CORA Cork Open Research Archive
Cartlann Taighde Oscailte Chorcaí

Title	Leveraging community-based sequencing approaches to characterise the dairy microbiome
Authors	Yap, Min
Publication date	2022
Original Citation	Yap, M. 2022. Leveraging community-based sequencing approaches to characterise the dairy microbiome. PhD Thesis, University College Cork.
Type of publication	Doctoral thesis
Rights	© 2022, Min Yap. - https://creativecommons.org/licenses/by-nc-nd/4.0/
Download date	2026-11-16 09:56:53
Item downloaded from	https://hdl.handle.net/10468/14131



UCC
University College Cork, Ireland
Coláiste na hOllscoile Corcaigh

University College Cork, Ireland
Coláiste na hOllscoile Corcaigh

Leveraging community-based sequencing approaches to characterise the dairy microbiome

A thesis presented to the National University of Ireland for the degree of Doctor of Philosophy

By

Min Yap B.Sc

Teagasc Food Research Centre, Moorepark, Fermoy Co. Cork, Ireland

School of Microbiology, University College Cork, Cork, Ireland

October 2022

Department head: Dr David Clarke

Research supervisors: Professor Paul Cotter, Dr Orla O'Sullivan and Professor Paul O'Toole

Agriculture and Food Development Authority

Table of contents

Declaration i

Abstract ii

Publications v

List of Abbreviations vi

List of Figures ix

List of Tables xi

Chapter 1. Next-generation food research: use of meta-omic approaches for characterising microbial communities along the food chain 1

1.1 Abstract 2

1.2 Introduction 2

1.3 Importance of microorganisms through the food chain 4

1.3.1 Quality: flavour, texture, fermentation and spoilage 4

1.3.2 Safety: pathogens and microbial antagonism 8

1.4 Traditional approaches to characterisation of microbes 10

1.5 Meta-omic approaches (community approaches) 13

1.5.1 Sequencing-based meta-omic approaches: metagenetics, metagenomics and metatranscriptomics 15

1.5.2 Other meta-omic approaches: metaproteomics, meta-metabolomics... 17

1.5.3 Current challenges in sequencing-based meta-omic approaches 18

1.6 Applications of sequencing-based meta-omic approaches along the food chain 21

1.6.1 Public health applications 21

1.6.2 Food industry applications 27

1.6.3 A dairy perspective 36

1.7 Future issues 38

1.8 Summary points 40

1.9 Acknowledgements including funding source 41

1.10 References 41

Chapter 2. Seasonality and geography have a greater influence than the use of chlorine-based cleaning agents on the microbiota of bulk tank raw milk. 61

2.1 Abstract 62

2.1.1 Importance 63

2.2 Introduction 63

2.3 Materials and Methods 66

2.3.1 Sample collection and preparation. 66

2.3.2 DNA extraction. 67

2.3.3 DNA library preparation and shotgun metagenomic sequencing. 67

2.3.4 Bioinformatic analysis of shotgun metagenomic data. 67

2.3.5 Climate data. 68

2.3.6 Statistical analysis and data visualisation. 68

2.4 Results 69

2.4.1 Sampling month and location had greater influence on the taxonomic diversity and profiles of the bulk tank milk microbiota than cleaning method. 69

2.4.2 Similarly, functional characteristics of the bulk tank raw milk microbiome were influenced by sampling month and location 76

2.4.3 Higher number of antibiotic resistance gene marker reads were mapped to samples from farms which used chlorine-based cleaning agents compared to other cleaning methods 77

2.5 Discussion 78

2.6 Declarations 84

2.6.1 Data availability 84

2.6.2 Acknowledgements including funding source 84

2.7 References 85

2.8 Supplemental Material 91

Chapter 3. Evaluation of methods for the reduction of contaminating host reads when performing shotgun metagenomic sequencing of the milk microbiome... 111

3.1 Abstract 112

3.2 Introduction 113

3.3 Materials and Methods 116

3.3.1 Collection, storage and initial treatment of milk samples 116

3.3.2 Description of mock community (positive control) and negative controls utilised 117

3.3.3 Host depletion, microbial enrichment, and DNA extraction 117

3.3.4 DNA library preparation 118

3.3.5 Bioinformatic and statistical analysis of shotgun metagenomic data 119

3.4 Results 120

3.4.1 Efficiency of methods for improving the proportion of microbial reads. 120

3.4.2 Differential performance of taxonomic classification algorithms on spiked samples 122

3.4.3 Analysis of bovine and human milk samples – community structure and taxonomic and functional profiles 126

3.4.4 Characterisation of the milk microbiome 129

3.5 Discussion 132

3.6 Declarations 136

3.6.1 Ethics approval and consent to participate 136

3.6.2 Data availability 136

3.6.3 Acknowledgements including funding source 136

3.7 References 137

3.8 Supplemental Material 143

Chapter 4. Seasonal and geographical impact on the Irish bulk tank milk microbiota correlates with chemical composition and climatic variables. 149

4.1 Abstract 150

4.2 Introduction 151

4.3 Materials and Methods 153

4.3.1 Sample collection and preparation. 153

4.3.2 DNA extraction. 153

4.3.3 DNA library preparation and shotgun metagenomic sequencing. 154

4.3.4 Bioinformatic analysis of shotgun metagenomic data. 154

4.3.5 Statistical analysis and data visualisation. 155

4.4 Results 156

4.4.1 The Irish bulk tank milk microbiota is highly diverse, but there is a core microbiota. 156

4.4.2 Both season and location influence the bulk tank milk microbiota. .159

4.4.3 Functional and antibiotic resistance profiles differed by season and sampling location. 166

4.4.4 The bulk tank milk microbiota correlated with climatic factors and chemical composition. 169

4.5 Discussion 174

4.6 Declarations.....179

4.6.1 Data availability.....179

4.6.2 Acknowledgements including funding source 179

4.7 References.....180

4.8 Supplemental Material.....186

Chapter 5. Development of sequencing-based methodologies to distinguish viable from non-viable cells in a bovine milk matrix: a pilot study. 197

5.1 Abstract 198

5.1.1 Importance.....199

5.2 Introduction.....199

5.3 Materials and Methods 202

5.3.1 Model community and samples.....202

5.3.2 PMA treatment and DNA extraction 203

5.3.3 Illumina DNA library preparation and shotgun metagenomic sequencing 204

5.3.4 MDA amplification and Oxford Nanopore DNA metagenomic library preparation 205

5.3.5 RNA extraction 205

5.3.6 cDNA synthesis 206

5.3.7 Illumina 16S-rRNA library preparation and sequencing 206

5.3.8 Oxford Nanopore 16S-rRNA library preparation and sequencing. 207

5.3.9 rRNA depletion 207

5.3.10 Illumina metatranscriptomics library preparation and sequencing. 208

5.3.11 Oxford Nanopore metatranscriptomics library preparation and sequencing. 208

5.3.12 Bioinformatic analysis 209

5.3.13 Statistical analysis and data visualisation. 210

5.4 Results 211

5.4.1 Taxonomic classification of model community strains was accurate to genus level across all methods. 212

5.4.2 Methods were accurate though the precision and sensitivity varied. 215

5.4.3 Significant differences were observed between data depending on library types (DNA and RNA) and sequencing methods (Illumina and ONT)...216

5.5 Discussion 221

5.6 Declarations....227

5.6.1 Data availability....227

5.6.2 Acknowledgements including funding source 227

5.7 References....228

5.8 Supplemental Material....234

Chapter 6. General Discussion 240

6.1 Discussion 241

6.2 Potential applications....246

6.3 Future work 248

6.4 Conclusion 251

6.5 References....252

Acknowledgements....256

Declaration

I hereby certify that this material, which I now submit for assessment on the programme of study, leading to the award of PhD is entirely my own work (except as declared, hereafter) and has not been submitted for another degree, either at University College Cork or elsewhere.

Signed: 

Min Yap

Student number: 118222019

Date: 06/10/2022

Abstract

Microorganisms exist in communities along the food chain and contribute, positively and negatively, to the safety and quality of food products downstream. While traditional culture-based methods usually investigate specific groups of microorganisms, advances in high-throughput DNA sequencing have enabled the study of entire microbial communities in food or food-related samples. In this thesis, shotgun metagenomic sequencing was employed to investigate microbial communities in bulk tank milk. First, shotgun sequencing was used to determine the impact of farm cleaning methods on the microbiome. The results revealed that seasonality and geography significantly impacted the bulk tank milk microbiota but that the use of chlorine-based relative to chlorine-free cleaning methods did not result in differences. Next, further sampling was performed to determine if interactions exist between the bulk tank milk microbiota, raw milk chemical composition and climatic variables. Analysis of shotgun sequencing data from bulk tank milk over 12-months revealed a stable core microbiota and showed a significant and obvious influence of season and location on the raw milk microbiota. Associations were detected between pathogenic and mastitis-related taxa and chemical composition and climate/environmental variables. The interconnected nature of these variables suggests there is merit in considering the use of chemical and/or climate variables as markers for food safety and spoilage threats in bulk tank milk. While useful information can be obtained through shotgun metagenomic sequencing, the untargeted nature of this approach involves the sequencing of all DNA present in samples, including host-derived reads that are present in high proportions in milk (~90%), as well as both live and dead cells, which could alter

inferred community structure or diversity. To overcome these challenges, two studies were completed to improve current shotgun metagenomic sequencing methodology. First, three methods were evaluated to assess their relative ability to deplete host DNA or enrich microbial DNA. The MolYsis™ Complete5 kit yielded significantly higher proportions of microbial reads and improved sequencing depth, allowing for further characterisation of the bovine and human milk microbiomes through the recovery of metagenome-assembled genomes (MAGs). Tackling the issue of microbial viability, 4 methods were explored to compare their relative efficacy at distinguishing between viable and non-viable cells with a five-strain live and heat-inactivated model microbial community spiked into a bovine milk matrix and sequenced on Illumina and Oxford Nanopore Technologies (ONT) platforms. The results showed that all methods were relatively accurate, but significant differences between library type (DNA and RNA) and sequencing technologies (Illumina and ONT) were found. Variations between library types were due to the differing molecular targets, while sequencing depths, error rates and bioinformatic pipelines likely contributed to the differences between sequencing technologies. This pilot study provides insights into the potential enhancement of sequencing-based approaches to better characterise and differentiate live and dead microbes in food-related environments, but more work is required to optimise these methods before more widespread use. Overall, the use of shotgun metagenomics provided valuable insights into the bulk tank raw milk microbiota, improving our current understanding of the factors that influence the quality and safety of raw milk. Building on the work performed during this thesis, further improvements can enhance existing community-based sequencing methods to allow for better understanding of food

microbiomes, that in turn would be beneficial for food safety, quality, and public health.

Publications

Yap, M., Ercolini, D., Álvarez-Ordóñez, A., O'Toole, P. W., O'Sullivan, O., & Cotter, P. D. (2022). Next-generation food research: use of meta-omic approaches for characterizing microbial communities along the food chain. *Annual Review of Food Science and Technology*, 13, 361-384.

Yap, M., Feehily, C., Walsh, C. J., Fenelon, M., Murphy, E. F., McAuliffe, F. M., ... & Cotter, P. D. (2020). Evaluation of methods for the reduction of contaminating host reads when performing shotgun metagenomic sequencing of the milk microbiome. *Scientific reports*, 10(1), 1-11.

Yap, M., Gleeson, D., O'Toole, P. W., O'Sullivan, O., & Cotter, P. D. (2021). Seasonality and geography have a greater influence than the use of chlorine-based cleaning agents on the microbiota of bulk tank raw milk. *Applied and environmental microbiology*, 87(22), e01081-21.

Yap, M., O'Sullivan, O., O'Toole, P.W., & Cotter, P. D. (2022). Development of sequencing-based methodologies to distinguish viable from non-viable cells in a bovine milk matrix: A pilot study. *Frontiers in Microbiology*, 13.

List of Abbreviations

A	Accuracy
AMR	Antimicrobial resistance
ARGs	Antibiotic resistance genes
BTCF	Use of chlorine-free cleaning products for the bulk tank only
C	Use of traditional chlorine cleaning protocols
CARD	Comprehensive antibiotic resistance database
cDNA	Complementary DNA
CF	Use of chlorine-free cleaning methods throughout the system
CFU	Colony forming units
DNA	Deoxyribonucleic acid
DPC	Dairy Production Centre
EFSA	European Food Safety Authority
ELISA	Enzyme-linked immunosorbent assay
EMA	Ethidium monoazide
ENA	European Nucleotide Repository
F	F-score (the harmonic mean of precision and sensitivity)
FN	False negative
FP	False positive
Gbp	Giga base pairs
gDNA	Genomic DNA
HACCP	Hazard analysis critical control point
i-16s	Illumina-sequenced 16S rRNA

i-metaT	Illumina-sequenced metatranscriptomics
i-pma	Illumina-sequenced PMA-treated shotgun metagenomics
i-shotgun	Illumina-sequenced shotgun metagenomics
ITS	Internal transcribed spacer
LAB	Lactic acid bacteria
LCA	Lowest-common ancestor
MAGs	Metagenome-assembled genomes
MALDI-TOF MS	Matrix assisted laser desorption ionization-time of flight mass spectrometry
MAP	Modified atmosphere packaging
MDA	Multiple displacement amplification
ML	MolYsis™ Complete5 kit
mRNA	Messenger RNA
NEB	NEBNext Microbiome Enrichment kit, with extraction using the DNeasy PowerSoil Pro kit
NPN	Non-protein nitrogen
o-16S	ONT-sequenced 16S rRNA
o-metaT	ONT-sequenced metatranscriptomics
ONT	Oxford Nanopore Technologies
o-pma	ONT-sequenced PMA-treated shotgun metagenomics
o-shotgun	ONT-sequenced shotgun metagenomics
P	Precision
PBS	Phosphate buffered saline
PCoA	Principal coordinate analysis

PCR	Polymerase chain reaction
PDO	Protected designation of origin
PERMANOVA	Permutational multivariate analysis of variance
PMA	Propidium monoazide
PS	DNeasy PowerSoil Pro kit
qPCR	Quantitative polymerase chain reaction
RDA	Redundancy analysis
RNA	Ribonucleic acid
RPKM	Normalised reads per kilobase per million reads
rRNA	Ribosomal RNA
S	Sensitivity
SFI	Science Foundation Ireland
SNP	Single nucleotide polymorphism
STEC	Shiga toxin-producing E. coli
TA	Titratable acidity
TN	True negative
TP	True positive
TSA	Tryptic soy agar
UHT	Ultra-heat treatment
VBNC	Viable but not culturable

List of Figures

Figure 1.1. Current meta-omic approaches used in microbiome research. 14

Figure 1.2. Current applications of meta-omic sequencing-based approaches along the food chain. 27

Figure 2.1. Diversity analysis of the microbiota of bulk tank milk samples. 71

Figure 2.2. Taxonomic profiles of the bulk tank milk microbiota at relative abundances greater than 0.01%. 75

Figure 2.3. Functional profiles of the bulk tank milk microbiota. 77

Figure 2.4. Antibiotic resistance genes (ARGs) found in milk samples from farms with different cleaning methods. 78

Supplemental Figure 2.1. Alpha and beta diversity of raw milk stratified by cleaning method. 92

Supplemental Figure 2.2. Significantly differential relative abundances of species that differed between sampling month and location. 94

Supplemental Figure 2.3. Redundancy analysis of the relationship between climate factors and sampling locations of the bulk tank milk microbiota. 98

Supplemental Figure 2.4. Significantly differential abundances of functions that differed between sampling months. 99

Figure 3.1. Experimental procedure used for the study. 121

Figure 3.2. ML kit produced significantly higher proportions of microbial reads... 122

Figure 3.3. Mock community samples impacted by choice of bioinformatics classifier rather than method. 125

Figure 3.4. Both the bovine and human milk samples each had similar species level composition and community diversity. 128

Figure 3.5. ML kit generated higher numbers of observable species and metagenome assembled genomes in milk samples. ……. 130

Supplemental Figure 3.1. Initial extraction plan. …….. 143

Supplemental Figure 3.2. Functional analysis of bovine and human milk samples as based on gene ontology domains. ……..…. 144

Figure 4.1. Species level composition of the bulk milk microbiota. …… 159

Figure 4.2. Diversity of bulk tank milk microbiota. ……..…. 160

Figure 4.3. Taxonomic profiles and the core microbiota of the bulk tank milk microbiota. ……..…. 163

Figure 4.4. Functional characteristics of the bulk tank milk microbiota based on SUPERFOCUS analysis. ……..…. 168

Figure 4.5. Composition of the number of contigs mapped to genes from different drug resistance classes of the bulk tank milk microbiota. ……..…. 169

Figure 4.6. Correlation of climatic factors and chemical data and the raw milk microbiota. ……..…. 173

Figure 5.1 Experimental design of study. …… 212

Figure 5.2. Overall performance of methods based on the accuracy, F-score, precision and sensitivity. …… 216

Figure 5.3. Beta diversity and taxonomic profiles of samples. …… 218

Figure 5.4. Differences between Shotgun and PMA-shotgun methods. …….. 220

Supplemental Figure 5.1. Abundances of samples spiked into Live and Dead conditions. …….. 235

Supplemental Figure 5.2. Performance metrics of samples based on comparison of abundances with the live and dead model community controls. ……. 235

List of Tables

Table 1.1. Some classical examples of types of food spoilage that can be caused by different organisms. …… 5

Table 2.1. PERMANOVA analysis of the bulk tank milk microbiota. …….. 72

Supplemental Table 2.1. Sample numbers for each cleaning method and location. …… 91

Supplemental Table 2.2. PERMANOVA analysis of taxonomic composition and functional properties of the bulk tank milk microbiota stratified by cleaning method. …… 93

Supplemental Table 2.3. Relative abundances of species that differed by sampling month. …… 95

Supplemental Table 2.4. Relative abundances of species that differed by sampling location. …… 96

Supplemental Table 2.5. Climate data retrieved from the Irish Meteorological Service from the sampling days at the four sampling locations. …….. 97

Supplemental Table 2.6. Relative abundances of genes assigned to significant functional classifications between sampling months. …….. 100

Table 3.1. High-quality metagenome-assembled genomes from shotgun sequencing analysis. …… 130

Supplementary Table 3.1. Summary of sequencing quality control data for each sample. …… 145

Supplemental Table 3.2. List of species (> 1% relative abundance) assigned to negative control samples by each classifier. …… 147

Table 4.1. Sample metadata and the degree to which their variation explains differences in the taxonomic composition and functional profiles of the bulk tank milk microbiota. …… 161

Table 4.2. Sample size, chemical composition and climatic variables of bulk tank raw milk across season. 171

Table 4.3. Sample size, chemical composition, and climatic variables of bulk tank raw milk across locations. 172

Supplemental table 4.1. Species assignment into 5 categories (environment, host, pathogen, spoilage or technologically relevant). 186

Supplemental Table 4.2. High-quality MAGs. 188

Supplemental Table 4.3. Relative abundances of genera and subsystem level 1 functions that differed significantly by sampling season. 193

Supplemental Table 4.4. Relative abundances of genera and subsystem level 1 functions that differed significantly by sampling location. 195

Table 5.1. The highest resolution of taxonomic identity identified for the model community strains across all methods. 214

Supplemental Table 5.1. Sequencing quality metrics for each method. 234

Supplemental Table 5.2. Other taxa found in milk samples at relative abundances above 0.1%. 236

Supplemental Table 5.3. Taxa found in control milk samples at relative abundances above 0.1%. 238

Chapter 1. Next-generation food research: use of meta-omic approaches for characterising microbial communities along the food chain

Updated for this thesis, with the addition of section 1.6.3, since publication in Annual Review of Food Science and Technology
(doi: 10.1146/annurev-food-052720-010751)

Authors: Min Yap, Danilo Ercolini, Avelino Álvarez-Ordóñez, Paul W. O'Toole, Orla O'Sullivan and Paul D. Cotter

Contributions: Candidate drafted and edited the review.
Review was revised and edited by DE, AAO, PWO, OO and PDC.

1.1 Abstract

Microorganisms exist along the food chain and impact the quality and safety of foods in both positive and negative ways. Identifying and understanding the behaviour of these microbial communities enables the implementation of preventative or corrective measures in public health and food industry settings. Current culture-dependent microbial analyses are time consuming and only target specific subsets of microbes. However, the greater use of culture-independent meta-omic approaches has the potential to facilitate a thorough characterisation of the microbial communities along the food chain. Indeed, these methods have shown potential in contributing to outbreak investigation, ensuring food authenticity, assessing the spread of antimicrobial resistance, tracking microbial dynamics during fermentation and processing, and uncovering the factors along the food chain that impact food quality and safety. This review examines the community-based approaches, particularly the application of sequencing-based meta-omics strategies, for characterising microbial communities along the food chain.

