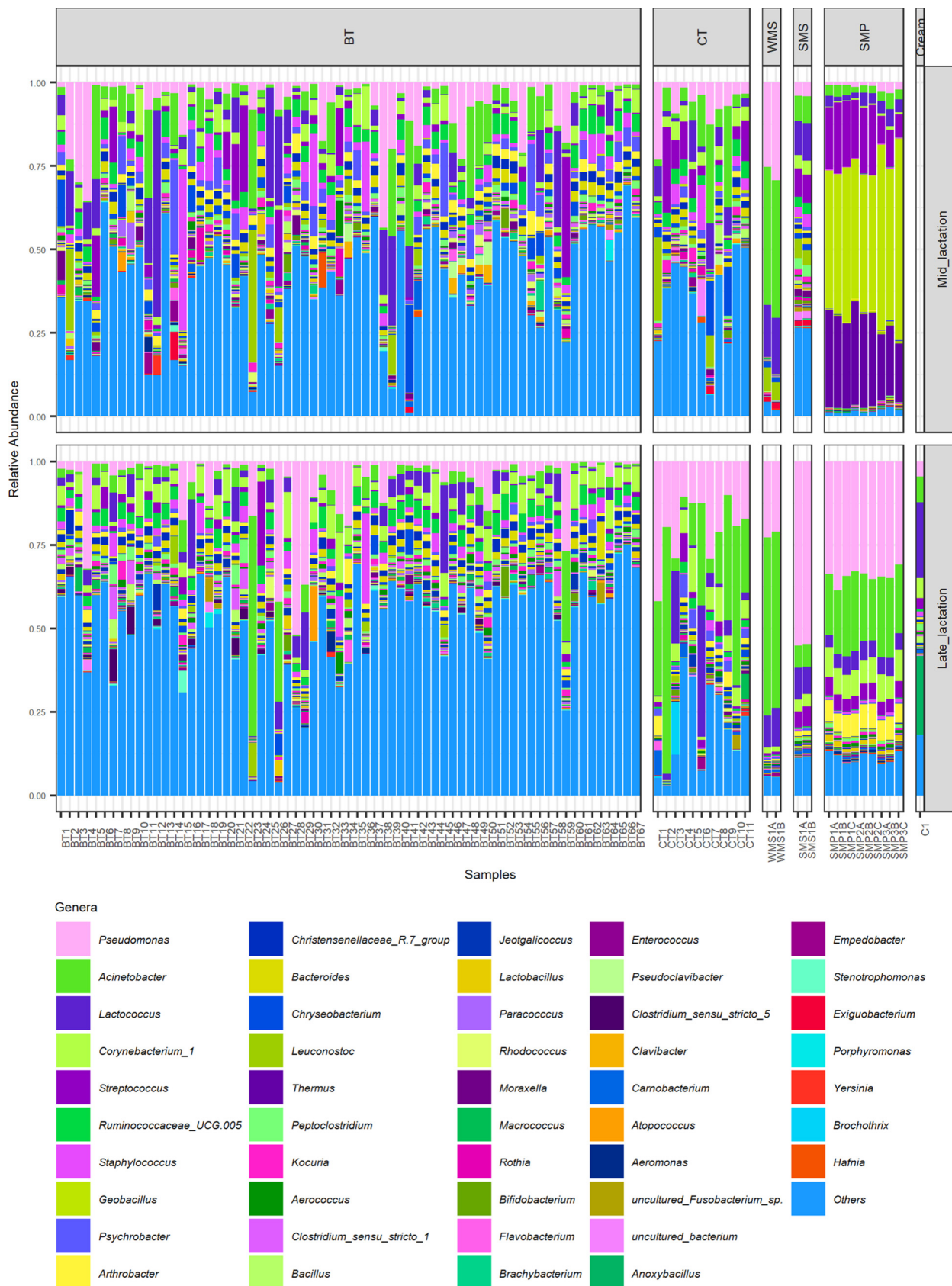


FIG 3 Genera present at a >5% relative abundance in at least one sample. Shown are 16S rRNA gene amplicon sequence data for mid- and late-lactation samples from bulk tanks (BT), collection tankers (CT), whole-milk silos (WMS), skimmed milk silos (SMS), skimmed milk powders (SMP), and cream (late lactation only). Genera shown are present at a >5% relative abundance in at least one sample.