1.2 Introduction

Microorganisms along the food chain from farm to fork influence food quality and safety. Historically, culture-based techniques have been used extensively to characterise these microbes. However, with the development of molecular methods and high-throughput sequencing technologies, culture-independent techniques have become more relevant to food microbiome analysis. This has resulted in a corresponding shift from the investigation of specific taxa or groups of

food microorganisms to a broader community-based analysis (Cocolin and Ercolini 2015). These methods are based on the extraction of nucleic acids, proteins, and/or metabolites, allowing for the detection and characterisation of microbes within an environment without the need for culturing. As the occurrence of and interactions between microorganisms impact the quality and safety of food, a deeper understanding of these processes allows for early interventions to adverse food safety events, ensure optimal food quality, and to identify the source of desirable or undesirable microorganisms.

Despite being labour-intensive and time-consuming, culture methods are still the methods employed most regularly in the food industry (Dwivedi and Jaykus 2011, Sohier, Pavan et al. 2014). However, one of the biggest limitations is that these approaches frequently detect only a fraction of the microbes that are present in the sample as they rely on the isolation and growth of single microbes on culture media whose metabolic and physiological requirements can be reproduced in vitro. This not only overlooks the portion of microbes that are viable but not culturable (VBNC), but also fails to consider the relationships within the community of bacteria present in the sample (Cao, Fanning et al. 2017). Uncultured microorganisms are estimated to account for up to 99% of the microorganisms in many environments, meaning that the use of traditional culture methods causes a gross underestimation of the microbial population (Handelsman 2004). While this level of underestimation may not be as considerable in food systems, nevertheless because the microbiomes of food and food processing environments are composed of complex, dynamic

microbial communities, meta-omic approaches have the potential to provide a more accurate and greater understanding of these communities.

In this review, the contribution of microorganisms to the quality and safety of food and the traditional approaches to microbial characterisation are briefly described. The main focus of the review is on sequencing-based meta-omic approaches and their contribution to understanding microbial community dynamics in food, food-associated environments and along the food processing microbiome. Other non-sequencing-based meta-omic approaches are also mentioned in brief.

1.3 Importance of microorganisms through the food chain

1.3.1 Quality: flavour, texture, fermentation and spoilage

Food quality is often associated with physical parameters such as pH and moisture content, which can influence the growth and survival of microorganisms within a food and the food chain. Food spoilage is a process or change that renders a product undesirable or unacceptable for consumption, which is impacted by both the food's intrinsic characteristics and the extrinsic environment (Blackburn 2006). It is a complex process, whereby food undergoes biochemical changes, often due to microbial activity according to ecological determinants (Nychas and Panagou 2011). Some common spoilage bacteria include *Pseudomonas* spp., *Shewanella* spp., *Bacillus* spp., *Clostridium* spp., lactic acid bacteria (LAB) and Enterobacteriaceae (Blackburn 2006). Ultimately, different bacteria cause varying quality problems in different types of food, with some examples presented in Table 1.1. Furthermore, despite their slower growth rate, yeasts and moulds are able to exploit many

ecological niches in food systems, and can utilise substrates and tolerate extreme conditions that are not possible for bacteria (in't Veld 1996, Petruzzi, Corbo et al. 2017). Common spoilage yeasts include species of *Zygosacchromyces* (in high sugar foods), *Saccharomyces* (a cause of gassiness and turbidity in wines), or *Candida* (off-flavours in meat and dairy products) and common spoilage moulds include *Zygomycetes* (in produce with high water content), *Penicillium* spp. (cause rot in fruits) and *Aspergillus* (in grains, spices and nuts) (Sahu and Bala 2017).

Table 1.1. Some classical examples of types of food spoilage that can be caused by different organisms.

Spoilage characteristic	Spoilage organism in food type	References
Off-odour and off-flavours	*Pseudomonas* spp., *Carnobacterium* spp., *Serratia* spp., *Leuconostoc* spp. and *Brochothrix thermosphacta* produces off-odours and off-flavours in meat, fish and poultry *Shewanella putrefaciens* causes rancid and sulphurous odours and *Aeromonas* spp. produces a sour flavour in smoked salmon Various Enterobacteriaceae causes off-odours and off-flavours in preserved seafood products *Citrobacter* and *Proteus* has been found to cause off-odours on poultry	Blackburn (2006) Stohr, Joffraud et al. (2001) Fleet (2011)

	Candida spp. and *Kluyveromyces* spp. cause off-odours and flavours in fermented dairy products.	
Changes in texture	*Pseudomonas* spp. causes meat and poultry to become slimy/mushy due to the action of degradative enzymes. LAB can cause poor texture in cheese *Bacillus* spp. is able to cause ropiness in breads and bakery products *Clostridium* spp. and *Bacillus* spp. cause softening in vegetables and fruit *Erwinia* and *Penicillium* spp. cause soft rots in vegetables, leading to a mushy texture. The texture of cheese and yoghurts are altered by *Candida* spp. and *Kluyveromyces* spp.	Blackburn (2006) Nychas and Panagou (2011) Fleet (2011)
Discolouration	*Pseudomonas fluorescens* is able to cause blue colouration in cheese *Carnobacterium viridans* causes green discolouration in cooked cured sausage	Nogarol, Acutis et al. (2013) Peirson, Guan et al. (2003)
Gas formation	*Clostridium* spp. causes gas formation resulting in bloating in canned or vacuum-packed goods and late blowing defects in cheese Enterobacteriaceae was responsible for gas production in salad products *Saccharomyces* causes gassiness in wines Several yeast species cause swelling in juice packets	Petruzzi, Corbo et al. (2017) Sahu and Bala (2017)

However, microbes can also improve food quality by changing its intrinsic characteristics. This is evident in fermented foods, where the activity of microbes can improve their organoleptic and nutritive qualities in addition to extending its shelf life. Fermented food microbes can either be introduced spontaneously (from the raw materials, production or processing environments) or inoculated as starter cultures and, over time, can produce enzymes, volatile compounds as well as antimicrobial molecules, such as organic acids, fatty acids, hydrogen peroxide, diacetyl and bacteriocins, which can help to slow down or prevent the growth of spoilage and pathogenic microbes (Reis, Paula et al. 2012). While the spontaneous introduction of microbes is of specific relevance to this review, here we briefly provide an overview of some of the most important microorganisms in fermented foods in general.

LAB are among the most important microbes in the production of several fermented foods (Hatti-Kaul, Chen et al. 2018). This is reflected in the natural adaptation of many LAB to fermented food environments but has been complemented by many years of research to better understand and enhance their contribution to product safety as well as organoleptic, nutritional and health properties (Leroy and De Vuyst 2004). Different species of LAB have been used in dairy products (cheese and fermented milks), meats (sausage), fish, vegetables (sauerkraut and pickles), soy sauce, cereals (sourdough) and alcoholic beverages (wine) (Leroy and De Vuyst 2004). Another group of bacteria associated with fermentation are acetic acid bacteria, which mainly consist of *Acetobacter* and *Gluconacetobacter*. This group of bacteria play important roles in coffee, cocoa and

vinegar fermentation due to their ability to oxidize carbon substrates (Schwan and Ramos 2014). *Bacillus subtilis* and *B. licheniformis* are important for the industrial scale fermentation of soybeans as they grow rapidly, resulting in short fermentation times (Schallmey, Singh et al. 2004). Yeast can also play an important role in the production of many fermented foods. *Saccharomyces cerevisiae* is used in alcoholic fermentation, and yeasts are ultimately used in many indigenous fermented foods as they are acid tolerant, able to grow at high temperatures, and are present in many environments (Schwan and Ramos 2014). In Asia, indigenous foods fermented with yeast, such as miso, soy sauce and wines, are commonly consumed (Aidoo, Rob Nout et al. 2006).

1.3.2 Safety: pathogens and microbial antagonism

Despite the value of fermented food and other microbes in contributing to food quality and safety, with respect to food safety, microbes are frequently regarded negatively, with foodborne pathogens responsible for foodborne illness and outbreaks across the globe annually. The consumption of contaminated food causes an estimated 4,500 deaths annually in Europe (World Health Organisation 2017). The causative agents of foodborne outbreaks in Europe in 2019 were bacterial pathogens (26.4%), bacterial toxins (19.3%), viruses (10.7%), and parasites and other agents (3.6%), while 40% of reported outbreaks had unknown causative agents (European Food Safety Authority and European Centre for Disease Prevention and Control 2019). Common pathogenic bacteria include *Bacillus cereus*, *Campylobacter jejuni*, *Clostridium botulinum*, *C. perfringens*, *Cronobacter sakazakii*, *Escherichia coli*, *Listeria monocytogenes*, *Salmonella* spp., *Shigella* spp.,

Staphylococcus aureus, *Vibrio* spp. and *Yersinia enterocolitica*. Viruses such as norovirus and hepatitis E as well as parasites, including *Toxoplasma gondii* and *Trichinella spiralis*, are also common causes of outbreaks, and have been recently reviewed (Bintsis 2017). There are various ways that pathogenic microorganisms can enter the food chain. They can be inherent to the raw ingredients, or introduced along the processing line via equipment, food handlers or packaging materials, among other routes. Microbial communities can also be present in the form of biofilms, which are microbial communities that adhere to solid surfaces and may contain pathogenic and spoilage species that can persist on surfaces in food processing facilities (Coughlan, Cotter et al. 2016). Once attached, these biofilms can be difficult to remove as they are embedded in a polymeric matrix and cells in the biofilm may be resistant to disinfectants or antimicrobials, particularly in mixed species biofilms (Yuan, Hansen et al. 2020). Research efforts on control strategies to prevent biofilm formation and remove existing biofilms are ongoing to overcome this challenge in the food industry. Food safety management systems including hazard analysis critical control point (HACCP) and risk assessment principles have been widely implemented to prevent foodborne illnesses and outbreaks, and control the spread of pathogens along the food chain. However, these management systems are reliant on having a thorough understanding of the microorganisms present and the risk they may pose.

Although microbes are often viewed negatively from a food safety perspective, some have been useful in biocontrol or biopreservation. Microbial antagonism has been applied in the food industry through the use of bacteriocins, phages and more

(Jordan, Dalmasso et al. 2014). Bacteriocins, which consist of antibacterial peptides, have been used to target spoilage and pathogenic bacteria in food and in turn prolong the shelf life and improve the safety of food (Galvez, López et al. 2008). Bacteriocins from LAB such as nisin, has been approved for use in foods and is the most commonly used in foods such as meat, dairy and vegetable products (Jordan, Dalmasso et al. 2014). Bacteriophages or phages are virus predators of bacteria that have shown great promise as they are naturally occurring and control for specific pathogenic bacteria without impacting the quality and microbiota of foods (O'Sullivan, Bolton et al. 2019). Phages have been applied to a range of foods at various stages from farm to fork to eliminate common pathogenic bacteria such as *Campylobacter jejuni*, *Salmonella* spp., *E. coli* O157:H7 and more (Vikram, Woolston et al. 2020). Additionally, biofilms from some species can aid in improving food safety by outcompeting undesirable bacteria. Some LAB strains were found to exhibit antagonistic properties against unwanted bacteria, to act as a natural barrier and to alter biofilm formation of spoilage microbes (Ouali, Al Kassaa et al. 2014). In food, the microbial community and the interactions between microbes play essential roles in food quality and safety.

1.4 Traditional approaches to characterisation of microbes

As noted above, culture-based assays have historically been used for the detection, enumeration and isolation of viable foodborne pathogens or spoilage microbes in food and environmental samples (Dwivedi and Jaykus 2011). In general, samples are first homogenized, then often undergo enrichment steps (pre-enrichment and selective enrichment), followed by selective or differential plating to distinguish

from other microbes present and, finally, confirmation with biochemical, serological, or other methods. Pre-enrichment is used to recover injured cells and dilute inhibitory compounds in food samples, while selective enrichment increases the concentration of a target pathogen while suppressing the growth of other microflora (Dwivedi and Jaykus 2011). These conventional methods are inexpensive but time consuming and labour-intensive and can take from 2-3 days to a week for inoculation, isolation and confirmation depending on the targeted microorganism (Mandal, Biswas et al. 2011). Furthermore, due to the possible presence of viable but non-culturable (VBNC) bacteria, false negatives may occur, which could mean the unsuccessful detection of pathogens or spoilage bacteria in food.

With the need for more timely detection of bacteria for foodborne outbreak response, rapid methods that replace conventional plating steps with faster immunology or molecular-based approaches have been investigated in depth and adopted more widely in recent years (Wang and Salazar 2016). High levels of sensitivity and specificity are needed for food pathogen detection (Feng 2007). Immunology-based methods like enzyme-linked immunosorbent assay (ELISA), based on the specific binding of antigens with antibodies, have shown potential but their lack of sensitivity and relatively high limit of detection (10^3 – 10^5 colony forming units (CFU) per mL) has meant that enrichment is generally first required before detection. Nevertheless, when ELISA is coupled with nanotechnology, its application for food analysis has achieved greater sensitivity, specificity and stability (Wu, Li et al. 2019).

Nucleic acid-based techniques such as polymerase chain reaction (PCR), quantitative PCR (qPCR) and loop-mediated isothermal amplification (LAMP), compared to traditional culture methods, require shorter time, and through multiplexing, can facilitate concurrent detection, real time monitoring and quantification of multiple microbial targets (Liu, Zou et al. 2017, Tao, Liu et al. 2020). Development and optimisation of each assay is important as complex food matrices may hinder nucleic acid extraction and contain inhibitors that may interfere with reactions in the assay. Particularly when multiplexing, primer design is crucial as primer sets require similar annealing temperatures for a successful assay (Wang and Salazar 2016). Unfortunately, these methods, like immunology-based assays, often still require enrichment or concentration steps because of their limit of detection (10^3 – 10^4 CFU per mL) and as pathogens are often in low concentrations in foods, direct detection is difficult (Ceuppens, Li et al. 2014, Wang and Salazar 2016). Another rapid method used largely in clinical settings is matrix assisted laser desorption ionization-time of flight mass spectrometry (MALDI-TOF MS), which analyses signals from ribosomal proteins after ionization and mass spectra that are distinctive for each strain, allowing for rapid microbial identification (de Koster and Brul 2016). Although these rapid methods are generally preferable to culture-based approaches, thorough validation is required by industry and regulatory bodies before routine adoption.

While culture-based and rapid methods are useful in identifying microbes in complex food or food-environment related samples for public health and commercial purposes, they create an unbalanced emphasis on specific

microorganisms and, despite multiplexing, still only capture a small percentage of the microbial community as a whole (Fleet 1999). As microorganisms exist in communities, it is important to study them as such, because the growth, survival, and activity of one species or strain may impact or be associated with the presence of another. Furthermore, from a practical perspective, approaches that could theoretically allow the simultaneous identification of all pathogens and spoilage microbes in a sample from the food chain could have a disruptive positive influence.

1.5 Meta-omic approaches (community approaches)

Advances in technologies have provided the opportunity for faster and superior characterisation of food chain microbiomes, with a shift towards replacing or supplementing culture-dependent methods with culture-independent, molecular-based methods (Sohier, Pavan et al. 2014). Whole genome sequencing has successfully complemented culture-dependent methods by providing deeper discrimination of microbial strains than previous typing methods, with regulatory bodies in the US and Europe now including it as a tool for pathogen typing and antimicrobial resistance surveillance (Rantsiou, Kathariou et al. 2018). In *E. coli* O157:H7 outbreak investigations, genome sequencing stood out from other typing methods, providing insights that enabled improved epidemiological case and cluster identification, geographical origin tracking and information of potential emerging strains (Jenkins, Dallman et al. 2019).

In contrast, culture-independent methods including the use of DNA sequencing technologies have enabled identification and characterisation of multiple microbes in foods or along the food chain at the same time, while also bypassing the need to culture microbes (Cocolin and Ercolini 2015). Some current community-based approaches are shown in Figure 1.1.

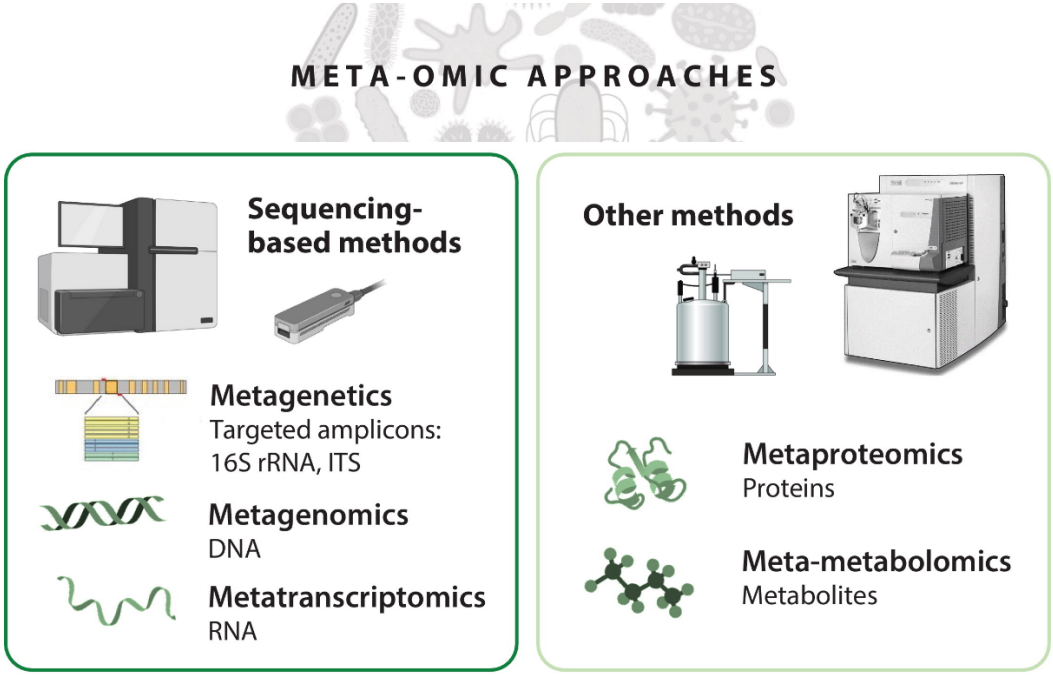


Figure 1.1. Current meta-omic approaches used in microbiome research. Sequencing-based approaches include Metagenetics, Metagenomics and Metatranscriptomics and other community-based methods include Metaproteomics and Meta-metabolomics, which are currently being used in human, environment and food microbiome studies.

1.5.1 Sequencing-based meta-omic approaches: metagenetics, metagenomics and metatranscriptomics

Metagenetics, also known as amplicon sequencing, metataxonomics, metabarcoding and sometimes as 16S metagenomics or 16S rRNA gene sequencing, is a targeted approach which involves the amplification of marker genes from mixed genomic DNA by PCR, followed by direct sequencing and alignment against a reference database to identify the taxonomic composition of whole microbial communities (Franzosa, Hsu et al. 2015). The 16S ribosomal RNA (rRNA) gene is most frequently used in the identification of bacteria as it is universally found in bacteria, and the gene contains nine hypervariable regions, some or all of which can be targeted through amplification and sequencing to identify the corresponding bacterial taxonomy. A similar approach can be applied to fungi, through targeting the 18S or 23S rRNA genes or the internal transcribed spacer (ITS) regions of the rRNA operon.

Shotgun metagenomics, commonly referred to as metagenomics, is the untargeted genomic analysis of a population of microorganisms by sequencing the entire DNA sample extracted from a mixed microbial community (Quince, Walker et al. 2017). This method involves fragmentation of the sample DNA, followed by preparation of a library that is sequenced, with the resulting data analysed to provide information on both taxonomic composition and functional potential of the entire microbial community. Due to the untargeted nature of metagenomics, information relating to all categories of microbes, including bacteria, viruses, archaea and single-celled eukaryotes like fungi, can be derived from the sample (Quince, Walker et al. 2017),

assuming the DNA extraction method is appropriate. The lack of an amplification step removes the bias that metagenetics may have and has greater sensitivity, enabling taxonomic classification up to the strain level. Another advantage of shotgun metagenomics is the potential for the recovery, if sufficient sequencing depth is applied, of metagenome assembled genomes (MAGs), which can provide more genomic information, revealing functional and safety-related properties of specific taxa (Bowers, Kyrpides et al. 2017) and allow for the investigation of strain-level diversity in food-related microbial species such as LAB (Pasolli, De Filippis et al. 2020).

Metatranscriptomics relates to the untargeted sequencing of total mRNA isolated from a sample, which allows for the identification of transcriptionally active microbes in the sample and may provide further insights into the potential functional characteristics of the microbial community. This approach reveals the microbes that are viable and, indeed, most active within a community while also enabling a deeper understanding of how microbial communities in complex food microbiomes or food-related environments interact with each other. This approach can be also used to look at in situ gene expression in food, collecting information on the metabolic activities potentially related to food fermentation and/or spoilage that are currently expressed in a food ecosystem. An additional advantage of metatranscriptomics is the ability to detect RNA-based viruses, including foodborne pathogens such as norovirus (Lewis, Hudson et al. 2020).

1.5.2 Other meta-omic approaches: metaproteomics, meta-metabolomics

Other non-sequencing community-based methods, i.e., metaproteomics and meta-metabolomics, also have the potential to be used in food microbiome studies. Metaproteomics is the large-scale study of the entire protein complement produced by microbial communities within a sample at a given time point, which can aid in linking genomic and transcriptomic data to biological function, deepening the understanding of phenotypic changes as conditions change (Soggiu, Piras et al. 2016). Metaproteomics provides information on the microbial communities and their abundances and function, the interaction within the community, the changes in metabolism and physiology of the community and the substrate utilisation, carbon sources and assimilation pathways of the microbes in the sample (Kleiner 2019). Mass spectrometry is used for metaproteomics and its application in the food microbiome has mainly been in the characterisation of fermented foods such as fermented soybean and cheese (Soggiu, Piras et al. 2016, Xie, Wu et al. 2019). Meta-metabolomics, on the other hand, involves the use of chemistry, biochemistry, and bioinformatics to detect and analyse small weight metabolites in samples and provide insights on microbial phenotypic characteristics. There are two categories of meta-metabolomic analyses, untargeted, which focuses on the detection of as many groups of metabolites as possible, and targeted, which focuses on a specific user-selected group of metabolites under determined conditions (Li, Mann et al. 2020). Researchers in this field typically use either mass spectrometry or nuclear magnetic resonance to evaluate food ingredients, quality, safety, authenticity, and traceability (Kim, Kim et al. 2016). The integration of metaproteomics and meta-metabolomics with other omic approaches has been

used to provide novel insights and link genomic information with phenotypes (Kim, Kim et al. 2016, Pinu, Beale et al. 2019).

1.5.3 Current challenges in sequencing-based meta-omic approaches

Despite their promise, these sequencing-based approaches have challenges to overcome in order to achieve wider application. The major challenge for these meta-omic approaches is the lack of standardisation, causing variation in results due to the use of different extraction methods, sequencing platforms, databases, and bioinformatics tools. To highlight this point, we refer to several studies that have found differing conclusions depending on the approaches taken for analysis, thereby highlighting the need to identify those that provide the greatest accuracy and, ultimately, their use in a standardised manner (Walsh, Crispie et al. 2018, Lewis, Hudson et al. 2020, Yang, Cousineau et al. 2020, McHugh, Yap et al. 2021).

Similarly in terms of analysis, the fact that results are typically presented in terms of relative abundances may lead to misinterpretations, as an increase in the relative abundance of one taxon results in the concurrent decrease of others. For this reason, it is necessary to quantify microbial communities by using complementary methods and efforts to do so have included digital PCR, qPCR, flow cytometry and culture. It is notable that the use of synthetic standards in sequencing has produced varying results for different methods, again highlighting a need for caution during analysis (Galazzo, van Best et al. 2020).

Furthermore, for metagenetics, there is a difficulty when endeavouring to compare outcomes across different studies arising due to a lack of consistency relating to the hypervariable region of the 16S rRNA gene targeted (Claesson, Wang et al. 2010). Additionally, the focus on one marker gene can cause other issues. In particular, the operon copy number for 16S rRNA genes differs across taxa, which may inadvertently affect quantitative estimation. Single-copy target genes like *recA*, *rpoB* and *gyrB* have been suggested as alternatives, but their use is limited due to their relevance to specific taxa of bacteria only and/or the absence of databases that are sufficiently populated (Poirier, Rue et al. 2018, Ogier, Pages et al. 2019)

From the perspective of methodology, extracting DNA/RNA of sufficient concentration and quality is essential for sequencing, which may be challenging in some circumstances, such as from environmental swab samples taken from areas with a low microbial load (De Filippis, Valentino et al. 2020). Amplification methods such as multiple displacement amplification (MDA) have been applied to generate DNA of sufficient quantities, but can lead to biases (Marine, McCarren et al. 2014, McHugh, Yap et al. 2021).

For both standard metagenetic and metagenomic sequencing, there is no differentiation between DNA extracted from living and dead organisms within a microbiome, which is of key importance with respect to food or food environment microbiome studies. Propidium monoazide (PMA) and the previously more commonly used ethidium monoazide (EMA) are DNA-binding dyes that selectively bind to accessible DNA present in the matrix, essentially binding to DNA from dead

bacteria and other cells and preventing its amplification during library preparation (Nocker, Cheung et al. 2006). Treatment with PMA before DNA extraction thereby selects for the subsequent sequencing of DNA from viable cells. The successful application of PMA with sequencing allowed the selective analysis of viable cells during milk processing and cheese manufacturing (Erkus, de Jager et al. 2016, Kable, Srisengfa et al. 2019). However, its performance can be influenced by the microbial community and sample biomass (Wang, Yan et al. 2021). Another option is the sequencing of RNA in place of or alongside DNA, as measuring RNA copies targets the active microbial fraction, which would allow the differentiation of viable and non-viable microbes (Mira Miralles, Maestre-Carballa et al. 2019), though the instability of mRNA can again present challenges.

It is also important to note that for metagenetics in particular, the short reads generated by some sequencing platforms, such as those developed by the market leader, Illumina, can be limiting with respect to assigning taxonomy at the species level. Other sequencing platforms that produce longer reads, such as those from Oxford Nanopore Technologies (ONT) or Pacific Biosciences, may address this but lower read accuracy and higher sequencing costs remain problematic.

Although shotgun metagenomic sequencing overcomes many of the biases associated within amplicon-based approaches, one of its biggest challenges is the reduced microbial sequencing depth that occurs when randomly sequencing samples that contain high amounts of host DNA. While most studies remove the reads from host DNA during bioinformatic analysis, a more efficient alternative is to

deplete host DNA or enrich microbial DNA through various chemical methods and commercially available kits (Marotz, Sanders et al. 2018, Yap, Feehily et al. 2020). Similarly, in metatranscriptomics, highly abundant rRNA can result in increasing costs and complex downstream analysis. To overcome this challenge, rRNA depletion or mRNA enrichment strategies before sequencing and/or post sequencing removal during downstream analysis have been adopted (Shakya, Lo et al. 2019).

Additionally, with the use of sequencing technologies, advanced computational power and bioinformatics skills are necessary for their use, which add to the challenges when considering the application of these approaches.

1.6 Applications of sequencing-based meta-omic approaches along the food chain

1.6.1 Public health applications

The sequencing-based meta-omic approaches mentioned above have contributed significantly to the study of various diverse microbiomes, including by facilitating significant advances in food microbiome research. From a public health perspective, these meta-omic approaches have provided insights relating to pathogen detection, outbreak investigation, determining antimicrobial resistance and food authenticity and source tracking.

Pathogen detection and outbreak investigation. Meta-omic approaches are advantageous as they bypass the need for culturing and enrichment of pathogens from samples before identification and characterisation of putative etiological agents. They also are able to reveal the presence of uncultured and hard to culture microbes, which may be useful in surveillance, source attribution, risk assessment and epidemiological analysis, when traditional methods fall short (EFSA Panel on Biological Hazards, Koutsoumanis et al. 2019). Both metagenetic and metagenomic approaches have enabled the detection and characterisation of pathogens in various foods, including vegetables, meat and dairy products (Aw, Wengert et al. 2016, Yang, Noyes et al. 2016, McHugh, Feehily et al. 2018, Mira Miralles, Maestre-Carballa et al. 2019). Metatranscriptomics, though less widely applied due to the challenges in RNA isolation, also has great potential for identifying viable pathogens in food (Yang, Cousineau et al. 2020).

Use of metagenomic approaches can extend beyond the food chain, where metagenomic sequencing of patient stool samples collected during the outbreak in Germany of STEC (Shiga toxin-producing *E. coli*) *O104*:H4 assisted the recovery of genomes of the outbreak strain (Loman, Constantinidou et al. 2013). Moreover, metagenomics is useful when a viral agent is the cause of the outbreak or, in the case of multi-strain outbreaks, it is able to discriminate and characterise several strains, allowing them to be distinguished considerably faster than traditional culture-based methods (Buytaers, Saltykova et al. 2020). Compared to metagenetics, which may be more useful for low-biomass samples due to the amplification of the target, metagenomics facilitates more sensitive

characterisation to species level and further investigation of the functional potential of microbes present (Grützke, Malorny et al. 2019).

Despite the potential of these meta-omic approaches, they are currently not widely used. One reason is due to the lack of harmonised methods and standardised, accredited workflows/pipelines that would allow for consistent detection and characterisation of outbreak-causing agents (EFSA Panel on Biological Hazards 2019). However, the utility of metagenomic analyses can be enhanced when they are complemented with further quantitative molecular assays, highlighting their effectiveness in determining pathogen contamination or outbreak events. A big technical challenge that hinders greater adoption of meta-omic techniques as a routine screening tool for pathogens is that these techniques are not always sufficiently sensitive (Leonard, Mammel et al. 2015, Lewis, Hudson et al. 2020). With low numbers of pathogenic cells in samples, substantial sequencing depth would be required, particularly for shotgun sequencing, as samples would contain other DNA (Yang, Noyes et al. 2016). With sufficient sequencing depth, shotgun metagenomics can be a faster and more valuable tool that provides more information than current conventional workflows, which permit linking food/environment outbreak-related samples with clinical samples (Grützke, Malorny et al. 2019, Buytaers, Saltykova et al. 2020, Li, Mann et al. 2020). Although the complexity of various food matrices can be a challenge, this is not as great an issue for less biologically complex matrices, such as water used in food production or some minimally processed foods (Fernandez-Cassi, Timoneda et al. 2017).

Identification of antimicrobial resistance-encoding genes. Over the past decades, antimicrobial resistance (AMR) has been identified as a serious public health threat and because of this, more tools have been published for the detection of genetic determinants of AMR from sequencing data. Although whole-genome sequencing of cultured isolates is usually utilised, metagenomic sequencing shows great potential for monitoring AMR, out-performing culture-based methods in quantifying resistance in swine herds (Munk, Andersen et al. 2017). Shotgun sequencing has shown success in monitoring of AMR genes in the environment from farm to slaughter (Noyes, Yang et al. 2016, Pitta, Dou et al. 2016). It has also been used to understand the association between antimicrobial use and resistance and the effect of processing on the resistome and virulome (Campos Calero, Caballero Gómez et al. 2018, Van Gompel, Luiken et al. 2019, Mencía-Ares, Cabrera-Rubio et al. 2020).

As with other metagenomic approaches, sequencing depth and the presence of host DNA should be considered as they have been found to affect resistome profiling in environmental and food samples (Gweon, Shaw et al. 2019, Rubiola, Chiesa et al. 2020). Other challenges include the difficulty in assigning ARGs (antibiotic resistance genes) to their host species or strains, which may be addressed by sequencing with long-read technology, and the choice of reference resistance gene database, where differences were found between gene variants from the same reference sequence from different databases, reiterating the need for comprehensive databases and standardised workflows (Doyle, O'Sullivan et al. 2020, Slizovskiy, Mukherjee et al. 2020). It is also important to note that the AMR

data obtained may not always be phenotypically relevant, as these genes might not be expressed or the choice of bioinformatic tools can result in false positives or negatives (Doyle, O'Sullivan et al. 2020). From the perspective of gene expression, metatranscriptomics can potentially be employed to complement the analysis (Wang, Hu et al. 2020). The analysis of the mobilome (all mobile genetic elements of the microbiome) has also been paired with resistome analysis to understand the potential spread of AMR genes and virulence factors through horizontal gene transfer (Slizovskiy, Mukherjee et al. 2020).

Food authenticity. Food fraud is a global issue that has many consequences including possible health risks, economic losses, and hindering sustainability efforts. Metabarcoding has been used to determine the authenticity and origin of honey, traditional Chinese medicines, fish and more (Coghlan, Haile et al. 2012, Carvalho, Palhares et al. 2017, Khansaritoreh, Salmaki et al. 2020, Liu, Teixeira et al. 2020). The basic concept is that the microbiome associated with a traditional food is closely linked to the geographical origin and mode of production of the food as the microbes are typical of raw materials and environment. Although there have been some successes, there are challenges associated with using microbiomes as a means of determining the provenance of food. These include the need for the existence of databases containing the components of the expected microbiome of the food and the potential alteration of the microbiome due to storage or processing conditions (Liu, Teixeira et al. 2020). Similar to other meta-omic applications, the reliance on the completeness of reference databases together with the accuracy of food matrix authentication are important to avoid inaccurate

conclusions. Haiminen, Edlund et al. (2019) found both DNA and RNA shotgun sequencing to be accurate untargeted methods for food authentication and contaminant detection, which has been applied by Kamilari, Tomazou et al. (2019) to characterise Protected Designation of Origin (PDO) cheeses with complementary metabolomics to define product origin differentiating factors.

Other public health-related fields. Besides the food industry, other fields have also found benefits in the application of community-based microbiome analysis methods. Community-based approaches have contributed to the increasing knowledge of the indigenous microbial community and antimicrobial resistance patterns in both healthcare settings and water systems that have provided evidence for the greater need for surveillance (King, Pham et al. 2016, O'Hara, Reed et al. 2017, Zhang, Kitajima et al. 2017). In hospital settings, meta-omic approaches have provided clues to the routes of entry and relationships between pathogens and non-pathogens, as well as helped in environmental surveillance to fight hospital-acquired infections and AMR (Comar, D'Accolti et al. 2019, Rampelotto, Sereia et al. 2019). Similarly, shotgun metagenomics when supplemented with other techniques was effective in uncovering the presence of virulence factors and novel biomarkers of pathogen-related species in drinking water distribution systems (Zhang, Kitajima et al. 2017). Additionally, on an international scale, urban sewage and waste from aircraft flights have been cited as economically and ethically acceptable datasets for continuous global surveillance and prediction of AMR using metagenomics (Petersen, Rasmussen et al. 2015, Hendriksen, Munk et al. 2019).

1.6.2 Food industry applications

Microbial communities exist throughout the food chain and understanding their dynamics and the conditions that promote or hinder their growth would be useful for food safety and quality purposes. Research efforts using meta-omic approaches have looked into foods, food-associated environments, and food processing steps, as presented in Figure 1.2, which will be elaborated in the following sections.

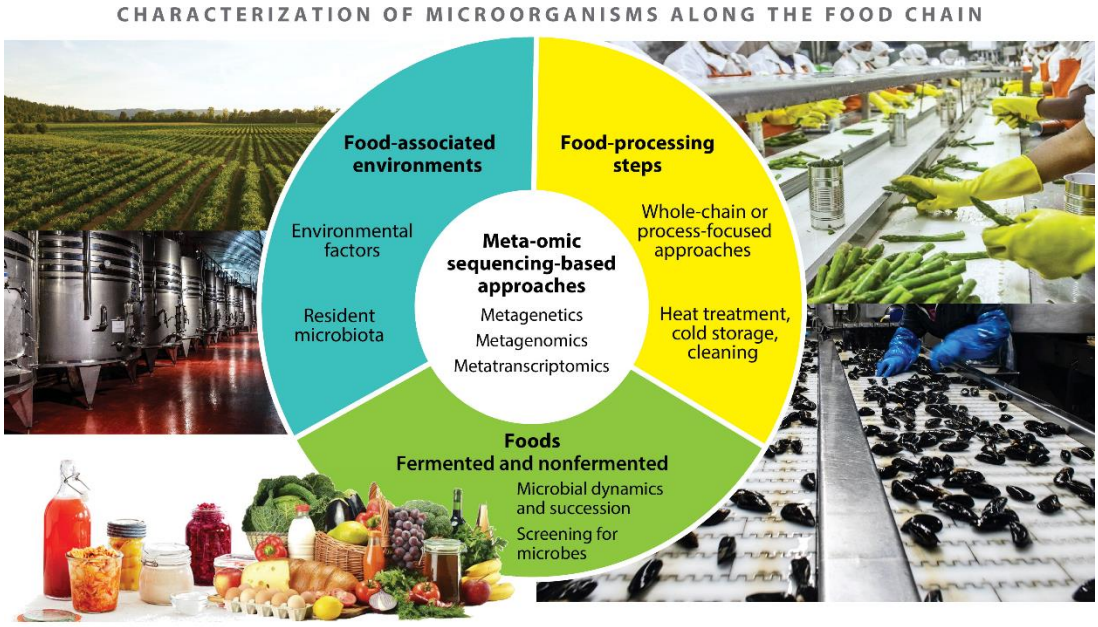


Figure 1.2. Current applications of meta-omic sequencing-based approaches along the food chain. Metagenetics, metagenomics, and metatranscriptomics have been used in studies investigating the microbial community of food, food-processing steps and food-associated environments.

1.6.2.1 Foods (fermented and non-fermented)

One of the main applications of community-based approaches is in the study of fermented foods. Previous reviews noted that most of the early studies on fermented foods employed metagenetics to monitor the activity of microorganisms

during fermentation (De Filippis, Parente et al. 2017). In recent years, more studies have utilised metagenomics and metatranscriptomics to understand the changes in microbial community diversity and activity during fermentation in a broad range of foods, including vegetables, cheese and more (Jung, Lee et al. 2013, De Filippis, Genovese et al. 2016, Duru, Laine et al. 2018, Pham, Landaud et al. 2019, Kim, Chun et al. 2020, Liu, Liu et al. 2020, Xiao, Huang et al. 2020). Metatranscriptomic analysis revealed the changes in gene expression and metabolic properties of LAB during fermentation of vegetables (Jung, Lee et al. 2013, Xiao, Huang et al. 2020). Likewise, from metatranscriptomic analysis of cheese, metabolic interactions within the microbial community and temperature-driven functional changes during ripening were revealed (De Filippis, Genovese et al. 2016, Pham, Landaud et al. 2019). The use of both metagenomic and metatranscriptomic analyses allowed for the detection of active microbes during fermentation and of microbes responsible for biogenic amine production in fermented soy products (Kim, Chun et al. 2020, Liu, Liu et al. 2020). This parallel approach was also useful in understanding the dynamics of the microbial community during ripening, revealing the impact of temperature on the microbial community and genes expressed (Dugat-Bony, Straub et al. 2015, Duru, Laine et al. 2018). These are selected examples of studies within the continuously growing pool of research that employ these methods to study the microbial consortia in fermented foods. Unsurprisingly, it has been suggested that multiple meta-omic approaches facilitate the improved, efficient and sustainable production of fermented foods through detailed functional characterisation of their microbiomes (Chen, Chen et al. 2017).

Though the number of studies using meta-omic approaches to study non-fermented foods are considerably lower, those that have been completed highlight the great potential of such approaches. Most of these applications have related to the characterisation of food-associated environments or processing steps, which will be elaborated in the following sections. Other than those studies, there have been promising studies involving the use of community-based methods to screen for spoilage or pathogenic microorganisms. However, due to the complex nature of food samples and the frequently low pathogen abundances, direct sequencing of DNA or RNA of food has, to date, been found to be less sensitive than conventional culture-based or amplicon-based methods (Lewis, Hudson et al. 2020, Yang, Cousineau et al. 2020). It should also be noted that, even though both short and long read sequencing technologies have shown promise with respect to accurate classification of microbes to the family and genus levels, not all approaches sufficiently classify to the species or strain level needed for pathogen detection (Grützke, Malorny et al. 2019). This is sometimes a significant limitation, especially in terms of food safety, where identifying only at genus level may not be informative enough to understand the actual species present that could cause food safety or quality issues along the food chain. The need for sensitive and specific tests coupled with other challenges prove that these community-based approaches are currently not applicable at regulatory compliance level, but with further development are likely to become standard in the future (Yang, Noyes et al. 2016).

1.6.2.2 Food-associated environments

Food-associated environments, from farm to processing facility, have repeatedly been found to impact, both positively and negatively, on the final product microbiome.