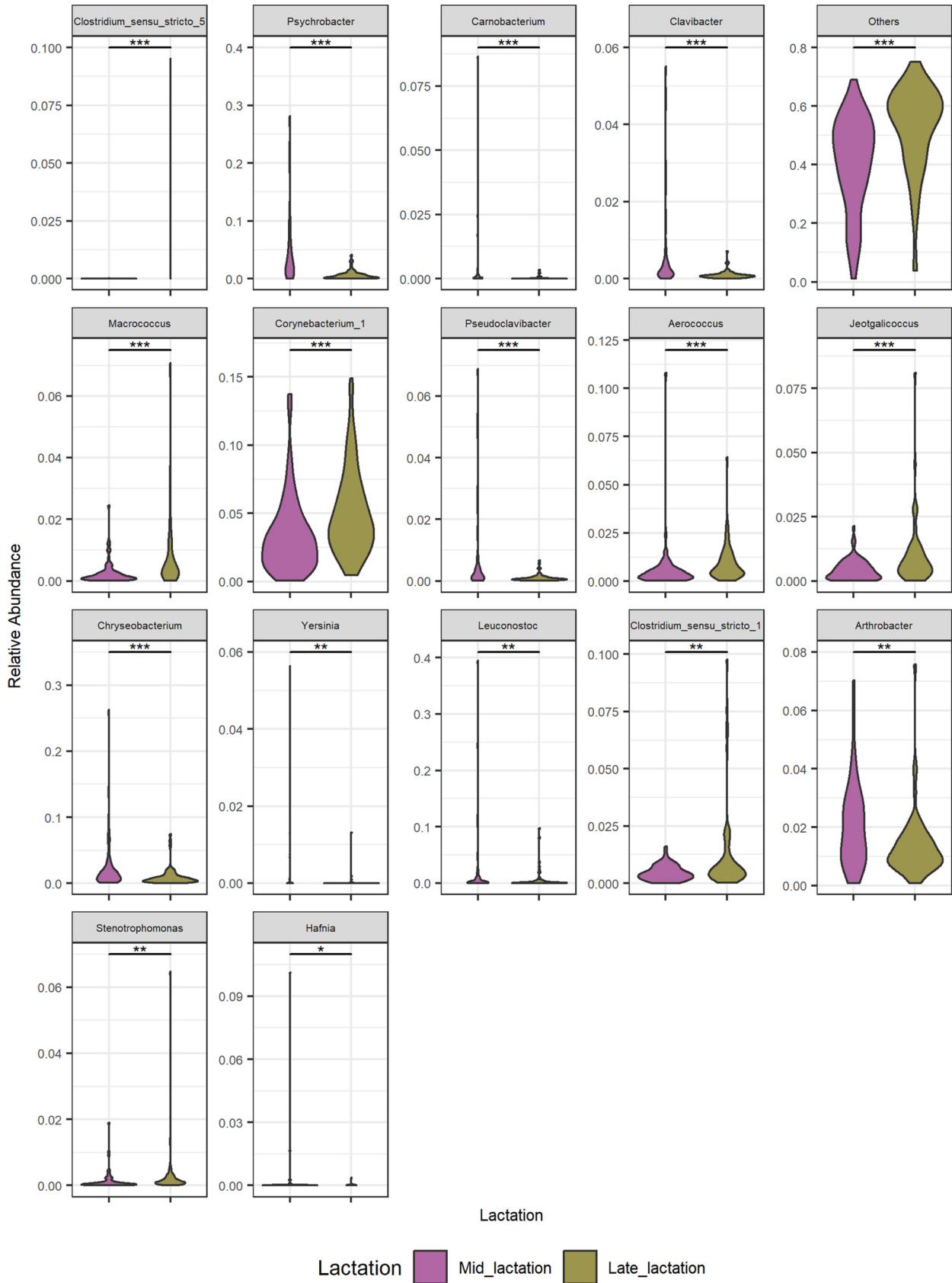


FIG 4 Significantly differential relative abundances of taxa constituting at least a 5% relative abundance in at least one sample between mid- and late-lactation bulk tanks. Shown are violin plots of genera that differ significantly in relative abundances between mid- and late-lactation bulk tanks.

(Continued on next page)

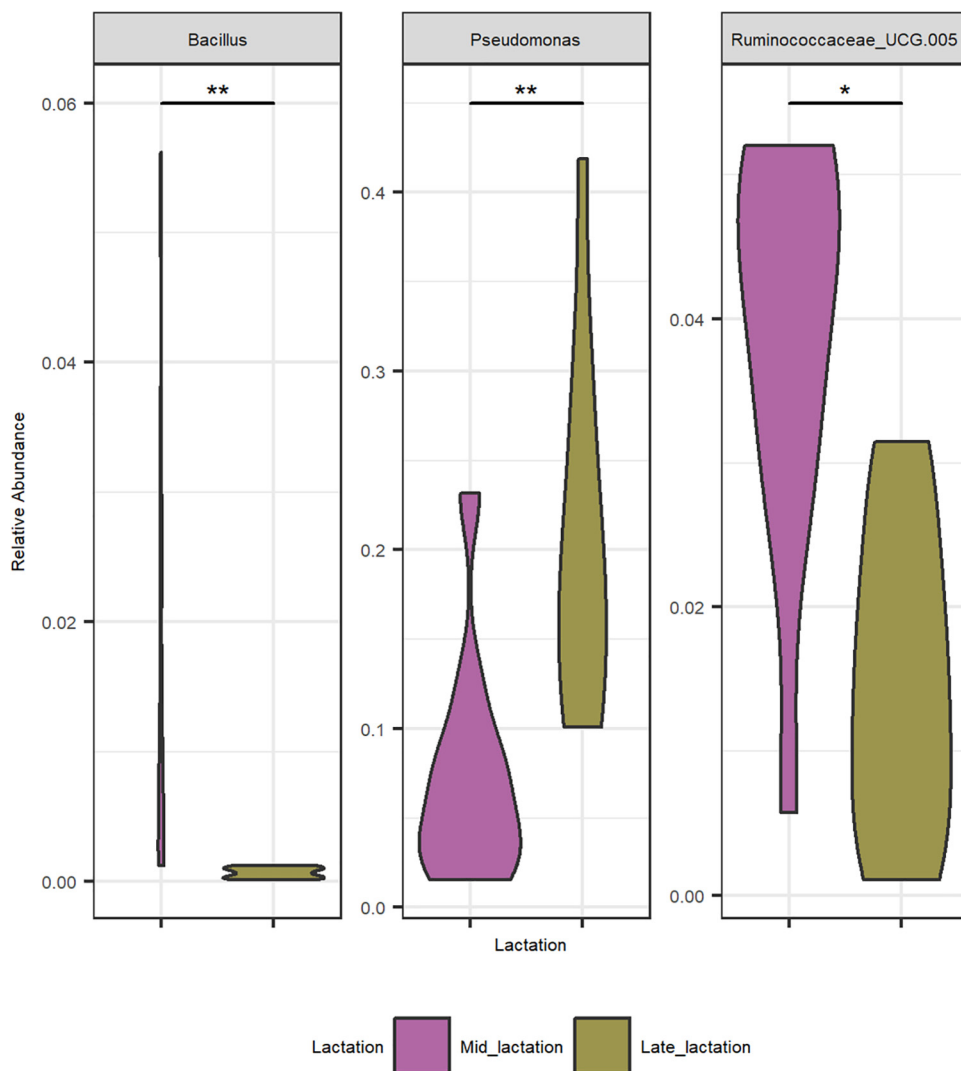


FIG 5 Significantly differential relative abundances of taxa constituting at least a 5% relative abundance in at least one sample between mid- and late-lactation collection tankers. Shown are violin plots of genera that differ significantly in relative abundances between mid- and late-lactation collection tankers. This highlights genera that are present at a >5% relative abundance in at least one sample. A pairwise Wilcoxon rank sum test using Benjamini-Hochberg *P* value correction/FDR correction analysis was performed, and samples are ordered by the lowest *P* value to the highest (**, *P* < 0.01; *, *P* < 0.05).

When SUPER-FOCUS was used to assign a functional classification to the shotgun metagenomic reads (Fig. S3), a higher relative abundance of virulence functions was noted for raw milks in the WMS and pooled tankers than for all other samples (Fig. 7).

The skimmed milk silo microbial composition differs across samples collected in different seasons. The alpha diversity of the skimmed milk silo (SMS) microbiota was higher than that of the preceding whole-milk silo samples (Fig. 2A). Beta diversity analysis showed that the microbiota of mid-lactation SMS samples clustered separately from WMS samples. In contrast, late-lactation samples clustered closely with their corresponding WMS samples (Fig. 2B). From a taxonomic perspective, the relative abundance of the dominant psychrotrophic bacteria detected at the preceding WMS

FIG 4 Legend (Continued)

This highlighted genera that are present at a >5% relative abundance in at least one sample. A pairwise Wilcoxon rank sum test using Benjamini-Hochberg *P* value correction/FDR correction analysis was performed, and samples are ordered by the lowest *P* value to the highest (***, *P* < 0.001; **, *P* < 0.01; *, *P* < 0.05).

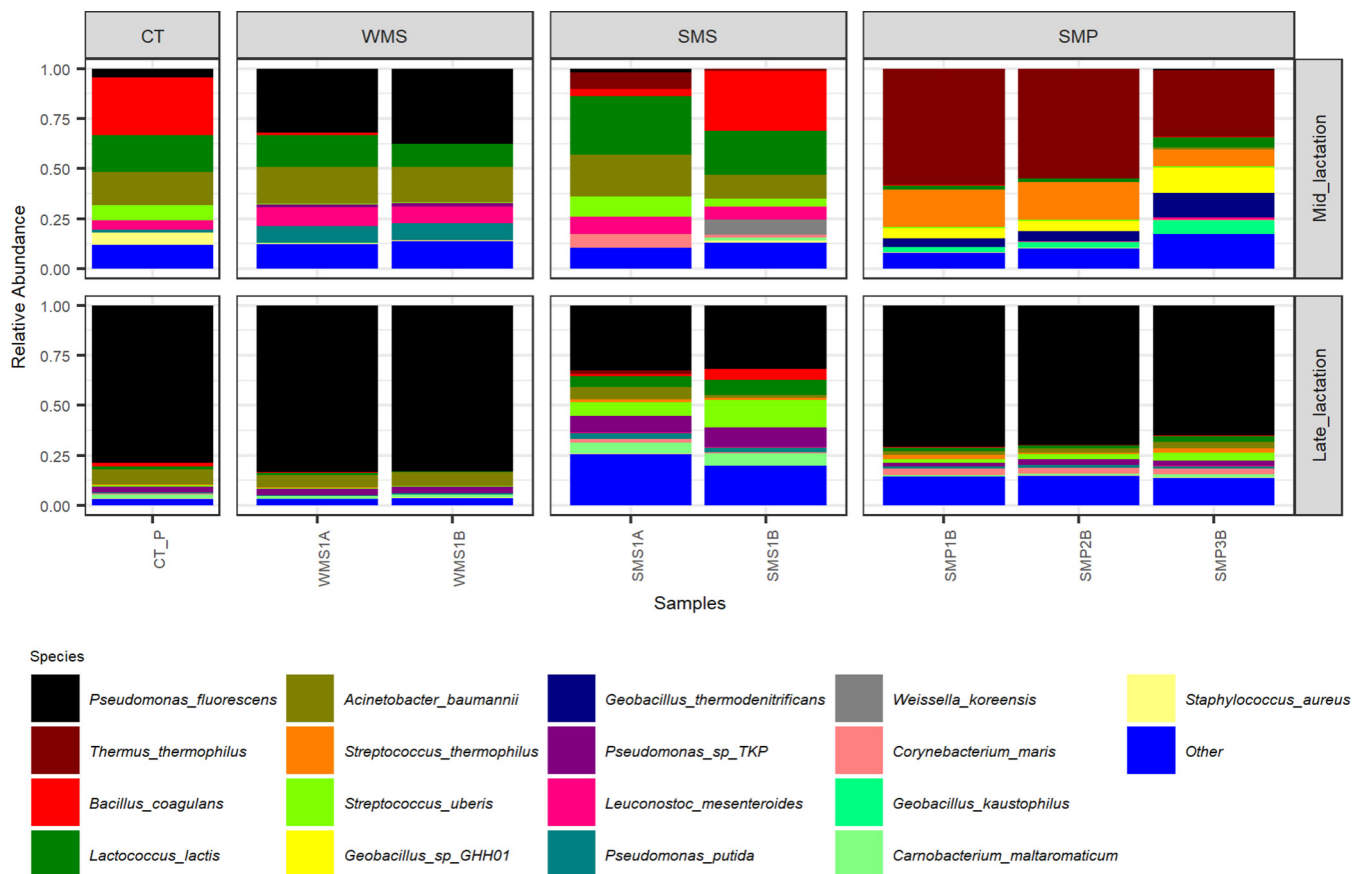


FIG 6 Relative abundances of species present at a >5% relative abundance in at least one sample. Shown are shotgun species-level taxonomic classifications from Kraken analysis using a filter threshold of 0.2 on a subset of samples from the mid- and late-lactation processing pipelines, including a pooled representative collection tanker (CT_P) sample, whole-milk silo (WMS) samples, skimmed milk silo (SMS) samples, and a subset of skimmed milk powder (SMP) samples from each lactation stage.

stage was reduced in the mid-lactation SMS samples following pasteurization and separation of that milk (Fig. 3). This decrease was not as evident in late-lactation samples, with *Pseudomonas*, *Lactococcus*, and *Acinetobacter* remaining dominant (Fig. 3). SMS samples from both time points contained *Lactococcus*, *Acinetobacter*, and *Streptococcus* (Fig. 3); the former two genera were present at high relative abundances at all previous processing stages, whereas the latter, *Streptococcus*, was present at a high relative abundance in the preceding BT and CT samples but not in the WMS samples. Species-level taxonomic analysis revealed that these primarily corresponded to *Lactococcus lactis*, *Acinetobacter baumannii*, *Streptococcus thermophilus*, and *Streptococcus uberis* at the species level (Fig. 6). Mid-lactation SMS samples had a higher proportion of low- and very-low-abundance bacteria than the late-lactation SMS samples (Fig. 3, Table S1, and Fig. S1). When SUPER-FOCUS was used for the functional classification of shotgun metagenomic reads (Fig. S3), higher relative abundances of genes associated with phage, prophage, and transposable elements were noted in mid-lactation SMS samples than in all other samples (Fig. 7).

The cream silo microbiota differs from those of other dairy processing samples.

The late-lactation processing pipeline provided the only cream sample available for analysis within this study. This sample had a higher alpha diversity than the whole-milk silo sample from which it was produced and the skimmed milk silo contents from which it was separated (Fig. 2A). The microbial communities in the sample did not cluster with either the late-lactation skimmed milk silo or the whole-milk silo samples (Fig. 2B). This sample was dominated by *Anoxybacillus* (Fig. 3), a genus present at a <1% relative abundance at previous stages of the processing pipeline, as well as *Lactococcus*,