Environmental factors. Microorganisms can enter the food chain at a number of different points. This includes the crops and animals from which the foods are sourced/derived, as well as environmental factors such as soil, water, farming systems, pests, and climate conditions. Meta-omic approaches have found that factors such as pasture systems, animal housing, airborne dust, irrigation water and several others can influence the microbiota diversity and composition of food (Doyle, Gleeson et al. 2017, Allard, Callahan et al. 2019, Wu, Nguyen et al. 2019). Besides diversity and composition, meta-omic approaches used to characterise the resistome reveal that farm environments are potential vehicles for AMR bacteria and genes, originating from dust and animal faeces that contribute to AMR spread and workers' exposure (Noyes, Yang et al. 2016, Luiken, Van Gompel et al. 2020). The use of animal waste as fertiliser (manure/wastewater) can also cause the dissemination of AMR bacteria and genes in the environment, which in turn affect the microbiota of crops grown or animals raised on the land (Allard, Callahan et al. 2019, He, He et al. 2019). Seasonality is another contributing factor to the microbiota of the animal and plant environment. Seasonal impacts were evident in certain products, like milk and beef, where the use of metagenetics and metagenomics has revealed seasonal variations in the microbiota of final products (Kable, Srisengfa et al. 2016, Hwang, Choi et al. 2020, McHugh, Feehily et al. 2020).

Food processing environments. Meta-omic techniques have been adopted in the characterisation of several environments involved in the processing of foods such as meat (De Filippis, La Storia et al. 2013, Hultman, Rahkila et al. 2015, Stellato, La Storia et al. 2016), dairy (Anvarian, Cao et al. 2016, Kable, Srisengfa et al. 2016, Doyle, Gleeson et al. 2017) and alcoholic beverages (Bokulich, Ohta et al. 2013, Bokulich, Bergsveinson et al. 2015, Wang, Du et al. 2018). One key observation from using meta-omic approaches for such studies is the presence of a resident microbiome that persists within the processing environments and has the potential to affect final food product quality and safety. This was highlighted in a recent review, relating to the use of high-throughput sequencing to characterise dominant taxa found in both processing environments and food products, which summarised the evidence that the processing environment can act as a reservoir and source of microbial transfer to food (De Filippis, Valentino et al. 2020). This can be both beneficial and detrimental with, for example, beneficial effects apparent in fermented food production. In this regard, microbes in the environment were found to contribute positively to the production of fermented vegetables, wine and Chinese liquor (Bokulich, Ohta et al. 2013, Einson, Rani et al. 2018, Wang, Du et al. 2018). In contrast, spoilage or pathogenic microorganisms have been found on surfaces of various dairy, meat and vegetable processing facilities using different meta-omic approaches (Hultman, Rahkila et al. 2015, Pothakos, Stellato et al. 2015, Stellato, La Storia et al. 2016, McHugh, Feehily et al. 2020, Zwirzitz, Wetzels et al. 2020). For example, *Pseudomonas* spp. was found in drain biofilms of cheese and salmon processing plants (Dzieciol, Schornsteiner et al. 2016, Langsrud, Moen et al. 2016) and pathogens like *Staphylococcus* and *Yersinia* were found on surfaces of

milk and meat processing plants (Hultman, Rahkila et al. 2015, Kable, Srisengfa et al. 2019). Indeed, correlation of microbial communities in biofilms, as determined by metagenetics, with environmental factors has been used to track persistence over time, showing that bacterial communities were location-specific in meat and fish processing plants (Rodríguez-López, Rodríguez-Herrera et al. 2020). Additionally, microbial co-occurrences of pathogens with other microbes and microbial interactions within complex ecosystems can be evaluated through meta-omic approaches, which may determine patterns which favour or prevent the growth or survival of foodborne pathogens (Illeghems, Weckx et al. 2015, den Besten, Amézquita et al. 2018). This was investigated through 16S rRNA sequencing that examined interactions between *Listeria* spp. and the microbiome within a food production facility and identified species that acted as apparent protagonists or antagonists that impacted on the presence of *L. monocytogenes* within the processing plant (Fox, Solomon et al. 2014).

Handling can be a potential source of contamination or microbial transfer, whereby microbes can be unknowingly transferred from surfaces to the food product. *Moraxella* spp., a prominent meat spoilage bacteria, was found on gloves of employees, which were identified as a potential source of contamination using full-length 16S rRNA gene sequencing throughout a pork processing plant (Zwirzitz, Wetzels et al. 2020). Similarly, handling was identified as a catalyst in the proliferation of spoilage bacteria in beef products after high-throughput sequencing uncovered the origin of spoilage-associated bacteria from carcasses and their persistence in the environment (De Filippis, La Storia et al. 2013).

1.6.2.3 Food processing steps

Using meta-omic approaches to monitor the changes in food microbiomes during food processing has been useful in understanding the impact of processes on the quality and safety of foods. This has been studied through two approaches. One approach has involved profiling the entire food processing chain, where samples were taken from the start to the end of the process and meta-omic methods were used to track the changes in microbial community dynamics, which can facilitate the generation of mitigation measures. This whole chain approach often involves sampling of both food and environmental samples and has highlighted areas where contamination or spoilage can potentially occur, e.g., in meat processing, animal carcasses or hides were identified as possible sources of contamination and measures taken during and after slaughter were found to be key in reducing bacterial load and transmission of AMR genes to meat products (De Filippis, La Storia et al. 2013, Noyes, Yang et al. 2016, Yang, Noyes et al. 2016, Calero, Gómez et al. 2020). A similar approach to study sausage production showed that the emulsification step selected for gram-positive spoilage bacteria (Hultman, Rahkila et al. 2015). Other investigations have highlighted the impact of low temperatures, storage and equipment on the milk microbiota in dairy processing (Kable, Srisengfa et al. 2016, Falardeau, Keeney et al. 2019, McHugh, Feehily et al. 2020), while in breweries, food contact surfaces were noted as areas that could allow transmission of spoilage bacteria or genes (Bokulich, Bergsveinson et al. 2015).

The other approach that has been taken when using meta-omic methodologies is process-focused, where specific processing steps that are often considered critical

points in food safety management systems are examined. Processes like heat treatment, cold storage, packaging, cleaning and others have been studied to understand the microbial dynamics during these processes and to ensure their efficacy at eliminating or reducing growth of bacteria. Metagenetics used to investigate heat treatments unsurprisingly found a reduction of bacterial abundance and diversity in meatballs and cheese but also an influence on the quality of the final products (Kamilari, Anagnostopoulos et al. 2020, Li, Wang et al. 2021). Similarly, monitoring the ripening processes of cheese using metagenomics and metatranscriptomics has provided a better understanding of the temperature-driven differences in flavour development (De Filippis, Genovese et al. 2016, Duru, Laine et al. 2018), while metagenetics, proteomics and complementary physicochemical and sensory analysis revealed the efficacy of high pressure processing in improving the quality and shelf life of fish fillets and led to the identification of quality markers for further study (Tsironi, Anjos et al. 2019). For storage in particular, metagenetic and metagenomic analysis revealed cold temperature storage is an area along the processing chain that allowed for the proliferation and dominance of certain psychrotrophic spoilage microorganisms in meat and dairy (Stellato, La Storia et al. 2016, McHugh, Feehily et al. 2020). Monitoring microbial dynamics to understand the effect of storage temperature on the microbial community has been performed using metagenetics coupled with sensory assessment or culture-dependent methods in sausage and fish, which has resulted in the development of models to infer spoilage dynamics and associations of bacterial species during storage (Benson, David et al. 2014, Zotta, Parente et al. 2019). Modified atmosphere packaging (MAP) is currently used to extend the shelf

life of various foods like fresh and processed meat and seafood, fruit and vegetables, but optimisation of the gas composition is required to keep the product's quality. In the evaluation of MAP for poultry, Wang, Zhang et al. (2017) identified a shift in the bacterial community compared to other packaging conditions using metagenomics and Höll, Hilgarth et al. (2020) used metatranscriptomics to monitor the regulation responses of two spoilage bacteria to different atmospheric conditions. Similarly, evaluation of shelf life of fish fillets in MAP and vacuum packaging at low temperatures has been performed with metagenetics and sensory analysis or metabolomics to understand the dynamics of spoilage bacteria over time (Jääskeläinen, Jakobsen et al. 2019, Sørensen, Bøknæs et al. 2020). The efficacy of cleaning and disinfection has been investigated with metagenetics, with evidence of bacterial diversity and abundance altered after cleaning in dairy and pig facilities (Dass, Wang et al. 2018, Bridier, Le Grandois et al. 2019). Similarly, RNA-based 16S rRNA sequencing showed current cleaning practices with ozonation were effective and caused shifts in potentially active microbiota in meat processing plants (Botta, Ferrocino et al. 2020). In contrast, sanitation in salmon processing plants, determined by metagenetics, was found to be inadequate as *Pseudomonas* spp. persisted in biofilms on conveyor belts (Langsrud, Moen et al. 2016).

Ultimately, sequencing-based meta-omic approaches have been found to be an effective tool in identifying microorganisms along the processing chain and routine implementation can help to uncover the factors that influence microbial population dynamics (McHugh, Feehily et al. 2020, Zwirzitz, Wetzels et al. 2020). The numerous

studies carried out to date show that there is great potential for the use of meta-omic approaches in tracking microbial communities along the food chain.

1.6.3 A dairy perspective

From a dairy perspective, meta-omic approaches have been widely applied across the processing chain, from farm to fork. At the farm level, the rumen microbiome has been widely studied due its importance with respect to animal productivity, the environment and the meat or milk end product (Matthews, Crispie et al. 2019). The abundance of specific taxa in the rumen have been found to correlate with feed conversion efficiency (Li and Guan 2017). Many studies have also employed high-throughput meta-omic approaches to investigate the microbial basis for mastitis in animals, as it affects animal welfare and productivity and, thus, results in potential economic losses. Specifically, metagenetics and metagenomics have been used to understand host-pathogen interactions, factors contributing to mastitis occurrence, and the efficacy of treatments (Falentin, Rault et al. 2016, Ganda, Bisinotto et al. 2016, Hoque, Istiaq et al. 2019). Comparison of the bovine milk microbiome from healthy and mastitis-affected animals using shotgun metagenomics has shown clear differences in composition, as well as differences in the abundances of microbial metabolic pathways related to colonisation, proliferation, immune-diseases, oxidative stress, among others (Hoque, Istiaq et al. 2019).

Along the dairy processing chain, 16S rRNA gene amplicon sequencing has been widely used to study bacterial composition. Studies investigating the changes in the raw milk microbiota during common processes such as transfer from tanker trucks

to the processing facility, concentration, separation or blending, showed that the bacterial communities were highly variable (Kable, Srisengfa et al. 2016, Kable, Srisengfa et al. 2019). These findings helped highlight critical control points during processing that in particular would benefit from further monitoring, such as blended silos (Kable, Srisengfa et al. 2019). Metagenetics is also the most common approach used for research to investigate the factors influencing the raw milk microbiome. While short read technologies, in particular Illumina-based approaches, are commonly used (Doyle, Gleeson et al. 2017, Rajawardana, Fernando et al. 2022), long read approaches such as Pacific Biosciences single-molecule real time-based sequencing has also been applied more recently (Guo, Yu et al. 2021, Liang, Wang et al. 2022). A 'whole processing chain' approach has also been used to profile microbial communities of relevance to the production of dairy products such as cheese and dairy powders, with the processing environment frequently being found to contribute to the microbiota of finished products (Falardeau, Keeney et al. 2019, McHugh, Feehily et al. 2020). These investigations are being complemented by metatranscriptomic studies of the cheese microbiota that have provided insights into the active microbiota and genes expressed during ripening/maturation of cheeses that, for example, contribute to flavour development (De Filippis, Genovese et al. 2016, Quijada, Dzieciol et al. 2022). Finally, from a food safety view, meta-omic approaches have been used to reveal the distribution patterns of antimicrobial resistance genes in dairy products and environments. In specific instances, the use of metagenomics has uncovered the potential horizontal transfer of antimicrobial resistance genes between bacterial hosts from different areas of the farm (Pitta, Dou et al. 2016) and also revealed that

retail raw milk was a reservoir of AMR, because it contained a higher prevalence of ARGs than pasteurised milk (Liu, Zhu et al. 2020). In summary, the application of meta-omic approaches to dairy environments and/or products has expanded knowledge on these microbiomes, with the potential for further studies with metagenomics and metatranscriptomics to reveal additional insights.

1.7 Future issues

- With increasing adoption of these meta-omic approaches to uncover the microbiome of food and food-related environments, there is a great need for standardised workflows/pipelines for methodology and analysis.
- Large amounts of data are generated by sequencing. This requires good data management practices and systematic metadata documentation to facilitate data sharing of research outputs. Additionally, bioinformatics expertise for the analysis of the data generated is currently essential to draw accurate and correct interpretations from the sequencing data. Future efforts will need to focus on accurate, automated analytical tools.
- As substantial parts of the analysis require referencing to available databases, the results from sequencing studies are only as good as these databases. Databases are currently compiled mainly from human microbiome studies as more research has been done in that field, which may result in a bias towards human-related microbes. The ongoing increase in microbiome studies on food and other fields should correct this imbalance to enable better characterisation of microbiomes.

- With the further development of assays to overcome the challenges of meta-omic approaches, such as host DNA depletion and the ability to distinguish viable microbes in the microbial community, there will be an even wider application of meta-omic approaches for the characterisation of microbes along the food processing chain.
- From metagenomic data, the recovery of MAGs could make way for more single strain studies that can contribute to greater understanding of strains responsible for fermentation or spoilage in foods or the resident microflora of food environments. Additionally, increasing number of studies into the functional properties of microorganisms within food environments using metatranscriptomics or metagenomics with complementary approaches like metabolomics can provide a greater insight into the active microorganisms and metabolic pathways involved in processes along the food chain.
- Portable sequencing devices from ONT have allowed for field/onsite sequencing that has proven to be useful in clinical outbreak investigations and environmental sampling. These portable devices could enable rapid detection of microbiological contaminants or pathogens in food production or processing environments. Although some studies have explored this possibility, further comparisons with other sequencing technologies and platforms are required to determine accuracy and comparability (Yang, Cousineau et al. 2020, McHugh, Yap et al. 2021).

1.8 Summary points

- Microorganisms are important contributors to the quality and safety of a food product, and they exist throughout the whole food chain.
- As microbes exist in communities, it is valuable to study them as such. Meta-omic approaches bypass the need for culturing and isolating microbes and allow for the greater characterisation of microbial communities.
- Metagenetics and metagenomics are two sequencing-based meta-omic approaches that are already being used in the characterisation of foods, food-associated environments, and food processing microbiomes. Though only a few studies have used metatranscriptomics, results show potential in assessing the dynamics of viable microbes along the food chain.
- The use of sequencing-based meta-omic approaches show promise in better characterisation of microbiomes along the food chain and would allow for greater understanding of the factors contributing to food safety and quality. However, standardised workflows/pipelines are necessary to allow for data sharing and comparability and widespread adoption at a regulatory and industry level.

1.9 Acknowledgements including funding source

MY is funded by the Irish Dairy Levy. Research in the Cotter laboratory is also funded by Science Foundation Ireland (SFI) under grant number SFI/12/RC/2273 (APC Microbiome Ireland) and by SFI together with the Irish Department of Agriculture, Food and the Marine under grant number SFI/16/RC/3835 (VistaMilk) and by the European Commission under the Horizon 2020 program under grant number 818368 (MASTER).

1.10 References

Aidoo, K. E., M. Rob Nout and P. K. Sarkar (2006). "Occurrence and function of yeasts in Asian indigenous fermented foods." FEMS yeast research 6(1): 30-39.

Allard, S. M., M. T. Callahan, A. Bui, A. M. C. Ferelli, J. Chopyk, S. Chattopadhyay, E. F. Mongodin, S. A. Micallef and A. R. Sapkota (2019). "Creek to Table: Tracking fecal indicator bacteria, bacterial pathogens, and total bacterial communities from irrigation water to kale and radish crops." Science of The Total Environment 666: 461-471.

Anvarian, A. H., Y. Cao, S. Srikumar, S. Fanning and K. Jordan (2016). "Flow cytometric and 16S sequencing methodologies for monitoring the physiological status of the microbiome in powdered infant formula production." Frontiers in microbiology 7: 968.

Aw, T. G., S. Wengert and J. B. Rose (2016). "Metagenomic analysis of viruses associated with field-grown and retail lettuce identifies human and animal viruses." International journal of food microbiology 223: 50-56.

Benson, A. K., J. R. David, S. E. Gilbreth, G. Smith, J. Nietfeldt, R. Legge, J. Kim, R. Sinha, C. E. Duncan and J. Ma (2014). "Microbial successions are associated with changes in chemical profiles of a model refrigerated fresh pork sausage during an

80-day shelf life study." Applied and environmental microbiology 80(17): 5178-5194.

Bintsis, T. (2017). "Foodborne pathogens." AIMS microbiology 3(3): 529.

Blackburn, C. W. (2006). Food spoilage microorganisms. Cambridge, UK, Woodhead Publishing.

Bokulich, N. A., J. Bergsveinson, B. Ziola and D. A. Mills (2015). "Mapping microbial ecosystems and spoilage-gene flow in breweries highlights patterns of contamination and resistance." Elife 4: e04634.

Bokulich, N. A., M. Ohta, P. M. Richardson and D. A. Mills (2013). "Monitoring seasonal changes in winery-resident microbiota." PloS one 8(6).

Botta, C., I. Ferrocino, A. Pessione, L. Cocolin and K. Rantsiou (2020). "Spatiotemporal Distribution of the Environmental Microbiota in Food Processing Plants as Impacted by Cleaning and Sanitizing Procedures: The Case of Slaughterhouses and Gaseous Ozone." Applied and Environmental Microbiology 86(23).

Bowers, R. M., N. C. Kyrpides, R. Stepanauskas, M. Harmon-Smith, D. Doud, T. Reddy, F. Schulz, J. Jarett, A. R. Rivers and E. A. Eloe-Fadrosh (2017). "Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea." Nature biotechnology 35(8): 725-731.

Bridier, A., P. Le Grandois, M.-H. Moreau, C. Prénom, A. Le Roux, C. Feurer and C. Soumet (2019). "Impact of cleaning and disinfection procedures on microbial ecology and Salmonella antimicrobial resistance in a pig slaughterhouse." Scientific reports 9(1): 1-13.

Buytaers, F. E., A. Saltykova, S. Denayer, B. Verhaegen, K. Vanneste, N. H. Roosens, D. Piérard, K. Marchal and S. C. De Keersmaecker (2020). "A Practical Method to Implement Strain-Level Metagenomics-Based Foodborne Outbreak Investigation and Source Tracking in Routine." Microorganisms 8(8): 1191.

Calero, G. C., N. C. Gómez, L. L. Lerma, N. Benomar, C. W. Knapp and H. Abriouel (2020). "In silico mapping of microbial communities and stress responses in a porcine slaughterhouse and pork products through its production chain, and the efficacy of HLE disinfectant." Food Research International 136: 109486.

Campos Calero, G., N. Caballero Gómez, N. Benomar, B. Pérez Montoro, C. W. Knapp, A. Gálvez and H. Abriouel (2018). "Deciphering resistome and virulome diversity in a porcine slaughterhouse and pork products through its production chain." Frontiers in microbiology 9: 2099.

Cao, Y., S. Fanning, S. Proos, K. Jordan and S. Srikumar (2017). "A review on the applications of next generation sequencing technologies as applied to food-related microbiome studies." Frontiers in microbiology 8: 1829.

Carvalho, D. C., R. M. Palhares, M. G. Drummond and M. Gadanho (2017). "Food metagenomics: Next generation sequencing identifies species mixtures and mislabeling within highly processed cod products." Food Control 80: 183-186.

Ceuppens, S., D. Li, M. Uyttendaele, P. Renault, P. Ross, M. V. Ranst, L. Cocolin and J. Donaghy (2014). "Molecular methods in food safety microbiology: interpretation and implications of nucleic acid detection." Comprehensive Reviews in Food Science and Food Safety 13(4): 551-577.

Chen, G., C. Chen and Z. Lei (2017). "Meta-omics insights in the microbial community profiling and functional characterization of fermented foods." Trends in Food Science & Technology 65: 23-31.

Claesson, M. J., Q. Wang, O. O'Sullivan, R. Greene-Diniz, J. R. Cole, R. P. Ross and P. W. O'Toole (2010). "Comparison of two next-generation sequencing technologies for resolving highly complex microbiota composition using tandem variable 16S rRNA gene regions." Nucleic acids research 38(22): e200-e200.

Cocolin, L. and D. Ercolini (2015). "Zooming into food-associated microbial consortia: a 'cultural'evolution." Current Opinion in Food Science 2: 43-50.