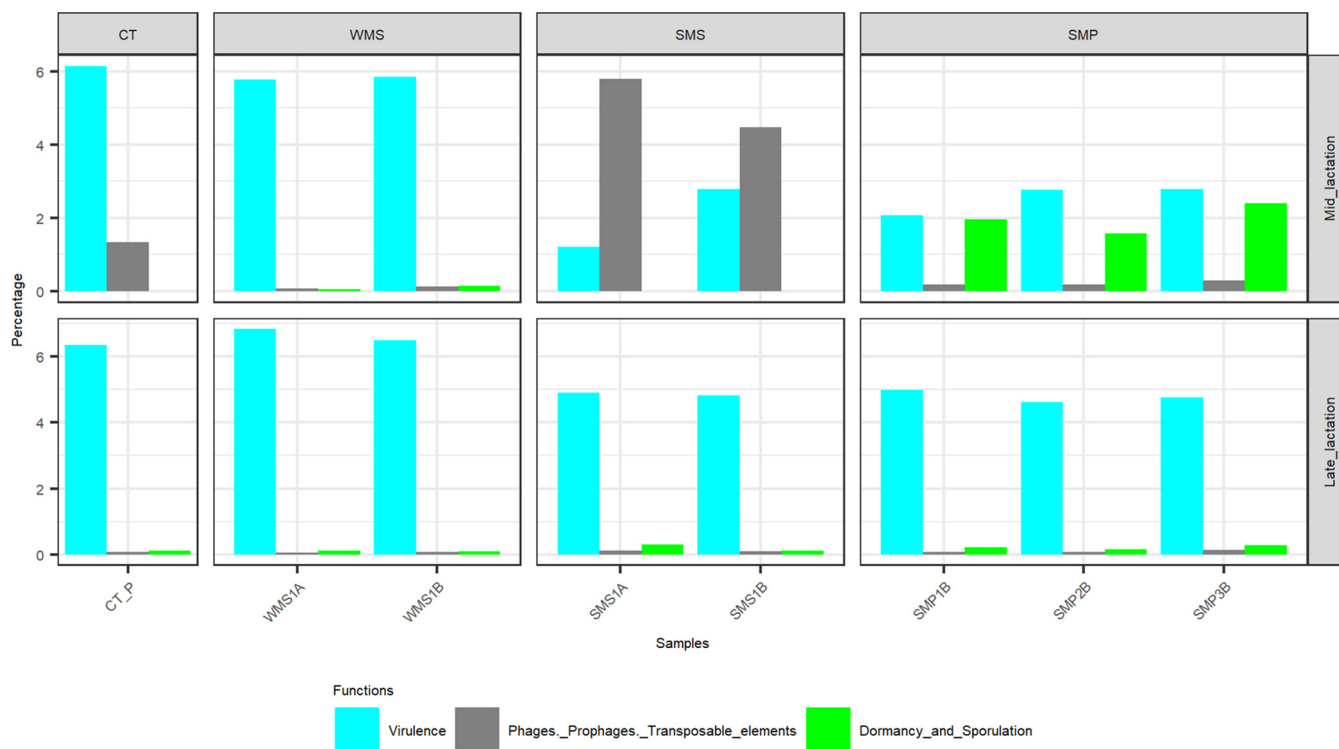


FIG 7 Percentages of a subset of functional classifications from SUPER-FOCUS L1 analysis attributed to prophage, phage, and transposable elements; dormancy and sporulation; as well as virulence in mid- and late-lactation pooled collection tanker (CT_P), whole-milk silo (WMS), skimmed milk silo (SMS), and skimmed milk powder (SMP) samples.

Corynebacterium, and *Acinetobacter*, among other genera (Fig. 3 and Table S1). Low-abundance genera accounted for 18.0% of the cream silo microbiota (Table S1), and 6.4% of the microbiota corresponded to very-low-abundance genera.

The dairy powder microbiota can vary, reflecting either the original raw-milk microbiota or microbes selected for during processing. The microbiotas of skimmed milk powder (SMP) samples differed in a manner that reflected the impact of processing on particular days. The microbiota of mid-lactation skimmed milk powder samples had a lower alpha diversity than that of the SMS milk from which it was produced (Fig. 2A). Furthermore, the microbial communities in the mid-lactation skimmed milk powders clustered separately from both the late-lactation skimmed milk powder samples and the milks from which they were derived (Fig. 2B). More specifically, the mid-lactation powders showed a shift in microbial taxonomic dominance, with thermophilic bacteria such as *Thermus*, *Geobacillus*, and *Streptococcus* being more dominant in these samples (Fig. 3). Notably, *Thermus* and *Geobacillus* had not been seen in any previous stages of the processing pipeline. Shotgun metagenomic sequencing assigned these as *Thermus thermophilus*, *Geobacillus* sp. strain GHH01, *Geobacillus thermodenitrificans*, and *Streptococcus thermophilus*. SUPER-FOCUS functional classification highlighted corresponding increases in the relative abundances of sporulation- and dormancy-associated genes in mid-lactation skimmed milk powder samples (Fig. 7), reflecting the proportions of spore-forming bacteria present. In these samples, low-relative-abundance genera (<5%) accounted for 0.9% to 2.9% of reads, and very-low-abundance genera (<1%) accounted for 0.45% to 1.3% of reads. The microbiotas of late-lactation skimmed milk powders differed considerably. An increase in diversity was observed in late-lactation skimmed milk powders compared to the SMS milk from which they were produced (Fig. 2A), and these samples clustered closely to the skimmed milk samples from which they were derived (Fig. 2B). Taxonomic analysis revealed the dominance of psychrotrophic genera and, more specifically, of the same genera and species (*P. fluorescens* and *A. baumannii*) that had dominated previous WMS and SMS samples

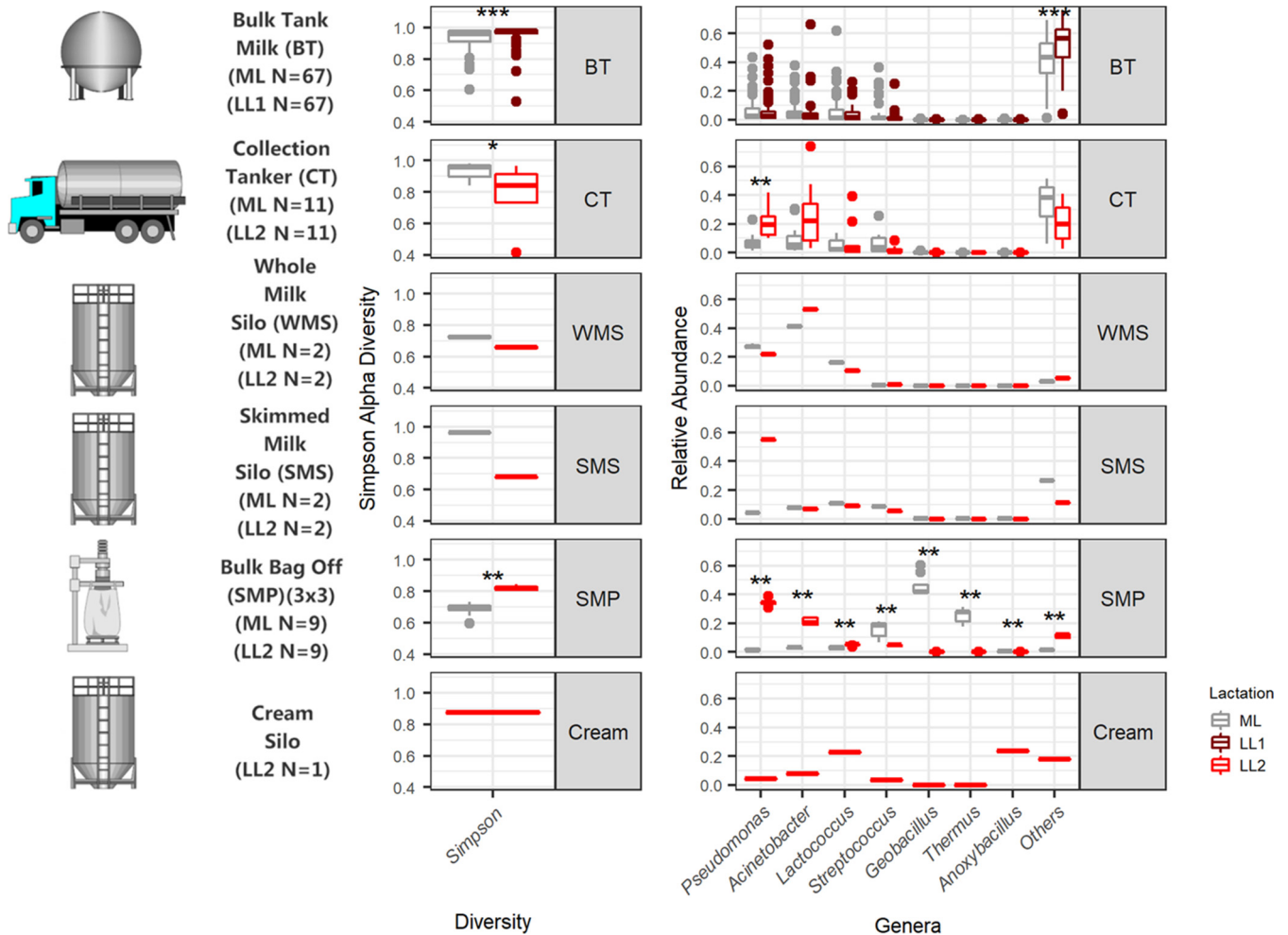


FIG 8 Summary of differences in results due to sampling location and lactation stage. A brief sampling schematic with numbers, Simpson alpha diversity values, as well as genera with an average of at least a 10% relative abundance per sample location and lactation stage are shown, and significant differences are highlighted (***, $P < 0.001$; **, $P < 0.01$; *, $P < 0.05$).

(Fig. 3 and 5). Late-lactation powder had significantly higher ($P < 0.01$) microbial alpha diversity than the mid-lactation powder (Fig. 2A).

Sample source and lactation stage significantly influence differences between the microbiota of the dairy samples. Overall, Adonis analysis from the R vegan package for Bray-Curtis beta diversity analysis showed significant differences between samples based on the lactation stage and source of the sample ($P \leq 0.001$). Twenty percent of the variation in the distance between samples was attributed to the sample source (BT/CT/WMS/SMS/SMP/cream), and 9.5% of the variation in the distance between samples was attributed to the lactation stage (ML, late lactation 1 [LL1], and LL2), with 4.7% being due to both the source and lactation stage.

DISCUSSION

This study set out to use molecular methods to provide an important description of the microbiota of a food processing pipeline by tracking the microbiota of raw milks on farms to a final skimmed milk powder. Through HTS, it was demonstrated that different production days and microbial selection by processing parameters can impact the microbiota during this process (Fig. 8). Through this approach, this study expands upon previous investigations (14, 19–21) to give an even greater understanding of the changes in the dairy microbiota from milk to skimmed milk powder. This investigation confirms the diversity of the raw-milk microbiota; however, it highlights that bulk tank

samples from mid- and late lactation broadly cluster together. Upon bulk refrigerated storage, there is a shift toward a psychrotolerance-dominant, processing facility-selected community that is dominated by *Pseudomonas*, *Acinetobacter*, and *Lactococcus*. This is consistent with a U.S. study by Kable et al. (14), which noted that large-scale silo storage led to the convergence of the microbiota into one of two community types. The subsequent fate of these bacteria following further product processing in our study showed that the relative abundances of these dominant psychrotrophic bacteria diminished following pasteurization. This was particularly apparent in mid-lactation samples, possibly due to a number of reasons not limited to the lactation stage and associated warmer weather, with an increase in thermophilic bacteria during the processing of these samples, with colder weather and lower production rates associated with late-lactation, psychrotroph-dominant samples. Continued processing of mid-lactation samples to a skimmed milk powder had a considerable impact on the microbiota. The resulting powder was dominated by the thermophiles *T. thermophilus* and *Geobacillus* sp. This dominance may be due to a number of factors, including, but not limited to, contamination from within the processing facility (5, 22) and/or their enrichment due to high-temperature treatment in SMP processing as well as the fact that the samples were collected from processing during a warmer time of year and when powder production was at its peak, with equipment running for longer periods and with higher throughput. The presence of *T. thermophilus* is notable as it was previously detected in environmental samples from a cheese manufacturing facility and has been shown to have a pinking defect in continental-type cheeses (18). *Geobacillus* is a high-heat-resistant thermophilic sporeformer known to contaminate dairy powders. The high heat in SMP manufacture allows *Geobacillus* to thrive when other microorganisms succumb to high-heat treatment, and its spore-forming ability facilitates subsequent survival in powdered dairy.

In contrast to mid-lactation samples, the microbiota of late-lactation samples did not change as considerably throughout processing. It is not clear if this was due to seasonality or potential process-related factors such as differences in cleaning practices used at the later sampling time point. It may also be that thermophilic or desiccation-tolerant bacteria were absent, or were present at only low levels, in the raw ingredients or processing system to take advantage of the high heat and drying in SMP processing on this processing day. Overall, significant differences between samples were accounted for by both the source of the sample and the lactation period. Ultimately, as this was the first such comprehensive study of its kind, further investigation is needed to determine how or why thermophilic species were selected for in one processing run on 1 day in a mid-lactation period and were not selected for in a processing run on 1 day in a late-lactation period. An increased understanding of why and how these differences occurred will undoubtedly aid dairy processors globally to implement effective measures and decrease undesirable microorganisms while increasing food safety and quality.

Although no previously undetected or unexpected taxa were identified, a number of potentially pathogenic bacteria were detected in the samples. One mid-lactation bulk tank sample in isolation had a high (>5%) relative abundance of *Yersinia*; however, this accounted for <0.4% of all processed samples, suggesting its elimination by heat treatments and processing. A number of *Clostridium sensu stricto* subgroups were determined to be present at significantly higher relative abundances in late lactation than in mid-lactation bulk tanks; however, the relative abundances of these subgroups also decreased following transport and processing. SUPER-FOCUS functional gene classification noted higher relative abundances of sporulation- and dormancy-associated functional gene groups in mid-lactation skimmed milk powder samples containing thermophilic spore-forming *Geobacillus* species than in all other samples. Although *Geobacillus* species are not pathogenic, they are extremely difficult to eliminate from the processing environment due to their heat tolerance and their ability to form resistant spores and biofilms, which can harbor pathogenic species (23, 24). *A. baumannii* was determined by shotgun metagenomic sequencing to be the dominant species of *Acineto-*

bacter. Although traditionally associated with opportunistic infections, drug resistance, and nosocomial infections, *A. baumannii* has been detected in animal products and was shown to have epidemiological characteristics that are different from those of strains that cause nosocomial infections (25). It has also been detected in bulk tank milks and dairy powders, leading the Food and Agriculture Organization of the United Nations (FAO) to classify it as one of the “category B organisms—causality plausible, but not yet demonstrated,” with respect to causing infant illness from powdered infant formula (26). The low number of shotgun metagenomic reads did not allow accurate strain-level identification or the identification of specific virulence-associated genes to facilitate a more in-depth investigation. Despite this, SUPER-FOCUS functional gene classification showed higher relative abundances of virulence-associated functional gene groups in raw-milk samples in tankers and whole-milk silos than in heat-processed samples in skimmed milk silos and skimmed milk powders. Overall, there is a general pattern of a greater relative abundance of potentially pathogenic genera and species in raw-milk samples and an increased relative abundance of spore-forming species present in mid-lactation skimmed milk powders. However, it should be emphasized that relative abundances are reported throughout this paper, and further analysis would be needed to confirm if there were increases and decreases in absolute abundances. It should be noted that the relative-abundance results obtained largely agree with data from corresponding culture analyses (27).