Coghlan, M. L., J. Haile, J. Houston, D. C. Murray, N. E. White, P. Moolhuijzen, M. I. Bellgard and M. Bunce (2012). "Deep sequencing of plant and animal DNA

contained within traditional Chinese medicines reveals legality issues and health safety concerns." PLoS Genet 8(4): e1002657.

Comar, M., M. D'Accolti, C. Cason, I. Soffritti, G. Campisciano, L. Lanzoni, M. Bisi, A. Volta, S. Mazzacane and E. Caselli (2019). "Introduction of NGS in environmental surveillance for healthcare-associated infection control." Microorganisms 7(12): 708.

Coughlan, L. M., P. D. Cotter, C. Hill and A. Alvarez-Ordóñez (2016). "New weapons to fight old enemies: novel strategies for the (bio) control of bacterial biofilms in the food industry." Frontiers in microbiology 7: 1641.

Dass, S. C., B. Wang, J. E. Stratton, A. Bianchini and A. Ababdappa (2018). "Food processing environment surveillance using amplicon metagenomics: assessing the change in the microbiome of a fluid milk processing facility before and after cleaning." BAOJ Food Sci & Tec 2(1): 12.

De Filippis, F., A. Genovese, P. Ferranti, J. A. Gilbert and D. Ercolini (2016). "Metatranscriptomics reveals temperature-driven functional changes in microbiome impacting cheese maturation rate." Scientific reports 6: 21871.

De Filippis, F., A. La Storia, F. Villani and D. Ercolini (2013). "Exploring the sources of bacterial spoilers in beefsteaks by culture-independent high-throughput sequencing." PLoS One 8(7).

De Filippis, F., E. Parente and D. Ercolini (2017). "Metagenomics insights into food fermentations." Microbial biotechnology 10(1): 91-102.

De Filippis, F., V. Valentino, A. Alvarez-Ordóñez, P. D. Cotter and D. Ercolini (2020). "Environmental microbiome mapping as a strategy to improve quality and safety in the food industry." Current Opinion in Food Science.

de Koster, C. G. and S. Brul (2016). "MALDI-TOF MS identification and tracking of food spoilers and food-borne pathogens." Current Opinion in Food Science 10: 76-84.

den Besten, H. M., A. Amézquita, S. Bover-Cid, S. Dagnas, M. Ellouze, S. Guillou, G. Nychas, C. O'Mahony, F. Pérez-Rodriguez and J.-M. Membré (2018). "Next generation of microbiological risk assessment: potential of omics data for exposure assessment." International journal of food microbiology 287: 18-27.

Doyle, C. J., D. Gleeson, P. W. O'Toole and P. D. Cotter (2017). "Impacts of seasonal housing and teat preparation on raw milk microbiota: a high-throughput sequencing study." Applied and environmental microbiology 83(2).

Doyle, R. M., D. M. O'Sullivan, S. D. Aller, S. Bruchmann, T. Clark, A. C. Pelegrin, M. Cormican, E. D. Benavente, M. J. Ellington and E. McGrath (2020). "Discordant bioinformatic predictions of antimicrobial resistance from whole-genome sequencing data of bacterial isolates: An inter-laboratory study." Microbial genomics 6(2).

Dugat-Bony, E., C. Straub, A. Teissandier, D. Onésime, V. Loux, C. Monnet, F. Irlinger, S. Landaud, M.-N. Leclercq-Perlat and P. Bento (2015). "Overview of a surface-ripened cheese community functioning by meta-omics analyses." PloS one 10(4): e0124360.

Duru, I. C., P. Laine, M. Andreevskaya, L. Paulin, S. Kananen, S. Tynkkynen, P. Auvinen and O.-P. Smolander (2018). "Metagenomic and metatranscriptomic analysis of the microbial community in Swiss-type Maasdam cheese during ripening." International journal of food microbiology 281: 10-22.

Dwivedi, H. P. and L.-A. Jaykus (2011). "Detection of pathogens in foods: the current state-of-the-art and future directions." Critical reviews in microbiology 37(1): 40-63.

Dzieciol, M., E. Schornsteiner, M. Muhterem-Uyar, B. Stessl, M. Wagner and S. Schmitz-Esser (2016). "Bacterial diversity of floor drain biofilms and drain waters in a Listeria monocytogenes contaminated food processing environment." International journal of food microbiology 223: 33-40.

EFSA Panel on Biological Hazards, K. Koutsoumanis, A. Allende, A. Alvarez-Ordóñez, D. Bolton, S. Bover-Cid, M. Chemaly, R. Davies, A. De Cesare and F. Hilbert (2019). "Whole genome sequencing and metagenomics for outbreak investigation, source

attribution and risk assessment of food-borne microorganisms." EFSA Journal 17(12): e05898.

Einson, J. E., A. Rani, X. You, A. A. Rodriguez, C. L. Randell, T. Barnaba, M. K. Mammel, M. L. Kotewicz, C. A. Elkins and D. A. Sela (2018). "A vegetable fermentation facility hosts distinct microbiomes reflecting the production environment." Appl. Environ. Microbiol. 84(22): e01680-01618.

Erkus, O., V. C. de Jager, R. T. Geene, I. van Alen-Boerrigter, L. Hazelwood, S. A. van Hijum, M. Kleerebezem and E. J. Smid (2016). "Use of propidium monoazide for selective profiling of viable microbial cells during Gouda cheese ripening." International journal of food microbiology 228: 1-9.

European Food Safety Authority and European Centre for Disease Prevention and Control (2019). "The European Union one health 2018 zoonoses report." EFSA Journal 17(12): e05926.

Falardeau, J., K. Keeney, A. Trmčić, D. Kitts and S. Wang (2019). "Farm-to-fork profiling of bacterial communities associated with an artisan cheese production facility." Food microbiology 83: 48-58.

Falentin, H., L. Rault, A. Nicolas, D. S. Bouchard, J. Lassalas, P. Lamberton, J.-M. Aubry, P.-G. Marnet, Y. Le Loir and S. Even (2016). "Bovine teat microbiome analysis revealed reduced alpha diversity and significant changes in taxonomic profiles in quarters with a history of mastitis." Frontiers in microbiology 7: 480.

Feng, P. (2007). Rapid methods for the detection of foodborne pathogens: current and next-generation technologies. Food Microbiology: Fundamentals and Frontiers, Third Edition, American Society of Microbiology: 911-934.

Fernandez-Cassi, X., N. Timoneda, E. Gonzales-Gustavson, J. Abril, S. Bofill-Mas and R. Girones (2017). "A metagenomic assessment of viral contamination on fresh parsley plants irrigated with fecally tainted river water." International journal of food microbiology 257: 80-90.

Fleet, G. H. (1999). "Microorganisms in food ecosystems." International Journal of Food Microbiology 50(1-1): 101-117.

Fleet, G. H. (2011). Yeast spoilage of foods and beverages. The yeasts, Elsevier: 53-63.

Fox, E. M., K. Solomon, J. E. Moore, P. G. Wall and S. Fanning (2014). "Phylogenetic profiles of in-house microflora in drains at a food production facility: comparison and biocontrol implications of Listeria-positive and-negative bacterial populations." Appl. Environ. Microbiol. 80(11): 3369-3374.

Franzosa, E. A., T. Hsu, A. Sirota-Madi, A. Shafquat, G. Abu-Ali, X. C. Morgan and C. Huttenhower (2015). "Sequencing and beyond: integrating molecular'omics' for microbial community profiling." Nature Reviews: Microbiology 13(6): 360-372.

Galazzo, G., N. van Best, B. Benedikter, K. Janssen, L. Bervoets, C. Driessen, M. Oomen, M. Lucchesi, P. van Eijck and H. Becker (2020). "How to count our microbes? The effect of different quantitative microbiome profiling approaches." Front. Cell. Infect. Microbiol. 10: 403. doi: 10.3389/fcimb.

Galvez, A., R. L. López, H. Abriouel, E. Valdivia and N. B. Omar (2008). "Application of bacteriocins in the control of foodborne pathogenic and spoilage bacteria." Critical reviews in biotechnology 28(2): 125-152.

Ganda, E. K., R. S. Bisinotto, S. F. Lima, K. Kronauer, D. H. Decter, G. Oikonomou, Y. H. Schukken and R. C. Bicalho (2016). "Longitudinal metagenomic profiling of bovine milk to assess the impact of intramammary treatment using a third-generation cephalosporin." Scientific reports 6(1): 1-13.

Grützke, J., B. Malorny, J. A. Hammerl, A. Busch, S. H. Tausch, H. Tomaso and C. Deneke (2019). "Fishing in the soup–Pathogen detection in food safety using Metabarcoding and Metagenomic sequencing." Frontiers in microbiology 10: 1805.

Guo, X., Z. Yu, F. Zhao, Z. Sun, L.-Y. Kwok and S. Li (2021). "Both sampling seasonality and geographic origin contribute significantly to variations in raw milk microbiota, but sampling seasonality is the more determining factor." Journal of Dairy Science 104(10): 10609-10627.

Gweon, H. S., L. P. Shaw, J. Swann, N. De Maio, M. AbuOun, R. Niehus, A. T. Hubbard, M. J. Bowes, M. J. Bailey and T. E. Peto (2019). "The impact of sequencing

depth on the inferred taxonomic composition and AMR gene content of metagenomic samples." Environmental Microbiome 14(1): 1-15.

Haiminen, N., S. Edlund, D. Chambliss, M. Kunitomi, B. C. Weimer, B. Ganesan, R. Baker, P. Markwell, M. Davis and B. C. Huang (2019). "Food authentication from shotgun sequencing reads with an application on high protein powders." NPJ science of food 3(1): 1-11.

Handelsman, J. (2004). "Metagenomics: application of genomics to uncultured microorganisms." Microbiology and molecular biology reviews 68(4): 669-685.

Hatti-Kaul, R., L. Chen, T. Dishisha and H. E. Enshasy (2018). "Lactic acid bacteria: From starter cultures to producers of chemicals." FEMS microbiology letters 365(20): fny213.

He, L.-Y., L.-K. He, Y.-S. Liu, M. Zhang, J.-L. Zhao, Q.-Q. Zhang and G.-G. Ying (2019). "Microbial diversity and antibiotic resistome in swine farm environments." Science of the Total Environment 685: 197-207.

Hendriksen, R. S., P. Munk, P. Njage, B. Van Bunnik, L. McNally, O. Lukjancenko, T. Röder, D. Nieuwenhuijse, S. K. Pedersen and J. Kjeldgaard (2019). "Global monitoring of antimicrobial resistance based on metagenomics analyses of urban sewage." Nature communications 10(1): 1124.

Höll, L., M. Hilgarth, A. J. Geissler, J. Behr and R. F. Vogel (2020). "Metatranscriptomic analysis of modified atmosphere packaged poultry meat enables prediction of Brochothrix thermosphacta and Carnobacterium divergens in situ metabolism." Archives of Microbiology 202: 1945-1955.

Hoque, M. N., A. Istiaq, R. A. Clement, M. Sultana, K. A. Crandall, A. Z. Siddiki and M. A. Hossain (2019). "Metagenomic deep sequencing reveals association of microbiome signature with functional biases in bovine mastitis." Scientific reports 9(1): 1-14.

Hultman, J., R. Rahkila, J. Ali, J. Rousu and K. J. Björkroth (2015). "Meat processing plant microbiome and contamination patterns of cold-tolerant bacteria causing

food safety and spoilage risks in the manufacture of vacuum-packaged cooked sausages." Appl. Environ. Microbiol. 81(20): 7088-7097.

Hwang, B. K., H. Choi, S. H. Choi and B.-S. Kim (2020). "Analysis of Microbiota Structure and Potential Functions Influencing Spoilage of Fresh Beef Meat." Frontiers in microbiology 11: 1657.

Illeghems, K., S. Weckx and L. De Vuyst (2015). "Applying meta-pathway analyses through metagenomics to identify the functional properties of the major bacterial communities of a single spontaneous cocoa bean fermentation process sample." Food microbiology 50: 54-63.

in't Veld, J. H. H. (1996). "Microbial and biochemical spoilage of foods: an overview." International journal of food microbiology 33(1): 1-18.

Jääskeläinen, E., L. M. Jakobsen, J. Hultman, N. Eggers, H. C. Bertram and J. Björkroth (2019). "Metabolomics and bacterial diversity of packaged yellowfin tuna (Thunnus albacares) and salmon (Salmo salar) show fish species-specific spoilage development during chilled storage." International journal of food microbiology 293: 44-52.

Jenkins, C., T. J. Dallman and K. A. Grant (2019). "Impact of whole genome sequencing on the investigation of food-borne outbreaks of Shiga toxin-producing Escherichia coli serogroup O157: H7, England, 2013 to 2017." EuroSurveillance 24(4): 1800346.

Jordan, K., M. Dalmasso, J. Zentek, A. Mader, G. Bruggeman, J. Wallace, D. De Medici, A. Fiore, E. Prukner-Radovcic and M. Lukac (2014). "Microbes versus microbes: control of pathogens in the food chain." Journal of the Science of Food and Agriculture 94(15): 3079-3089.

Jung, J. Y., S. H. Lee, H. M. Jin, Y. Hahn, E. L. Madsen and C. O. Jeon (2013). "Metatranscriptomic analysis of lactic acid bacterial gene expression during kimchi fermentation." International journal of food microbiology 163(2-3): 171-179.

Kable, M. E., Y. Srisengfa, M. Laird, J. Zaragoza, J. McLeod, J. Heidenreich and M. L. Marco (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(4): e00836-00816.

Kable, M. E., Y. Srisengfa, Z. Xue, L. C. Coates and M. L. Marco (2019). "Viable and total bacterial populations undergo equipment-and time-dependent shifts during milk processing." Applied and environmental microbiology 85(13): e00270-00219.

Kamilari, E., D. A. Anagnostopoulos, P. Papademas, A. Kamilaris and D. Tsaltas (2020). "Characterizing halloumi cheese's bacterial communities through metagenomic analysis." LWT: 109298.

Kamilari, E., M. Tomazou, A. Antoniades and D. Tsaltas (2019). "High throughput sequencing technologies as a new toolbox for deep analysis, characterization and potentially authentication of protection designation of origin cheeses?" International journal of food science.

Khansaritoreh, E., Y. Salmaki, E. Ramezani, T. A. Azirani, A. Keller, K. Neumann, K. Alizadeh, S. Zarre, G. Beckh and H. Behling (2020). "Employing DNA metabarcoding to determine the geographical origin of honey." Heliyon 6(11): e05596.

Kim, K. H., B. H. Chun, J. Kim and C. O. Jeon (2020). "Identification of biogenic amine-producing microbes during fermentation of ganjang, a Korean traditional soy sauce, through metagenomic and metatranscriptomic analyses." Food Control 121: 107681.

Kim, S., J. Kim, E. J. Yun and K. H. Kim (2016). "Food metabolomics: from farm to human." Current Opinion in Biotechnology 37: 16-23.

King, P., L. K. Pham, S. Waltz, D. Sphar, R. T. Yamamoto, D. Conrad, R. Taplitz, F. Torriani and R. A. Forsyth (2016). "Longitudinal metagenomic analysis of hospital air identifies clinically relevant microbes." PloS one 11(8).

Kleiner, M. (2019). "Metaproteomics: much more than measuring gene expression in microbial communities." MSystems 4(3): e00115-00119.

Langsrud, S., B. Moen, T. Møretrø, M. Løype and E. Heir (2016). "Microbial dynamics in mixed culture biofilms of bacteria surviving sanitation of conveyor belts in salmon-processing plants." Journal of applied microbiology 120(2): 366-378.

Leonard, S. R., M. K. Mammel, D. W. Lacher and C. A. Elkins (2015). "Application of metagenomic sequencing to food safety: detection of Shiga toxin-producing Escherichia coli on fresh bagged spinach." Appl. Environ. Microbiol. 81(23): 8183-8191.

Leroy, F. and L. De Vuyst (2004). "Lactic acid bacteria as functional starter cultures for the food fermentation industry." Trends in Food Science & Technology 15(2): 67-78.

Lewis, E., J. Hudson, N. Cook, J. Barnes and E. Haynes (2020). "Next-generation sequencing as a screening tool for foodborne pathogens in fresh produce." Journal of Microbiological Methods: 105840.

Li, F. and L. L. Guan (2017). "Metatranscriptomic profiling reveals linkages between the active rumen microbiome and feed efficiency in beef cattle." Applied and environmental microbiology 83(9): e00061-00017.

Li, R., C. Wang, G. Zhou, C. Li and K. Ye (2021). "The effects of thermal treatment on the bacterial community and quality characteristics of meatballs during storage." Food Science & Nutrition 9(1): 564-573.

Li, S., D. A. Mann, S. Zhang, Y. Qi, R. J. Meinersmann and X. Deng (2020). "Microbiome-Informed Food Safety and Quality: Longitudinal Consistency and Cross-Sectional Distinctiveness of Retail Chicken Breast Microbiomes." Msystems 5(5).

Liang, L., P. Wang, X. Zhao, L. He, T. Qu and Y. Chen (2022). "Single-molecule real-time sequencing reveals differences in bacterial diversity in raw milk in different regions and seasons in China." Journal of Dairy Science.

Liu, J., Y. Zhu, M. Jay-Russell, D. G. Lemay and D. A. Mills (2020). "Reservoirs of antimicrobial resistance genes in retail raw milk." Microbiome 8(1): 1-15.

Liu, N., D. Zou, D. Dong, Z. Yang, D. Ao, W. Liu and L. Huang (2017). "Development of a multiplex loop-mediated isothermal amplification method for the simultaneous detection of Salmonella spp. and Vibrio parahaemolyticus." Scientific reports 7(1): 1-7.

Liu, X.-F., C.-J. Liu, X.-Q. Zeng, H.-Y. Zhang, Y.-Y. Luo and X.-R. Li (2020). "Metagenomic and metatranscriptomic analysis of the microbial community structure and metabolic potential of fermented soybean in Yunnan Province." Food Science and Technology 40(1): 18-25.

Liu, X., J. S. Teixeira, S. Ner, K. V. Ma, N. Petronella, S. Banerjee and J. Ronholm (2020). "Exploring the potential of the microbiome as a marker of the geographic origin of fresh seafood." Frontiers in Microbiology 11.

Loman, N. J., C. Constantinidou, M. Christner, H. Rohde, J. Z.-M. Chan, J. Quick, J. C. Weir, C. Quince, G. P. Smith and J. R. Betley (2013). "A culture-independent sequence-based metagenomics approach to the investigation of an outbreak of Shiga-toxigenic Escherichia coli O104: H4." Jama 309(14): 1502-1510.

Luiken, R. E., L. Van Gompel, A. Bossers, P. Munk, P. Joosten, R. B. Hansen, B. E. Knudsen, S. García-Cobos, J. Dewulf and F. M. Aarestrup (2020). "Farm dust resistomes and bacterial microbiomes in European poultry and pig farms." Environment International 143: 105971.

Mandal, P. K., A. K. Biswas, K. Choi and U. K. Pal (2011). "Methods for rapid detection of foodborne pathogens: an overview." American Journal of Food Technology 6: 87-102.

Marine, R., C. McCarren, V. Vorrasane, D. Nasko, E. Crowgey, S. W. Polson and K. E. Wommack (2014). "Caught in the middle with multiple displacement amplification: the myth of pooling for avoiding multiple displacement amplification bias in a metagenome." Microbiome 2(1): 3.

Marotz, C. A., J. G. Sanders, C. Zuniga, L. S. Zaramela, R. Knight and K. Zengler (2018). "Improving saliva shotgun metagenomics by chemical host DNA depletion." Microbiome 6(1): 42.

Matthews, C., F. Crispie, E. Lewis, M. Reid, P. W. O'Toole and P. D. Cotter (2019). "The rumen microbiome: a crucial consideration when optimising milk and meat production and nitrogen utilisation efficiency." Gut microbes 10(2): 115-132.

McHugh, A. J., C. Feehily, M. A. Fenelon, D. Gleeson, C. Hill and P. D. Cotter (2020). "Tracking the dairy microbiota from farm bulk tank to skimmed milk powder." Msystems 5(2).

McHugh, A. J., C. Feehily, J. T. Tobin, M. A. Fenelon, C. Hill and P. D. Cotter (2018). "Mesophilic sporeformers identified in whey powder by using shotgun metagenomic sequencing." Appl. Environ. Microbiol. 84(20): e01305-01318.