This study highlights that changes in the dairy microbiota throughout processing vary across processing days. It highlights the merits of using HTS to monitor processing-facility-induced changes in the dairy microbiota and potential future benefits of applying this technology more specifically to elucidate the basis for differences in dairy powder composition. Species-level composition analysis from shotgun metagenomic sequencing enables a more in-depth analysis than previously possible with 16S rRNA gene amplicon sequencing. Further development of this method opens the possibilities of routine microbiology testing, improved detection of sources of contamination, tracking of microorganisms throughout the food chain, and, in general, enhancing the ability of processors to make informed decisions to reduce risk and waste. There are a number of ways in which the approach taken in this study can be improved upon for commercial applications, such as combining it with quantitative approaches to obtain absolute numbers. Untargeted shotgun analysis can be subject to host DNA contamination (28), and dairy samples can have particularly high levels of host DNA, as somatic cell counts can be high in raw dairy. This can result in decreased yields of microbial DNA sequence, thereby limiting the number of reads available for comprehensive microbial analysis (see Fig. S4 in the supplemental material). The removal of host DNA prior to sequencing is possible by utilizing microbiome enrichment kits (29) and can be considered for performing a similar analysis in the future. However, this approach increases the overall cost. Another approach could be to enrich for specific subpopulations prior to DNA extraction (30). Developments such as these, combined with advances in portable sequencing technologies and further advances in the speed and accuracy of *in silico* tools, have the potential to greatly assist decision-making in this and other food chains.

Conclusions. In conclusion, this study provides detailed insight into the changes in the microbiota of dairy samples throughout a milk powder manufacturing process, on distinct sampling days. A notable change was observed upon large-volume pooling in the processing facility, which resulted in the dairy microbiota becoming dominated by psychrotolerant, spoilage-associated bacteria. Also of note were the impacts of processing and the processing facility on the microbiota. A pattern of particular note is that low levels of thermophilic bacteria present in raw ingredients, or within the processing facility, can potentially proliferate in the absence of competitors during and following processing and dominate the processed dairy product. With the routine implementation of these methods, an understanding of the reasons that lead to different species being dominant in the final product can be determined and lead to informed decisions

regarding product fate, in turn leading to increased food safety, reduced risk, and reduced economic losses.

MATERIALS AND METHODS

Sample collection. Raw milks, pasteurized milks, and powdered dairy products were sampled from within a commercial milk processing pipeline (Fig. 1). Raw-milk bulk tanks were sampled on one day during the early- to mid-lactation period (May 2016), here referred to as mid-lactation. These milks were combined and further processed to a skimmed milk powder. During this process, samples were also collected from collection tankers, the processor's whole-milk silo, the skimmed milk silo, and the resultant skimmed milk powders. Samples from raw-milk bulk tanks (October 2016) and the processing pipeline (December 2016), which in this instance also included sampling of the cream silo, were also collected later in the year to represent the late-lactation period. In order to complete the full process, a minimum capacity of milk was required, and considering the lower rate of herd production in late lactation, a greater number of farm bulk tanks was required. Therefore, during the late-lactation period, the original farms were resampled separately from the process (collection tanker to skimmed milk powder), which contained different farm bulk tank inputs ($n = 150$, not sampled). Samples were collected by personnel at the processing facility using standard collection procedures, and liquid samples were transported on ice to the laboratory, where DNA extraction was performed immediately. Skimmed milk powders were transported at room temperature and stored for up to 1 month prior to DNA extraction.

DNA extraction. DNA was extracted from fresh liquid dairy samples. Powders were stored at room temperature, and upon reconstitution at 10% (wt/vol) in one-quarter-strength Ringer's solution, DNA was extracted from 30-ml milk samples and 50-ml reconstituted skimmed milk powder samples using the MoBio PowerFood DNA isolation kit according to the manufacturer's instructions, with some minor adjustments. Samples were centrifuged at $5,000 \times g$ for 20 min at 4°C. Fat was removed (from raw-milk samples), and the supernatant was discarded. From there, pellets were washed and subjected to lysozyme treatment as previously reported (30). Twenty-eight microliters of proteinase K was added, and the mixture was incubated at 55°C for 15 min. Samples were centrifuged at $13,000 \times g$ for 1 min, and the supernatant was discarded. Pellets were resuspended in 450 μ l PF1 solution from the PowerFood kit, and from this point, the kit protocol was followed, including the recommended alternative lysis step for difficult-to-lyse cells. DNA was eluted in 50 μ l of elution buffer.

16S rRNA amplicon sequencing. Template DNA was quantified and checked for quality by using the Qubit double-stranded DNA (dsDNA) high-sensitivity assay kit as well as by running 2 μ l on a 1% agarose gel. DNA was normalized to 5 ng μ l⁻¹. The V3-V4 variable region of the 16S rRNA genes was amplified in triplicate from each sample as previously described (7), with a few changes: 35 PCR cycles were used instead of 32, and Kapa2G Robust (Kapa Biosystems Ltd.) was used instead of Kapa HiFi Hotstart. Two microliters of each PCR mixture was run on a 1% agarose gel to check for quality before pooling triplicate PCR mixtures, with cleaning with a 0.8 \times volume of AMPure XP beads. Fifty microliters of the cleaned-up sample was stored at -20°C. Five microliters was aliquoted and subjected to index PCR and cleanup according to Illumina 16S metagenomic sequencing library preparation guidelines, as previously described (7). DNA concentrations were quantified using the Qubit dsDNA high-sensitivity assay kit before diluting to 20 nM, pooling, and performing a final 1:1 AMPure XP cleanup step. The samples were pooled into 4 pools. Samples from each processing step were contained in each pool, with mid-lactation skimmed milk powder samples included in each pool as a control. The samples were sequenced on the Illumina MiSeq sequencing platform in the Teagasc sequencing facility using a 2-by-250 V2 kit according to Illumina sequencing protocols.

Whole-metagenome shotgun sequencing. A subset of samples was selected for whole-metagenome shotgun sequencing. These included 4 WMS (2 mid-lactation and 2 late lactation), 4 SMS (2 mid-lactation and 2 late lactation), 6 SMP (3 mid-lactation and 3 late lactation [1 from each bag]), as well as 2 pooled collection tanker (CT_P) samples, in which equal volumes of DNA from each of the 11 mid-lactation samples were pooled into one sample and equal volumes of each of the late-lactation tankers were pooled into a second sample. Samples were prepared according to the Nextera XT DNA library preparation guide from Illumina. Samples were sequenced on the Illumina MiSeq sequencing platform at the Teagasc sequencing facility with a 2-by-250 V2 kit with standard Illumina sequencing protocols.