McHugh, A. J., M. Yap, F. Crispie, C. Feehily, C. Hill and P. D. Cotter (2021). "Microbiome-based environmental monitoring of a dairy processing facility highlights the challenges associated with low microbial-load samples." npj Science of Food 5(4).

Mencía-Ares, O., R. Cabrera-Rubio, J. F. Cobo-Díaz, A. Álvarez-Ordóñez, M. Gómez-García, H. Puente, P. D. Cotter, F. Crispie, A. Carvajal and P. Rubio (2020). "Antimicrobial use and production system shape the fecal, environmental, and slurry resistomes of pig farms." Microbiome 8(1): 1-17.

Mira Miralles, M., L. Maestre-Carballa, M. Lluesma-Gomez and M. Martinez-Garcia (2019). "High-throughput 16S rRNA sequencing to assess potentially active bacteria and foodborne pathogens: a case example in ready-to-eat food." Foods 8(10): 480.

Munk, P., V. D. Andersen, L. de Knegt, M. S. Jensen, B. E. Knudsen, O. Lukjancenko, H. Mordhorst, J. Clasen, Y. Agersø and A. Folkesson (2017). "A sampling and metagenomic sequencing-based methodology for monitoring antimicrobial resistance in swine herds." Journal of Antimicrobial Chemotherapy 72(2): 385-392.

Nocker, A., C.-Y. Cheung and A. K. Camper (2006). "Comparison of propidium monoazide with ethidium monoazide for differentiation of live vs. dead bacteria by selective removal of DNA from dead cells." Journal of microbiological methods 67(2): 310-320.

Nogarol, C., P. Acutis, D. Bianchi, C. Maurella, S. Peletto, S. Gallina, D. Adriano, F. Zuccon, S. Borrello and M. Caramelli (2013). "Molecular characterization of Pseudomonas fluorescens isolates involved in the Italian “blue mozzarella” event." Journal of food protection 76(3): 500-504.

Noyes, N. R., X. Yang, L. M. Linke, R. J. Magnuson, A. Dettenwanger, S. Cook, I. Geornaras, D. E. Woerner, S. P. Gow and T. A. McAllister (2016). "Resistome diversity in cattle and the environment decreases during beef production." Elife 5: e13195.

Nychas, G. E. and E. Panagou (2011). Microbiological spoilage of foods and beverages. Food and Beverage Stability and Shelf Life. D. Kilcast and P. Subarmaniam. Cambridge, UK, Woodhead Publishing: 3-28.

O'Sullivan, L., D. Bolton, O. McAuliffe and A. Coffey (2019). "Bacteriophages in food applications: From foe to friend." Annual review of food science and technology 10: 151-172.

O’Hara, N. B., H. J. Reed, E. Afshinnekoo, D. Harvin, N. Caplan, G. Rosen, B. Frye, S. Woloszynek, R. Ounit and S. Levy (2017). "Metagenomic characterization of ambulances across the USA." Microbiome 5(1): 125.

Ogier, J.-C., S. Pages, M. Galan, M. Barret and S. Gaudriault (2019). "rpoB, a promising marker for analyzing the diversity of bacterial communities by amplicon sequencing." BMC microbiology 19(1): 171.

Ouali, F. A., I. Al Kassaa, B. Cudennec, M. Abdallah, F. Bendali, D. Sadoun, N.-E. Chihib and D. Drider (2014). "Identification of lactobacilli with inhibitory effect on biofilm formation by pathogenic bacteria on stainless steel surfaces." International journal of food microbiology 191: 116-124.

Pasolli, E., F. De Filippis, I. E. Mauriello, F. Cumbo, A. M. Walsh, J. Leech, P. D. Cotter, N. Segata and D. Ercolini (2020). "Large-scale genome-wide analysis links lactic acid bacteria from food with the gut microbiome." Nature communications 11(1): 1-12.

Peirson, M. D., T. Y. Guan and R. A. Holley (2003). "Thermal resistances and lactate and diacetate sensitivities of bacteria causing bologna discolouration." International journal of food microbiology 86(3): 223-230.

Petersen, T. N., S. Rasmussen, H. Hasman, C. Carøe, J. Bælum, A. C. Schultz, L. Bergmark, C. A. Svendsen, O. Lund and T. Sicheritz-Pontén (2015). "Meta-genomic analysis of toilet waste from long distance flights; a step towards global surveillance of infectious diseases and antimicrobial resistance." Scientific reports 5: 11444.

Petruzzi, L., M. R. Corbo, M. Sinigaglia and A. Bevilacqua (2017). Microbial spoilage of foods: Fundamentals. The microbiological quality of food, Elsevier: 1-21.

Pham, N.-P., S. Landaud, P. Lieben, P. Bonnarme and C. Monnet (2019). "Transcription Profiling Reveals Cooperative Metabolic Interactions in a Microbial Cheese-Ripening Community Composed of Debaryomyces hansenii, Brevibacterium aurantiacum, and Hafnia alvei." Frontiers in microbiology 10: 1901.

Pinu, F. R., D. J. Beale, A. M. Paten, K. Kouremenos, S. Swarup, H. J. Schirra and D. Wishart (2019). "Systems biology and multi-omics integration: viewpoints from the metabolomics research community." Metabolites 9(4): 76.

Pitta, D. W., Z. Dou, S. Kumar, N. Indugu, J. D. Toth, B. Vecchiarelli and B. Bhukya (2016). "Metagenomic evidence of the prevalence and distribution patterns of antimicrobial resistance genes in dairy agroecosystems." Foodborne pathogens and disease 13(6): 296-302.

Poirier, S., O. Rue, R. Peguilhan, G. Coeuret, M. Zagorec, M.-C. Champomier-Verges, V. Loux and S. Chaillou (2018). "Deciphering intra-species bacterial diversity of meat and seafood spoilage microbiota using gyrB amplicon sequencing: A comparative analysis with 16S rDNA V3-V4 amplicon sequencing." PLOS ONE 13(9): e0204629.

Pothakos, V., G. Stellato, D. Ercolini and F. Devlieghere (2015). "Processing environment and ingredients are both sources of Leuconostoc gelidum, which emerges as a major spoiler in ready-to-eat meals." Applied and environmental microbiology 81(10): 3529-3541.

Quijada, N. M., M. Dzieciol, S. Schmitz-Esser, M. Wagner and E. Selberherr (2022). "Metatranscriptomic Analyses Unravel Dynamic Changes in the Microbial and Metabolic Transcriptional Profiles in Artisanal Austrian Hard-Cheeses During Ripening." Frontiers in microbiology 13.

Quince, C., A. W. Walker, J. T. Simpson, N. J. Loman and N. Segata (2017). "Shotgun metagenomics, from sampling to analysis." Nature biotechnology 35(9): 833.

Rajawardana, D. U., P. C. Fernando, P. J. Biggs, I. G. N. Hewajulige, C. M. Nanayakkara, S. Wickramasinghe, X. X. Lin and L. Berry (2022). "An insight into tropical milk microbiome: Bacterial community composition of cattle milk produced in Sri Lanka." International Dairy Journal 126: 105266.

Rampelotto, P. H., A. F. Sereia, L. F. V. de Oliveira and R. Margis (2019). "Exploring the hospital microbiome by high-resolution 16S rRNA profiling." International journal of molecular sciences 20(12): 3099.

Rantsiou, K., S. Kathariou, A. Winkler, P. Skandamis, M. J. Saint-Cyr, K. Rouzeau-Szynalski and A. Amézquita (2018). "Next generation microbiological risk assessment: opportunities of whole genome sequencing (WGS) for foodborne pathogen surveillance, source tracking and risk assessment." International journal of food microbiology 287: 3-9.

Reis, J., A. Paula, S. Casarotti and A. Penna (2012). "Lactic acid bacteria antimicrobial compounds: characteristics and applications." Food Engineering Reviews 4(2): 124-140.

Rodríguez-López, P., J. J. Rodríguez-Herrera and M. L. Cabo (2020). "Tracking bacteriome variation over time in Listeria monocytogenes-positive foci in food industry." International journal of food microbiology 315: 108439.

Rubiola, S., F. Chiesa, A. Dalmasso, P. Di Ciccio and T. Civera (2020). "Detection of Antimicrobial Resistance Genes in the Milk Production Environment: Impact of Host DNA and Sequencing Depth." Frontiers in Microbiology 11: 1983.

Sahu, M. and S. Bala (2017). "Food processing, food spoilage and their prevention: an overview." Int. J. Life. Sci. Scienti. Res 3(1): 753-759.

Schallmey, M., A. Singh and O. P. Ward (2004). "Developments in the use of Bacillus species for industrial production." Canadian journal of microbiology 50(1): 1-17.

Schwan, R. F. and C. L. Ramos (2014). Role of microbes and their diversity in fermented foods. Beneficial microbes in fermented and functional foods. V. Ravishankar Rai and J. A. Bai. Boca Raton, FL, CRC Press: 524-547.

Shakya, M., C.-C. Lo and P. S. Chain (2019). "Advances and challenges in Metatranscriptomic analysis." Frontiers in Genetics 10: 904.

Slizovskiy, I. B., K. Mukherjee, C. J. Dean, C. Boucher and N. R. Noyes (2020). "Mobilization of antibiotic resistance: are current approaches for colocalizing resistomes and mobilomes useful?" Frontiers in microbiology 11: 1376.

Soggiu, A., C. Piras, S. L. Mortera, I. Alloggio, A. Urbani, L. Bonizzi and P. Roncada (2016). "Unravelling the effect of clostridia spores and lysozyme on microbiota dynamics in Grana Padano cheese: a metaproteomics approach." Journal of proteomics 147: 21-27.

Sohier, D., S. Pavan, A. Riou, J. Combrisson and F. Postollec (2014). "Evolution of microbiological analytical methods for dairy industry needs." Frontiers in Microbiology 5: 16.

Sørensen, J. S., N. Bøknæs, O. Mejlholm and P. Dalgaard (2020). "Superchilling in combination with modified atmosphere packaging resulted in long shelf-life and limited microbial growth in Atlantic cod (Gadus morhua L.) from capture-based-aquaculture in Greenland." Food microbiology 88: 103405.

Stellato, G., A. La Storia, F. De Filippis, G. Borriello, F. Villani and D. Ercolini (2016). "Overlap of spoilage-associated microbiota between meat and the meat processing environment in small-scale and large-scale retail distributions." Appl. Environ. Microbiol. 82(13): 4045-4054.

Stohr, V., J.-J. Joffraud, M. Cardinal and F. Leroi (2001). "Spoilage potential and sensory profile associated with bacteria isolated from cold-smoked salmon." Food Research International 34(9): 797-806.

Tao, J., W. Liu, W. Ding, R. Han, Q. Shen, Y. Xia, Y. Zhang and W. Sun (2020). "A multiplex PCR assay with a common primer for the detection of eleven foodborne pathogens." Journal of food science 85(3): 744-754.

Tsironi, T., L. Anjos, P. I. Pinto, G. Dimopoulos, S. Santos, C. Santa, B. Manadas, A. Canario, P. Taoukis and D. Power (2019). "High pressure processing of European sea bass (Dicentrarchus labrax) fillets and tools for flesh quality and shelf life monitoring." Journal of Food Engineering 262: 83-91.

Van Gompel, L., R. E. Luiken, S. Sarrazin, P. Munk, B. E. Knudsen, R. B. Hansen, A. Bossers, F. M. Aarestrup, J. Dewulf and J. A. Wagenaar (2019). "The antimicrobial resistome in relation to antimicrobial use and biosecurity in pig farming, a metagenome-wide association study in nine European countries." Journal of Antimicrobial Chemotherapy 74(4): 865-876.

Vikram, A., J. Woolston and A. Sulakvelidze (2020). "Phage Biocontrol Applications in Food Production and Processing." Bacterial Viruses: Exploitation for Biocontrol and Therapeutics: 283-334.

Walsh, A. M., F. Crispie, O. O'Sullivan, L. Finnegan, M. J. Claesson and P. D. Cotter (2018). "Species classifier choice is a key consideration when analysing low-complexity food microbiome data." Microbiome 6(1): 50.

Wang, H., X. Zhang, G. Wang, K. Jia, X. Xu and G. Zhou (2017). "Bacterial community and spoilage profiles shift in response to packaging in yellow-feather broiler, a highly popular meat in Asia." Frontiers in microbiology 8: 2588.

Wang, X., H. Du, Y. Zhang and Y. Xu (2018). "Environmental microbiota drives microbial succession and metabolic profiles during Chinese liquor fermentation." Appl. Environ. Microbiol. 84(4): e02369-02317.

Wang, Y., Y. Hu, F. Liu, J. Cao, N. Lv, B. Zhu, G. Zhang and G. F. Gao (2020). "Integrated metagenomic and metatranscriptomic profiling reveals differentially expressed resistomes in human, chicken, and pig gut microbiomes." Environment International 138: 105649.

Wang, Y. and J. K. Salazar (2016). "Culture-independent rapid detection methods for bacterial pathogens and toxins in food matrices." Comprehensive Reviews in Food Science and Food Safety 15(1): 183-205.

Wang, Y., Y. Yan, K. N. Thompson, S. Bae, E. K. Accorsi, Y. Zhang, J. Shen, H. Vlamakis, E. M. Hartmann and C. Huttenhower (2021). "Whole microbial community viability is not quantitatively reflected by propidium monoazide sequencing approach." Microbiome 9(1): 1-13.

World Health Organisation (2017). The burden of foodborne diseases in the WHO European Region. Copenhagen, Denmark, WHO Regional Office for Europe: 48.

Wu, H., Q. D. Nguyen, T. T. Tran, M. T. Tang, T. Tsuruta and N. Nishino (2019). "Rumen fluid, feces, milk, water, feed, airborne dust, and bedding microbiota in dairy farms managed by automatic milking systems." Animal Science Journal 90(3): 445-452.

Wu, L., G. Li, X. Xu, L. Zhu, R. Huang and X. Chen (2019). "Application of nano-ELISA in food analysis: recent advances and challenges." TrAC Trends in Analytical Chemistry 113: 140-156.

Xiao, Y., T. Huang, Y. Xu, Z. Peng, Z. Liu, Q. Guan, M. Xie and T. Xiong (2020). "Metatranscriptomics reveals the gene functions and metabolic properties of the major microbial community during Chinese Sichuan Paocai fermentation." Food Microbiology: 103573.

Xie, M., J. Wu, F. An, X. Yue, D. Tao, R. Wu and Y. Lee (2019). "An integrated metagenomic/metaproteomic investigation of microbiota in dajiang-meju, a traditional fermented soybean product in Northeast China." Food research international 115: 414-424.

Yang, M., A. Cousineau, X. Liu, Y. Luo, D. Sun, S. Li, T. Gu, L. Sun, H. Dillow and J. Lepine (2020). "Direct metatranscriptome RNA-seq and multiplex RT-PCR amplicon sequencing on Nanopore MinION–promising strategies for multiplex identification of viable pathogens in food." Frontiers in microbiology 11: 514.

Yang, X., N. R. Noyes, E. Doster, J. N. Martin, L. M. Linke, R. J. Magnuson, H. Yang, I. Geornaras, D. R. Woerner and K. L. Jones (2016). "Use of metagenomic shotgun sequencing technology to detect foodborne pathogens within the microbiome of the beef production chain." Appl. Environ. Microbiol. 82(8): 2433-2443.

Yap, M., C. Feehily, C. J. Walsh, M. Fenelon, E. F. Murphy, F. M. McAuliffe, D. van Sinderen, P. W. O'Toole, O. O'Sullivan and P. D. Cotter (2020). "Evaluation of methods for the reduction of contaminating host reads when performing shotgun metagenomic sequencing of the milk microbiome." Scientific reports 10(1): 1-11.

Yuan, L., M. F. Hansen, H. L. Røder, N. Wang, M. Burmølle and G. He (2020). "Mixed-species biofilms in the food industry: current knowledge and novel control strategies." Critical reviews in food science and nutrition 60(13): 2277-2293.

Zhang, Y., M. Kitajima, A. J. Whittle and W.-T. Liu (2017). "Benefits of genomic insights and CRISPR-Cas signatures to monitor potential pathogens across drinking water production and distribution systems." Frontiers in microbiology 8: 2036.

Zotta, T., E. Parente, R. G. Ianniello, F. De Filippis and A. Ricciardi (2019). "Dynamics of bacterial communities and interaction networks in thawed fish fillets during chilled storage in air." International journal of food microbiology 293: 102-113.

Zwirzitz, B., S. U. Wetzels, E. D. Dixon, B. Stessl, A. Zaiser, I. Rabanser, S. Thalguter, B. Pinior, F.-F. Roch and C. Strachan (2020). "The sources and transmission routes of microbial populations throughout a meat processing facility." NPJ biofilms and microbiomes 6(1): 1-12.

Chapter 2. Seasonality and geography have a greater influence than the use of chlorine-based cleaning agents on the microbiota of bulk tank raw milk.

Included as published in Applied and Environmental Microbiology
(doi: 10.1128/AEM.01081-21)

Authors: Min Yap, Dave Gleeson, Paul W. O'Toole, Orla O'Sullivan and Paul D. Cotter

Contributions: Candidate performed sample preparation, DNA extraction, sequencing library preparation, bioinformatic and statistical analysis.
Candidate drafted and edited chapter.
All authors contributed, revised, and approved manuscript.

2.1 Abstract

Cleaning of the production environment is vital to ensure the safety and quality of dairy products. Although cleaning with chlorine-based agents is widely adopted, it has been associated with detrimental effects on milk quality and safety, which has garnered increasing interest in chlorine-free cleaning. However, the influence of these methods on the milk microbiota is not well documented. This study investigated the factors that influence the raw milk microbiota, with a focus on the differences when chlorine-based and chlorine-free cleaning of milking equipment are used. Bulk tank raw milk was sampled at three sampling months (Apr, Aug and Nov), from farms across Ireland selected to capture the use of different cleaning methods, i.e., exclusively chlorine-based (n = 51) and chlorine-free cleaning (n = 92), and farms that used chlorine-free agents for the bulk tank and chlorine-based cleaning agents for the rest of the equipment (n = 28). Shotgun metagenomic analysis revealed the significant influence of seasonal and geographic factors on the bulk tank milk microbiota, indicated by differences in diversity, taxonomic composition, and functional characteristics. Taxonomic and functional profiles of samples collected in November clustered separately from other months. In contrast, cleaning methods only accounted for 1% of the variation in the bulk tank milk bacterial community, and samples collected from farms using chlorine relative to chlorine-free cleaning did not differ significantly, suggesting that chlorine-free approaches used did not negatively impact microbiological quality. This study shows the value of shotgun metagenomics in advancing our knowledge of the raw milk microbiota.

2.1.1 Importance

The microbiota of raw milk is affected by many factors that can control or promote the introduction of undesirable microorganisms. Chlorine-based cleaning agents have been commonly used due to their effectiveness in controlling undesirable microorganisms but have been associated with the formation of chlorine residues that are detrimental to product quality and may impact consumer health. Chlorine-free alternatives have been recommended in some countries, but the influence of cleaning agents on the milk microbiota is unknown. Here we investigated the influence of cleaning methods and other factors on bulk tank raw milk. Results showed that season and location had a greater influence on the milk microbiota than the cleaning agents used. Indeed, the similar microbiota compositions of raw milk from farms that used chlorine-based and those that used chlorine-free cleaning methods supports the further use of chlorine-free cleaning agents in dairy production.

2.2 Introduction

The raw milk microbiota is complex, with many factors contributing to its composition (Quigley, O'Sullivan et al. 2013). Understanding the factors that positively or negatively impact the microbiota of raw milk is important as it can affect the safety and quality of foods produced. Cleaning of equipment is one important aspect of food processing that if done inefficiently, may cause the unwanted transfer of microorganisms from food production and processing surfaces to product. Chlorine-based cleaning products have been commonly used in many industries such as agriculture, water treatment and food processing due to

their efficient killing of a broad range of microorganisms (Reinemann, Wolters et al. 2003), including by disrupting protein synthesis and by increasing cell membrane permeability (Virto, Manas et al. 2005, Gómez-López, Rajkovic et al. 2009). In dairy production and processing, chlorine, in the form of hypochlorites or gaseous chlorine, is widely used in cleaning products for the disinfection of equipment (Reinemann, Wolters et al. 2003, Pandey, Kumari et al. 2014, McCarthy, O'Callaghan et al. 2018). However, the use of chlorine has been associated with the formation of total organic chlorine residues, which have been found to affect the end-product quality and may pose a health risk to consumers (EFSA Panel on Contaminants in the Food Chain 2015). Such chlorine residues include trichloromethane, which has been classified as a potential carcinogen, and chlorite, chlorate, and perchlorate, which have been found to cause oxidative stress in cells and/or to cause thyroid dysfunction (Tiefel and Guthy 1997, Pandey, Kumari et al. 2014, EFSA Panel on Contaminants in the Food Chain 2015). As a result, alternative cleaning methods involving the use of chlorine-free detergents and sanitisers have been recommended as substitutes. In this regard, it is notable that chlorine-free detergents and a detergent containing sodium hypochlorite resulted in similar post-treatment levels of bacteria, as determined by culture-based approaches, in bulk tank milk (Gleeson, O'Brien et al. 2013). However, other culture-based investigations have revealed that chlorine-based cleaning more effectively reduced levels of spore-forming bacteria than chlorine-free (Sundberg, Christiansson et al. 2011) or mechanical cleaning (Ostrov, Harel et al. 2016). Greater resolution of the influence of the cleaning detergents used in dairy production on the microbiological

community in raw milk will help determine if the removal of chlorine-based agents impacts on milk quality.