Bioinformatic and statistical analyses. 16S rRNA gene amplicon sequences were processed as previously described (7). Briefly, forward and reverse reads were joined using USEARCH FLASH (fast length adjustment of short reads to improve genome assemblies) (31). Paired-end reads were further processed by quality filtering based on a quality score of 25 and removing mismatched barcodes and sequences below length thresholds by QIIME (32). USEARCH v7 (64-bit) (33) was utilized for removing noisy data, detecting chimeras, and clustering into operational taxonomic units (OTUs) at 97% identity. OTUs were aligned using PyNASt (Python Nearest Alignment Space Termination) (a flexible tool for aligning sequences to a template alignment [34]), and taxonomy was assigned using BLAST (35) against the SILVA SSURef database, release 123 (36). QIIME data were further analyzed using Phyloseq in R (37), richness was estimated to obtain alpha diversity, and distances were obtained for beta diversity before visualization using ggplot2 (38). Taxonomy was also visualized using ggplot2. A pairwise Wilcoxon rank sum test using Benjamini-Hochberg *P* value correction/false discovery rate (FDR) correction analysis was used to compare sample groups from mid- and late lactation. Adonis from the R vegan package was used to determine differences in beta diversity. Shotgun metagenomic data were processed as previously described (30). Briefly, raw metagenomic reads were checked for the presence of bovine reads,

which were removed; filtered based on the presence of quality and quantity; and trimmed to 170 bp with a combination of Picard tools and SAMtools (39). Kraken with a filter threshold of 0.2 (40) and SUPER-FOCUS (41) were used to determine microbial composition to the species level and biological functions, respectively.

Data availability. All sequence data have been deposited in the European Nucleotide Archive (ENA) under study accession number [PRJEB31110](https://www.ebi.ac.uk/ena/record/PRJEB31110).

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

FIG S1, TIF file, 1.3 MB.

FIG S2, PDF file, 0.6 MB.

FIG S3, TIF file, 0.5 MB.

FIG S4, TIF file, 0.3 MB.

FIG S5, TIF file, 0.5 MB.

TABLE S1, DOCX file, 0.05 MB.

ACKNOWLEDGMENTS

This research was funded by the Department of Agriculture, Food and the Marine (DAFM), under the FIRM project SACCP, reference number 14/F/883. Research in the Cotter laboratory is also funded by Science Foundation Ireland (SFI) under grant numbers SFI/12/RC/2273 (APC Microbiome Ireland) and SFI/16/RC/3835 (Vistamilk) and by the European Commission under the Horizon 2020 program under grant number 818368 (Master).

Special thanks to Lizandra Paludetti for facilitating sample collection.

REFERENCES

- Quigley L, McCarthy R, O'Sullivan O, Beresford TP, Fitzgerald GF, Ross RP, Stanton C, Cotter PD. 2013. The microbial content of raw and pasteurized cow milk as determined by molecular approaches. *J Dairy Sci* 96:4928–4937. <https://doi.org/10.3168/jds.2013-6688>.
- Quijada NM, Mann E, Wagner M, Rodríguez-Lázaro D, Hernández M, Schmitz-Esser S. 2018. Autochthonous facility-specific microbiota dominates washed-rind Austrian hard cheese surfaces and its production environment. *Int J Food Microbiol* 267:54–61. <https://doi.org/10.1016/j.ijfoodmicro.2017.12.025>.
- Yeluri Jonnala BR, McSweeney PLH, Sheehan JJ, Cotter PD. 2018. Sequencing of the cheese microbiome and its relevance to industry. *Front Microbiol* 9:1020. <https://doi.org/10.3389/fmicb.2018.01020>.
- Fretin M, Martin B, Rifa E, Isabelle VM, Pomies D, Ferlay A, Montel MC, Delbes C. 2018. Bacterial community assembly from cow teat skin to ripened cheeses is influenced by grazing systems. *Sci Rep* 8:200. <https://doi.org/10.1038/s41598-017-18447-y>.
- Cho TJ, Kim HW, Kim NH, Park SM, Kwon JI, Kim YJ, Lee KW, Rhee MS. 2018. New insights into the thermophilic spore-formers in powdered infant formula: implications of changes in microbial composition during manufacture. *Food Control* 92:464–470. <https://doi.org/10.1016/j.foodcont.2018.05.036>.
- Wu S, Jiang Y, Lou B, Feng J, Zhou Y, Guo L, Forsythe SJ, Man C. 2018. Microbial community structure and distribution in the air of a powdered infant formula factory based on cultivation and high-throughput sequence methods. *J Dairy Sci* 101:6915–6926. <https://doi.org/10.3168/jds.2017-13968>.
- Doyle CJ, Gleeson D, O'Toole PW, Cotter PD. 2017. Impacts of seasonal housing and teat preparation on raw milk microbiota: a high-throughput sequencing study. *Appl Environ Microbiol* 83:e02694-16. <https://doi.org/10.1128/AEM.02694-16>.
- Sadiq FA, Flint S, He G. 2018. Microbiota of milk powders and the heat resistance and spoilage potential of aerobic spore-forming bacteria. *Int Dairy J* 85:159–168. <https://doi.org/10.1016/j.idairyj.2018.06.003>.
- Zhao Y, Kumar M, Caspers MPM, Nierop Groot MN, van der Vossen JMBM, Abee T. 2018. Short communication: growth of dairy isolates of *Geobacillus thermoglucosidans* in skim milk depends on lactose degradation products supplied by *Anoxybacillus flavithermus* as secondary species. *J Dairy Sci* 101:1013–1019. <https://doi.org/10.3168/jds.2017-13372>.
- Doyle CJ, O'Toole PW, Cotter PD. 2018. Genomic characterization of sulphite reducing bacteria isolated from the dairy production chain. *Front Microbiol* 9:1507. <https://doi.org/10.3389/fmicb.2018.01507>.
- Fricke M, Reissbrodt R, Ehling-Schulz M. 2008. Evaluation of standard and new chromogenic selective plating media for isolation and identification of *Bacillus cereus*. *Int J Food Microbiol* 121:27–34. <https://doi.org/10.1016/j.ijfoodmicro.2007.10.012>.
- Li N, Wang Y, You C, Ren J, Chen W, Zheng H, Liu Z. 2018. Variation in raw milk microbiota throughout 12 months and the impact of weather conditions. *Sci Rep* 8:2371. <https://doi.org/10.1038/s41598-018-20862-8>.
- Doyle CJ, Gleeson D, O'Toole PW, Cotter PD. 2017. High-throughput metataxonomic characterization of the raw milk microbiota identifies changes reflecting lactation stage and storage conditions. *Int J Food Microbiol* 255:1–6. <https://doi.org/10.1016/j.ijfoodmicro.2017.05.019>.
- Kable ME, Srisengfa Y, Laird M, Zaragoza J, McLeod J, Heidenreich J, Marco ML. 2016. The core and seasonal microbiota of raw bovine milk in tanker trucks and the impact of transfer to a milk processing facility. *mBio* 7:e00836-16. <https://doi.org/10.1128/mBio.00836-16>.
- Rodrigues MX, Lima SF, Canniatti-Brazaca SG, Bicalho RC. 2017. The microbiome of bulk tank milk: characterization and associations with somatic cell count and bacterial count. *J Dairy Sci* 100:2536–2552. <https://doi.org/10.3168/jds.2016-11540>.
- Walsh AM, Crispie F, Daari K, O'Sullivan O, Martin JC, Arthur CT, Claesson MJ, Scott KP, Cotter PD. 2017. Strain-level metagenomic analysis of the fermented dairy beverage nunu highlights potential food safety risks. *Appl Environ Microbiol* 83:e01144-17. <https://doi.org/10.1128/AEM.01144-17>.
- Wolfe BE, Button JE, Santarelli M, Dutton RJ. 2014. Cheese rind communities provide tractable systems for in situ and in vitro studies of microbial diversity. *Cell* 158:422–433. <https://doi.org/10.1016/j.cell.2014.05.041>.
- Quigley L, O'Sullivan DJ, Daly D, O'Sullivan O, Burdikova Z, Vana R, Beresford TP, Ross RP, Fitzgerald GF, McSweeney PL, Giblin L, Sheehan JJ, Cotter PD. 2016. Thermus and the pink discoloration defect in cheese. *mSystems* 1:e00023-16. <https://doi.org/10.1128/mSystems.00023-16>.
- Alessandria V, Ferrocino I, De Filippis F, Fontana M, Rantsiou K, Ercolini D, Cocolin L. 2016. Microbiota of an Italian grana-like cheese during manufacture and ripening, unraveled by 16S rRNA-based approaches. *Appl Environ Microbiol* 82:3988–3995. <https://doi.org/10.1128/AEM.00999-16>.
- Parente E, Cocolin L, De Filippis F, Zotta T, Ferrocino I, O'Sullivan O,

- Neviani E, De Angelis M, Cotter PD, Ercolini D. 2016. FoodMicrobionet: a database for the visualisation and exploration of food bacterial communities based on network analysis. *Int J Food Microbiol* 219:28–37. <https://doi.org/10.1016/j.ijfoodmicro.2015.12.001>.
21. Stellato G, Utter DR, Voorhis A, De Angelis M, Eren AM, Ercolini D. 2017. A few *Pseudomonas* oligotypes dominate in the meat and dairy processing environment. *Front Microbiol* 8:264. <https://doi.org/10.3389/fmicb.2017.00264>.
 22. Miller RA, Kent DJ, Watterson MJ, Boor KJ, Martin NH, Wiedmann M. 2015. Spore populations among bulk tank raw milk and dairy powders are significantly different. *J Dairy Sci* 98:8492–8504. <https://doi.org/10.3168/jds.2015-9943>.
 23. Burgess SA, Flint SH, Lindsay D. 2014. Characterization of thermophilic bacilli from a milk powder processing plant. *J Appl Microbiol* 116:350–359. <https://doi.org/10.1111/jam.12366>.
 24. Gopal N, Hill C, Ross PR, Beresford TP, Fenelon MA, Cotter PD. 2015. The prevalence and control of *Bacillus* and related spore-forming bacteria in the dairy industry. *Front Microbiol* 6:1418. <https://doi.org/10.3389/fmicb.2015.01418>.
 25. Hamouda A, Findlay J, Al Hassan L, Amyes SGB. 2011. Epidemiology of *Acinetobacter baumannii* of animal origin. *Int J Antimicrob Agents* 38:314–318. <https://doi.org/10.1016/j.ijantimicag.2011.06.007>.
 26. Cho G-S, Li B, Rostalsky A, Fiedler G, Rösch N, Igbinosa E, Kabisch J, Bockelmann W, Hammer P, Huys G, Franz CMAP. 2018. Diversity and antibiotic susceptibility of *Acinetobacter* strains from milk powder produced in Germany. *Front Microbiol* 9:536. <https://doi.org/10.3389/fmicb.2018.00536>.
 27. Paludetti LF, Kelly AL, O'Brien B, Jordan K, Gleeson D. 2019. Microbiological quality of milk from farms to milk powder manufacture: an industrial case study. *J Dairy Res* 86:242–247. <https://doi.org/10.1017/S0022029919000347>.
 28. Sharpton TJ. 2014. An introduction to the analysis of shotgun metagenomic data. *Front Plant Sci* 5:209. <https://doi.org/10.3389/fpls.2014.00209>.
 29. Feehery GR, Yigit E, Oyola SO, Langhorst BW, Schmidt VT, Stewart FJ, Dimalanta ET, Amaral-Zettler LA, Davis T, Quail MA, Pradhan S. 2013. A method for selectively enriching microbial DNA from contaminating vertebrate host DNA. *PLoS One* 8:e76096. <https://doi.org/10.1371/journal.pone.0076096>.
 30. McHugh AJ, Feehily C, Tobin JT, Fenelon MA, Hill C, Cotter PD. 2018. Mesophilic sporeformers identified in whey powder by using shotgun metagenomic sequencing. *Appl Environ Microbiol* 84:e01305-18. <https://doi.org/10.1128/AEM.01305-18>.
 31. Magoc T, Salzberg SL. 2011. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* 27:2957–2963. <https://doi.org/10.1093/bioinformatics/btr507>.
 32. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Peña AG, Goodrich JK, Gordon JJ, Huttley GA, Kelley ST, Knights D, Koenig JE, Ley RE, Lozupone CA, McDonald D, Muegge BD, Pirrung M, Reeder J, Sevinsky JR, Turnbaugh PJ, Walters WA, Widmann J, Yatsunenko T, Zaneveld J, Knight R. 2010. QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* 7:335–336. <https://doi.org/10.1038/nmeth.f.303>.
 33. Edgar RC. 2010. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26:2460–2461. <https://doi.org/10.1093/bioinformatics/btq461>.
 34. Caporaso JG, Bittinger K, Bushman FD, DeSantis TZ, Andersen GL, Knight R. 2010. PyNAST: a flexible tool for aligning sequences to a template alignment. *Bioinformatics* 26:266–267. <https://doi.org/10.1093/bioinformatics/btp636>.
 35. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol* 215:403–410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2).
 36. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO. 2013. The SILVA ribosomal RNA gene database project: improved data processing and Web-based tools. *Nucleic Acids Res* 41:D590–D596. <https://doi.org/10.1093/nar/gks1219>.
 37. McMurdie PJ, Holmes S. 2013. phyloseq: an R Package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One* 8:e61217. <https://doi.org/10.1371/journal.pone.0061217>.
 38. Wickham H. 2009. ggplot2: elegant graphics for data analysis. Springer, New York, NY.
 39. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25:2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>.
 40. Wood DE, Salzberg SL. 2014. Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biol* 15:R46. <https://doi.org/10.1186/gb-2014-15-3-r46>.
 41. Silva GG, Green KT, Dutilh BE, Edwards RA. 2016. SUPER-FOCUS: a tool for agile functional analysis of shotgun metagenomic data. *Bioinformatics* 32:354–361. <https://doi.org/10.1093/bioinformatics/btv584>.