High-throughput sequencing has been used to characterise the milk microbiome, including to identify mastitis-associated pathogens and to assess the safety and quality of milk for downstream use in dairy products (Parente, Ricciardi et al. 2020). While amplicon sequencing, such as the targeting of the 16S rRNA gene, has been most widely used, the greater taxonomic resolution, to species level, afforded by shotgun metagenomic sequencing is needed to distinguish harmless and undesirable taxa from the same genus (McHugh, Feehily et al. 2020) and to facilitate further characterization of the functional profiles of the microbiota, including antibiotic resistance and toxin determinants (Franzosa, Hsu et al. 2015). Regardless of the approach taken, studies have revealed that the composition of bulk tank milk is significantly affected by many factors such as season, lactation stage, farm system, feed and more (Kable, Srisengfa et al. 2016, Doyle, Gleeson et al. 2017). The fluctuating abundances of several taxa based on season can have important implications for the quality and safety of milk (Li, Wang et al. 2018, Porcellato, Aspholm et al. 2018, Kable, Srisengfa et al. 2019). Moreover, different farm management practices and other environmental factors also influence the microbiota of raw milk (Doyle, Gleeson et al. 2017). Understanding these factors and their influence on the raw milk microbiota can help inform food safety and quality decisions for downstream processing. This present study was run concurrently with Gleeson and colleagues who investigated samples under the same experimental conditions using culture-based and chemical analyses. The aim

of this study was to determine the factors influencing the bulk tank raw milk microbiota using shotgun metagenomics, with particular focus on differences in taxonomic or functional characteristics associated with the use of chlorine-based and chlorine-free cleaning agents.

2.3 Materials and Methods

2.3.1 Sample collection and preparation.

Raw bovine milk samples (100 ml) were collected from bulk milk tanks from farms across Ireland on single days in April (Apr), August (Aug), and November 2019 (Nov). Commercial farms, located mainly in the southern region of Ireland, were identified by 4 milk processors (locations A – D) to take part in the study. Farmers agreed to comply with three different cleaning methods: the use of chlorine-free cleaning products for the bulk tank only (BTCF), chlorine-free cleaning methods throughout the system (CF) and traditional chlorine cleaning protocols (C) throughout the study period. Chlorine-free cleaning involved the use of hot water washes with acid with five suggested protocols available online for farmer's to use for implementation (https://www.teagasc.ie/media/website/animals/dairy/research-farms/Chlorine-free-wash-routines_2020.pdf). The samples were transported in a cooler box to the laboratory and prepared as follows: 30 ml of the bovine milk sample was centrifuged at 4,500 x g for 20 min at 4°C. After centrifugation, the cream and supernatant were discarded and the pellets were subjected to two washing steps, whereby the pellets were resuspended in sterile PBS and centrifuged at 13,000 x g

for 1 minute, after which the supernatant was discarded, and the pellet used for DNA extraction.

2.3.2 DNA extraction.

DNA from samples were extracted on the same day of sample collection. Pellets from the removal of cream were resuspended in 800 µl CD1 solution from the DNeasy PowerSoil Pro kit (Qiagen, West Sussex, United Kingdom). From this point, extraction was carried out according to the manufacturer's instructions. The lysis step was performed using a TissueLyser II (Qiagen) for 10 minutes at 30 Hz and DNA was eluted in 70 µl and stored at -20^{o}C until use in library preparation.

2.3.3 DNA library preparation and shotgun metagenomic sequencing.

DNA was quantified using the Qubit double-stranded DNA (dsDNA) high sensitivity assay kit (Thermo Fisher Scientific, Ireland). All samples were prepared for shotgun metagenomic sequencing according to Illumina Nextera XT library preparation kit guidelines, with the use of unique dual indexes for multiplexing with the Nextera XT index kit (Illumina). Following indexing and clean up, samples were pooled to equimolar concentration of 1 nM. Samples were sequenced on an Illumina NextSeq 550 sequencing platform with a V2 kit, at the Teagasc DNA Sequencing Facility, using standard Illumina sequencing protocols. Sequencing controls and negative controls were used to ensure consistency between runs, and in total, 181 samples were sequenced (3 timepoints for 57 farms + controls).

2.3.4 Bioinformatic analysis of shotgun metagenomic data.

Raw metagenomic shotgun reads were quality checked and trimmed with Cutadapt (v. 1.18) and FastQC (v. 0.11.8). Reads were then aligned to the bovine genome (*Bos taurus*) to determine the number of host reads using Bowtie2 (v. 2.3.4). Kraken2 was used for taxonomic classification of sample data using the Genome Taxonomy Database (release 89) which contains Bacteria and Archaea (Parks, Chuvochina et al. 2018, Wood, Lu et al. 2019, Parks, Chuvochina et al. 2020). SUPER-FOCUS (Silva, Green et al. 2016) was used to characterise the microbiological functional potential of shotgun reads, through the alignment of reads against a reduced SEED (Overbeek, Begley et al. 2005) database using DIAMOND (Buchfink, Xie et al. 2015). Antibiotic resistance genes were quantified using ShortBRED, which maps shotgun reads against markers from the Comprehensive Antibiotic Resistance Databases (CARD) (Jia, Raphenya et al. 2016) and results are expressed as normalised reads per kilobase per million reads (RPKM) (Kaminski, Gibson et al. 2015).

2.3.5 Climate data.

Data on the rainfall (mm), maximum and minimum temperatures (°C), grass minimum temperature (°C) and mean wind speed (knots) from weather stations that were representative of the four sampling locations on the three sampling days were retrieved from the Irish Meteorological Service website (www.met.ie).

2.3.6 Statistical analysis and data visualisation.

Statistical analysis and data visualisation were performed in R (3.6.3) (R Core Team 2013). Kruskal-Wallis and pairwise Wilcoxon rank sum tests with Benjamini-Hochberg P-value correction were used for comparison between cleaning methods,

sampling months and sampling locations. Diversity analysis was done using the vegan package (Oksanen, Blanchet et al. 2019), with alpha diversity calculated as Observed Species, Shannon and Simpson metrics and beta diversity as Bray Curtis metrics, visualised in a principal coordinate analysis plot. The adonis function from the vegan package was used to determine differences in composition of the community between groups of samples (PERMANOVA, number of permutations = 999). Redundancy analysis (RDA) was performed using the vegan package to determine differences in the relative contribution of weather variables on microbiota composition between sampling locations. The multiplatt function from the indicspecies package was used to find species that were significantly associated to particular cleaning methods, sampling months and sampling locations, by calculating Pearson's phi coefficient of association and correcting for unequal group sizes using the parameter 'r.g' (Cáceres and Legendre 2009). Data cleaning and analysis and visualisation was done using tidyverse, ggplot2, ggord and ggpubr packages (Beck 2017, Wickham, Averick et al. 2019, Kassambara 2020).

2.4 Results

2.4.1 Sampling month and location had greater influence on the taxonomic diversity and profiles of the bulk tank milk microbiota than cleaning method.

A total of 171 raw milk samples collected from 57 bulk milk tanks across Ireland were sampled over each of the three sampling months. Some farms changed their cleaning routine during the sampling months and as a result, 51 samples were from farms that used chlorine-based agents in their cleaning routine (C), 92 samples

were from farms which used a chlorine-free cleaning routine (CF), and 28 samples were from farms that used chlorine-free agents for bulk tank cleaning and chlorine-based cleaning methods for the rest of the equipment (BTCF). There were varying sample sizes between sampling locations (A: 45, B: 45, C: 57, D: 24). Further details on the breakdown of samples are in Supplemental Table 2.1. Shotgun metagenomic sequencing of the raw milk samples generated an average of 7,204,026 (± 1,759,482) high-quality paired-end reads per sample. A high proportion (mean: 98.01%, range: 82.1 – 99.5%) of metagenomic reads sequenced were host reads that aligned to the *Bos taurus* genome. An average of 62,976 (± 106,273) reads were assigned as non-host reads (mean: 1.99%, range: 0.49 - 17.9%).

The alpha diversity (Observed species, Shannon, and Simpson metrics) of the bulk tank milk microbiota was not significantly different across samples from the 3 different cleaning methods (Figure 2.1). When compared by sampling month, the alpha diversity of raw milk sampled in April and August was not significantly different, and both were significantly lower ($P < 0.001$), across all three metrics used, than the alpha diversity of raw milk sampled in November. Between locations, the alpha diversity of location A was significantly higher than location B and D in all 3 metrics, and location C in Simpson scores ($P < 0.05$). For beta diversity, samples did not cluster based on cleaning method, which was found to only account for 1.6% of the variation of the bacterial community (Figure 2.1b, Table 2.1). On the other hand, significant differences were apparent by sampling month, where Nov samples clustered distinctly away from the Apr and Aug samples. Similarly, based on sampling location, location B samples clustered separately from

other locations. Both sampling month and location significantly contributed to the variation in microbiota composition ($P < 0.01$), accounting for 10.1% and 6.7% respectively (Table 2.1).

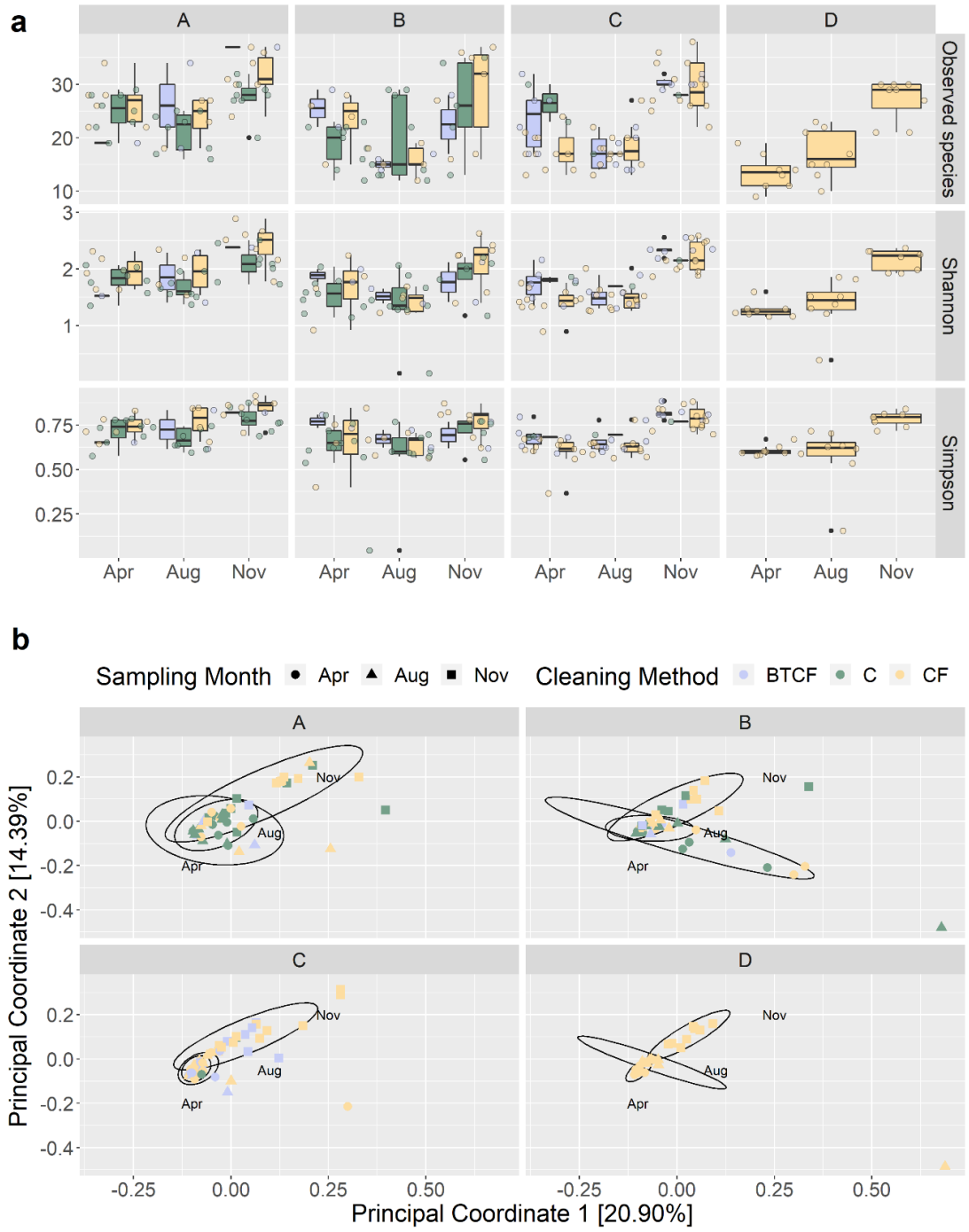


Figure 2.1. Diversity analysis of the microbiota of bulk tank milk samples. (a) Observed species, Shannon, and Simpson alpha diversity analysis of bulk tank milk samples. (b) Bray Curtis PCoA plots, with ellipses representing clustering by

Sampling Month. Samples are from farms from 3 different sampling months – April (Apr), August (Aug) and November (Nov) from 4 different sampling locations (A to D). Samples are classified in 3 different groups of cleaning methods, with samples from farms where chlorine agents are used for cleaning bulk tanks only (BTCF), farms which exclusively use chlorine (C) and chlorine-free agents (CF) in their cleaning.

Table 2.1. PERMANOVA analysis of taxonomic composition and functional properties (based on results from SUPER-FOCUS) of the bulk tank milk microbiota.

Taxonomic composition	R^2	*P* value
Cleaning Method	0.016	0.069
Sampling Month	0.106	0.001
Sampling Location	0.067	0.001
Cleaning Method*Sampling Month	0.014	0.901
Cleaning Method*Sampling Location	0.020	0.389
Sampling Month*Sampling Location	0.054	0.001
Cleaning Method*Sampling Month*Sampling Location	0.036	0.55
Residuals	0.687	
Functional profiles		
Cleaning Method	0.011	0.477
Sampling Month	0.041	0.001
Sampling Location	0.037	0.003
Cleaning Method*Sampling Month	0.017	0.859
Cleaning Method*Sampling Location	0.021	0.541
Sampling Month*Sampling Location	0.043	0.064
Cleaning Method*Sampling Month*Sampling Location	0.040	0.661
Residuals	0.790	

Given the possibility of the cleaning method confounding the impact of sampling month and location, results were stratified into the 3 cleaning methods and diversity and PERMANOVA analysis were done to determine the effect on community structure. Nov samples, irrespective of cleaning method, had a significantly higher alpha diversity and clustered separately from Apr and Aug samples ($P < 0.05$) (Supplemental figure 2.1). In terms of sampling location, alpha diversity was higher for location A across all cleaning methods, and particularly significantly higher for all metrics in farms that used chlorine-free cleaning routines (Supplemental figure 2.1). Location B clustered slightly further away from other locations, with significant differences between locations found the other 2 cleaning methods apart from samples from farms that used chlorine-based agents ($P <0.201$) (Supplemental table 2.2).

A total of 56 species were present at average relative abundances greater than 0.01% across all samples as shown in Figure 2.2. *Anaplasma phagocytophilum, Acinetobacter albensis* and *Murimonas intestini* were the three most abundant species in all samples. No differences were found in the species abundances between cleaning methods, but significant differences were found in 27 species between sampling months and 24 species between sampling locations ($P < 0.05$) (Supplemental Figure 2.2, Supplemental Table 2.3-2.4). Additional analysis for indicator species was performed to determine if any particular species was significantly associated with any cleaning method, sampling month or sampling location. While no species were found to be associated specifically with any cleaning method, 16 species were associated with at least one sampling month and

14 species were associated with at least one sampling location. For sampling months, *Ralstonia insidiosa* and *Microbacterium esteraromaticum_A* was associated with Apr ($P < 0.05$) and *Corynebacterium xerosis*, CAG-791 sp900101015, *Psychrobacter* sp002352555, *Microbacterium maritypicum*, *Aerococcus urinaeequi*, *Jeotgalicoccus* sp003513765, CAG-791 sp900320325, *Psychrobacter* sp001652315, *Rhodococcus erythropolis_D*, *Kocuria* sp002295155, CAG-791 sp900318375, *Facklamia_A* sp003521095, RUG420 sp900317985 and *Kocuria atrinae* were associated with Nov samples ($P < 0.05$). For sampling locations, CAG-791 sp900101015, *Aerococcus urinaeequi*, *Escherichia coli_D*, *Enterococcus faecalis*, CAG-791 sp900318375, *Facklamia_A* sp003521095, *Paracoccus* sp000787695, RUG420 sp900317985 and *Lactobacillus_C paracasei* were associated with location A ($P < 0.05$) and *Anaplasma phagocytophilum*, *Clostridium_F botulinum_A*, *Bacillus_AM* sp001989355, *Microbacterium esteraromaticum_A* and *Mycolicibacterium malmesburyense* with location C ($P < 0.05$). No particular species were associated with location B and D specifically, though *Acinetobacter* spp. was detected in greater relative abundances in both these locations alone.

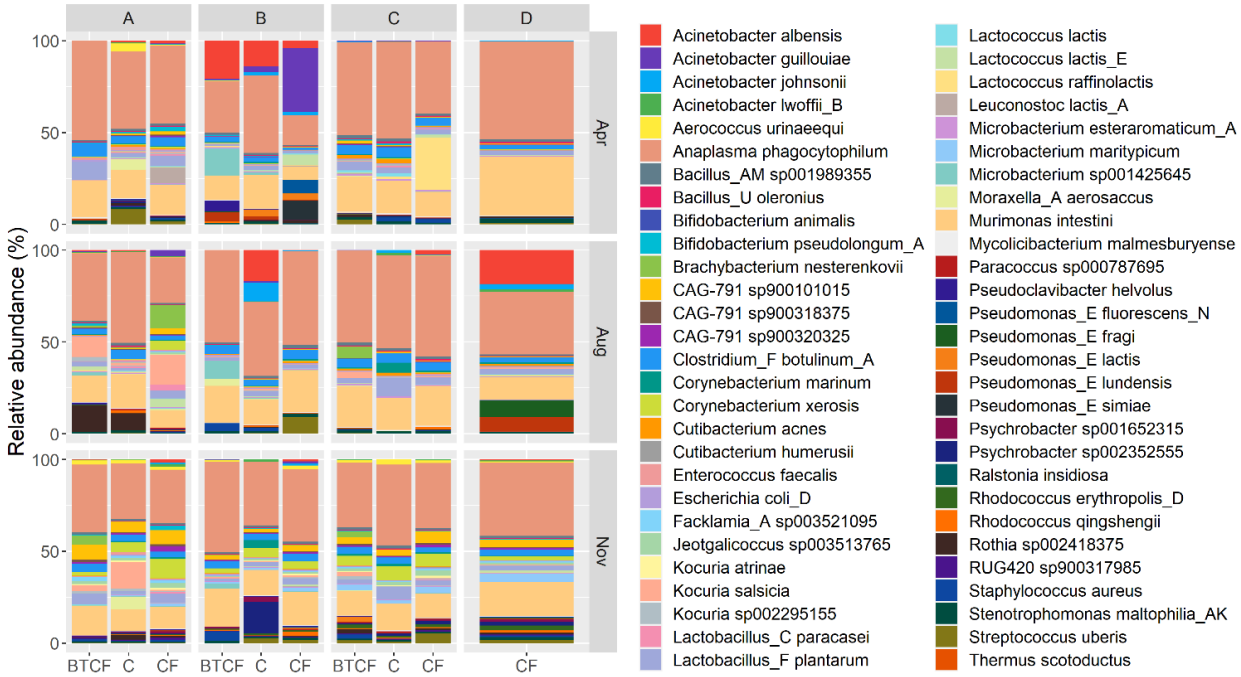


Figure 2.2. Taxonomic profiles of the bulk tank milk microbiota at relative abundances greater than 0.01%. Samples were taken at 3 different sampling months – April (Apr), August (Aug) and November (Nov) from 4 different sampling locations (A to D).

Climate data on the temperature, rainfall and wind from sampling days was retrieved in an effort to correlate this climatic data with seasonal and geographical differences (Supplemental Table 2.5). Analysis indicated that these climatic factors significantly influenced the microbiota (adjusted R^2: 0.0785, $P < 0.01$). Redundancy analysis showed that in terms of the climatic factors, the locations only differed slightly and still clustered close together (Supplemental Figure 2.3). Location B and D cluster more closely together and separately from A and C, though none of the climatic factors analysed can visibly account for it.

2.4.2 Similarly, functional characteristics of the bulk tank raw milk microbiome were influenced by sampling month and location

The functional potential of bulk tank milk microbiomes was predicted using SUPER-FOCUS, with genes involved in the metabolism of carbohydrates, protein and amino acids and derivatives being most abundant (Figure 2.3). Genes associated with oxidative stress response, such as glutathione-related genes or those with protection from reactive oxygen species, are potentially related to chlorine resistance. These genes were detected in very low relative abundances (0 – 0.1%), with no significant differences detected between any cleaning method, sampling month or sampling location. No genes encoding proteins that respond to reactive chlorine species were found in any samples. Overall, no significant differences in functional potential on the basis of cleaning method were found. Between sampling months, a total of 11 functional groups at subsystems level 1 significantly differed ($P < 0.05$) (Supplemental Figure 2.4, Supplemental Table 2.6). Significant differences were found between sampling months for 28 functional groups at subsystem level 2 and 98 groups at level 3 ($P < 0.01$) (Supplemental Table 2.6). Genes related to protein metabolism were seen to gradually increase, while those related to carbohydrates and fatty acids, lipids and isoprenoids gradually decreased over time (Supplemental Table 2.6). Between sampling locations, no differences were found at subsystem levels 1 and 2. However, at level 3, significant differences in abundances were noted for 1 functional group, ESAT-6 proteins secretion system in Mycobacteria (locus ESX-1), which was significantly higher in location A (0.0367%) compared to the other locations (B: 0.0148%, C: 0.0129% and D: 0.0075%).

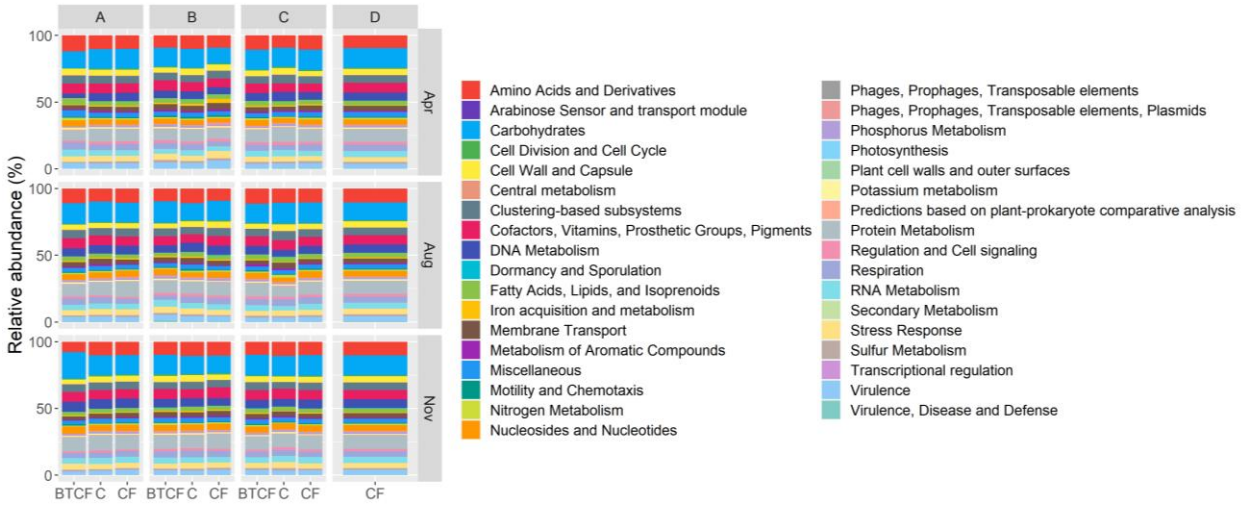


Figure 2.3. Functional profiles of the bulk tank milk microbiota, based on SUPER-FOCUS subsystem level 1 functions. Samples were taken at 3 different sampling months – April (Apr), August (Aug) and November (Nov) from 4 different sampling locations (A to D).

2.4.3 Higher number of antibiotic resistance gene marker reads were mapped to samples from farms which used chlorine-based cleaning agents compared to other cleaning methods

Genes encoding antibiotic resistance were identified using ShortBRED and within bulk tank milk samples, 32 unique antibiotic resistance gene (ARG) markers were detected. These gene markers were found in 19 milk samples (11.1% of all milk samples). Antibiotic resistance gene markers corresponded to 6 different antibiotic classes, i.e, tetracycline, lincosamide, aminoglycoside, sulfonamide, rifampin, and erythromycin. Although the same number of genes were detected between samples from farms that used chlorine-free cleaning and chlorine-based cleaning (Figure 2.4a), higher reads per kilobase of reference sequence per million sample

reads (RPKMs) were seen in milk from farms that used chlorine-based cleaning agents (Figure 2.4b). Although not statistically different, average reads mapped to samples from farms that used chlorine-free agents for bulk tank cleaning alone were lower (11.12 ± 5.27 RPKMs) compared to samples from farms that used a chlorine-free cleaning routine (9.34 ± 8.90 RPKMs) and samples from farms that used chlorine-based agents in the cleaning (30.06 ± 49.30 RPKMs).

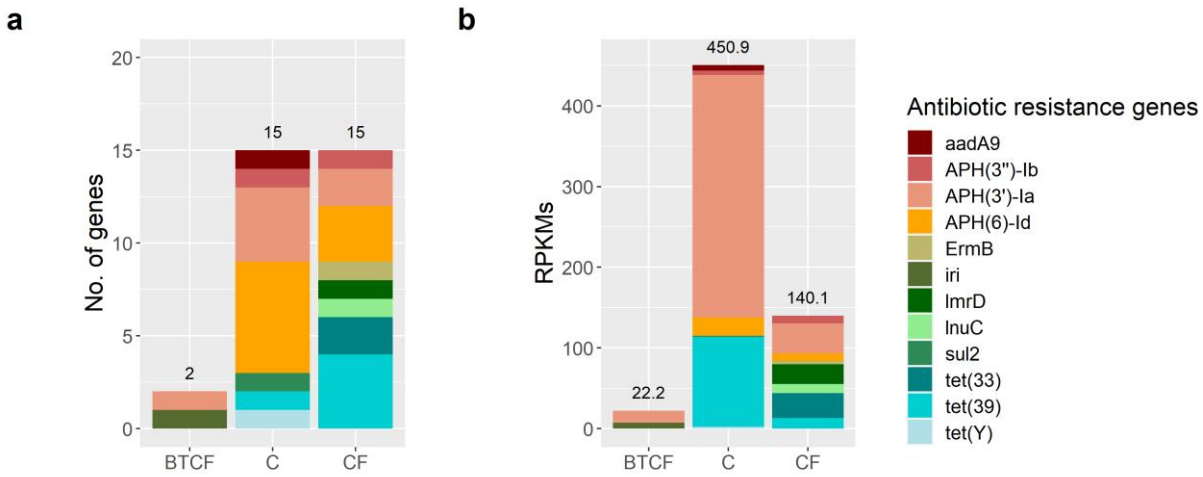


Figure 2.4. Antibiotic resistance genes (ARGs) found in milk samples from farms with different cleaning methods, expressed as (a) number of genes detected and (b) normalised reads per kilobase per million reads (RPKMs).

2.5 Discussion

Here we investigated the factors influencing the bulk tank milk microbiota and, particularly, whether there were taxonomic and functional differences between the use of chlorine-free or chlorine-based cleaning agents on the bulk tank raw milk microbiota. Milk samples were collected across three sampling months and four

different sampling locations, and their microbiome profile assessed. Overall, taxonomic profiles of the raw milk sampled were somewhat different from other studies on the raw milk microbiome previously reported. *Acinetobacter*, *Lactococcus*, *Streptococcus*, *Corynebacterium* and *Pseudomonas* have frequently been reported as the most commonly detected genera in raw milk that are present in high relative abundance (Parente, Ricciardi et al. 2020) and, although these genera were among the top 10 detected in this study, *Anaplasma* and *Murimonas* were also found in high relative abundances. Both *Anaplasma* and *Murimonas* are host-related and therefore their presence is not entirely unexpected (Silaghi, Nieder et al. 2018, Taponen, McGuinness et al. 2019). The identification of *Anaplasma* in raw milk had been previously reported, as several species have been known to cause disease in cattle (Zhang, Lv et al. 2016). *Murimonas* is part of the family *Lachnospiraceae*, which has been commonly found in bovine milk and dairy farm environments (Taponen, McGuinness et al. 2019, Du, Meng et al. 2020), though few studies have not assigned the particular genus or species. The improved taxonomic assignment of shotgun sequencing could be the reason for the identification of *Murimonas intestini* in bulk tank raw milk samples in this study.

The characterisation of the microbiota revealed similar taxonomic compositions and functional profiles regardless of the cleaning method used and only accounted for 1% of the variation in the bulk tank milk bacterial community. These results provide additional evidence supporting previous culture-based investigations by Gleeson et al. (2013), showing that the use of chlorine or non-chlorine cleaning does not affect the composition of raw milk. Interestingly, the taxonomic profile

and diversity of sampling locations B and D were more similar to one another than other locations, despite B having a mix of samples from farms that used all three cleaning methods and D comprising only farms that used chlorine-free cleaning. This result further shows that cleaning method (the use of chlorine or chlorine-free cleaning agents) does not significantly impact the milk microbiome and other factors have a greater influence, as evidenced by cleaning method accounting for a small percentage of the variation in taxonomic and functional profiles that was not found to be statistically significant (Table 2.1).

Seasonality is known to impact the microbiota composition of raw milk, where distinct seasonal differences in diversity and composition have been noted (Li, Wang et al. 2018, Kable, Srisengfa et al. 2019, Carafa, Navarro et al. 2020). In the current study, seasonality had a significant influence on the raw milk microbiota, accounting for 10% of the variation in microbiota composition, with significant effects of time also found in culture-based analysis of the same set of samples done in the concurrent study by Gleeson and colleagues. In Ireland, as milk production is based on seasonal calving, the collection dates corresponded to spring/early-lactation (Apr), summer/mid-lactation (Aug), and winter/late-lactation (Nov). Samples taken in Nov had the greatest diversity and a distinct microbiota composition compared to samples in Apr and Aug, corresponding to results previously reported (Doyle, Gleeson et al. 2017). Although other studies completed in the US and China noted a higher alpha diversity in spring/summer, this could be due to different farming practices, most notably Ireland's predominantly pasture-based system and the diverse nature of the raw milk microbiota (Kable, Srisengfa et

al. 2016, Li, Wang et al. 2018). Despite this, some similarities in specific taxa were evident, such as the increased abundance of species from the family *Lachnospiraceae* and phylum *Actinobacteria* in winter (Kable, Srisengfa et al. 2016, Doyle, Gleeson et al. 2017). Additionally, the species identified from indicator species analysis that were associated with Nov samples were mainly host-related species, which reflect Irish farming practices where cows are kept indoors in winter, which had been reported to impact the raw milk microbiota (Doyle, Gleeson et al. 2017). Functional genes related to the nutrient composition of milk were found to gradually increase (protein metabolism) or decrease (carbohydrates and fatty acids), which indicated the influence of seasonality or lactation stage on milk microbiota and subsequently nutrient composition.

Together with differences across sampling months, differences in the raw milk microbiota were found across sampling locations, which was expected because geographical location and associated environmental and farming practices have previously been shown to influence milk microbiota profiles (Oliver, Jayarao et al. 2005, Breitenwieser, Doll et al. 2020). In this study, analysis revealed that sampling location had a significant influence on the bulk tank raw milk microbiota (Table 2.1). As evidenced in the results, the microbiota of raw milk from locations B and D were more similar to each other compared to the other locations, though no definite correlation to climatic data was found (Figure 2.2 and Supplemental Figure 2.3). More information is required to better ascertain the influence of location on the milk microbiota. Farm practices, environment conditions, animal health and diet

are some factors that previous studies have found to be contributing factors (Doyle, Gleeson et al. 2017, Li, Wang et al. 2018).

Besides taxonomic profiling which adds support to previous culture-based studies, sequencing-based methods allow for further characterisation of the raw milk microbiota. Access to shotgun metagenomic sequencing facilitated functional profiling, with significant differences again being evident between sampling months and locations and not between cleaning methods. In terms of antibiotic resistance, results in this study show that milk is a reservoir of antibiotic resistance genes, which may proliferate in milk and be transferred to surfaces during transport or downstream processing. The most abundant ARGs detected in the milk samples were those conferring resistance to aminoglycosides and tetracycline, which have been previously reported in raw milk (Liu, Zhu et al. 2020, Tóth, Csabai et al. 2020). Notably, Acharya et al. (2020) found that even after effective disinfection, bacterial cells and their DNA can persist in samples, which could result in potential reservoirs for ARGs. Furthermore, incubation at room temperature for short periods of time have been found to enrich ARGs in milk, which highlights the importance of proper cold chain practices (Liu, Zhu et al. 2020). It has also been reported that chlorination can enrich ARGs in drinking water (Shi, Jia et al. 2013, Jia, Shi et al. 2015) and has been found to promote the horizontal transfer of plasmids by natural transformation, which result in the exchange of ARGs across bacterial genera (Jin, Liu et al. 2020). These phenomena may explain why numerically more ARGs were detected in raw milk from farms that used chlorine in their cleaning routine. Additionally, while more research is needed for a deeper understanding on the

resistome of bulk tank milk and the implications of the presence and contribution of ARGs to antimicrobial resistance along the dairy processing chain, this study demonstrates the use of sequencing as a potential tool for monitoring ARGs without isolation and culture of bacteria, which is valuable in terms of food safety.

It should be noted that while this study provides a deeper understanding of the raw milk microbiota and its relationship with cleaning and other environmental or climatic factors, it has some limitations. First, sampling was performed on single days in each month, and so we cannot conclusively state that the patterns are representative of broader seasonal differences, though the patterns are consistent with previous studies. In addition, although shotgun metagenomic sequencing is a valuable tool in the characterisation of the raw milk microbiota, it does not differentiate between viable and non-viable microorganisms. This issue, however, is of greater concern when dealing with samples with high levels of intact DNA from dead bacteria, such as recently pasteurised milk. Finally, the shotgun approach involves sequencing all DNA present in samples, which has resulted in the identification of high proportions of host DNA in each sample and low microbiological sequencing depth, which may have hindered the further resolution of taxonomic, functional and antibiotic resistance profiling, as previously shown in other studies (Pereira-Marques, Hout et al. 2019, Rubiola, Chiesa et al. 2020). Since the conclusion of this study, we have investigated approaches to overcome this problem (Yap, Feehily et al. 2020) that we will utilise in future studies.

Despite these challenges, the study clearly reveals that season and location have a greater impact on the bulk tank raw milk microbiota than the use of chlorine or chlorine-free cleaning methods. Our findings also show that chlorine-free cleaning methods are comparable with chlorine cleaning with respect to the net microbiota composition and functional potential and thus provide further reassurance when considering chlorine-free approaches. Furthermore, this study shows the value that shotgun metagenomic sequencing adds in advancing our understanding of the raw milk microbiota.

2.6 Declarations

2.6.1 Data availability

Sequence data generated during the current study have been deposited in the European Nucleotide Archive under the accession number PRJEB42046.

2.6.2 Acknowledgements including funding source

Research was conducted with the financial support of the Irish Dairy Levy. The authors thank Lizandra Paludetti for coordinating the collection of samples, Aoife McHugh, Laura Wosinska, Katie Falà and Amy Fitzpatrick for help with sample processing, Calum Walsh for help with bioinformatic analysis, and Fiona Crispie and Amanda Bréchon for their help with DNA sequencing.

2.7 References

Acharya, K., F. F. Halla, S. M. Massawa, S. M. Mgana, T. Komar, R. J. Davenport and D. Werner (2020). "Chlorination effects on DNA based characterization of water microbiomes and implications for the interpretation of data from disinfected systems." Journal of Environmental Management 276: 111319.

Beck, M. (2017). ggord: ordination plots with ggplot2. R package version 1.0. 0.

Breitenwieser, F., E. V. Doll, T. Clavel, S. Scherer and M. Wenning (2020). "Complementary Use of Cultivation and High-Throughput Amplicon Sequencing Reveals High Biodiversity Within Raw Milk Microbiota." Frontiers in microbiology 11: 1557.

Buchfink, B., C. Xie and D. H. Huson (2015). "Fast and sensitive protein alignment using DIAMOND." Nature methods 12(1): 59-60.

Cáceres, M. D. and P. Legendre (2009). "Associations between species and groups of sites: indices and statistical inference." Ecology 90(12): 3566-3574.

Carafa, I., I. C. Navarro, G. Bittante, F. Tagliapietra, L. Gallo, K. Tuohy and E. Franciosi (2020). "Shift in the cow milk microbiota during alpine pasture as analyzed by culture dependent and high-throughput sequencing techniques." Food Microbiology: 103504.

Doyle, C. J., D. Gleeson, P. W. O'Toole and P. D. Cotter (2017). "High-throughput metataxonomic characterization of the raw milk microbiota identifies changes reflecting lactation stage and storage conditions." International journal of food microbiology 255: 1-6.

Doyle, C. J., D. Gleeson, P. W. O'Toole and P. D. Cotter (2017). "Impacts of seasonal housing and teat preparation on raw milk microbiota: a high-throughput sequencing study." Applied and environmental microbiology 83(2).

Du, B., L. Meng, H. Liu, N. Zheng, Y. Zhang, X. Guo, S. Zhao, F. Li and J. Wang (2020). "Impacts of Milking and Housing Environment on Milk Microbiota." Animals 10(12): 2339.

EFSA Panel on Contaminants in the Food Chain (2015). "Risks for public health related to the presence of chlorate in food." EFSA Journal 13(6): 4135.

Franzosa, E. A., T. Hsu, A. Sirota-Madi, A. Shafquat, G. Abu-Ali, X. C. Morgan and C. Huttenhower (2015). "Sequencing and beyond: integrating molecular'omics' for microbial community profiling." Nature Reviews Microbiology 13(6): 360-372.

Gleeson, D., B. O'Brien and K. Jordan (2013). "The effect of using nonchlorine products for cleaning and sanitising milking equipment on bacterial numbers and residues in milk." International Journal of Dairy Technology 66(2): 182-188.

Gómez-López, V. M., A. Rajkovic, P. Ragaert, N. Smigic and F. Devlieghere (2009). "Chlorine dioxide for minimally processed produce preservation: a review." Trends in Food Science & Technology 20(1): 17-26.

Jia, B., A. R. Raphenya, B. Alcock, N. Waglechner, P. Guo, K. K. Tsang, B. A. Lago, B. M. Dave, S. Pereira and A. N. Sharma (2016). "CARD 2017: expansion and model-centric curation of the comprehensive antibiotic resistance database." Nucleic acids research: gkw1004.

Jia, S., P. Shi, Q. Hu, B. Li, T. Zhang and X.-X. Zhang (2015). "Bacterial community shift drives antibiotic resistance promotion during drinking water chlorination." Environmental Science & Technology 49(20): 12271-12279.

Jin, M., L. Liu, D.-n. Wang, D. Yang, W.-l. Liu, J. Yin, Z.-w. Yang, H.-r. Wang, Z.-g. Qiu and Z.-q. Shen (2020). "Chlorine disinfection promotes the exchange of antibiotic resistance genes across bacterial genera by natural transformation." The ISME Journal: 1-10.

Kable, M. E., Y. Srisengfa, M. Laird, J. Zaragoza, J. McLeod, J. Heidenreich and M. L. Marco (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(4).

Kable, M. E., Y. Srisengfa, Z. Xue, L. C. Coates and M. L. Marco (2019). "Viable and total bacterial populations undergo equipment-and time-dependent shifts during milk processing." Applied and environmental microbiology 85(13).

Kaminski, J., M. K. Gibson, E. A. Franzosa, N. Segata, G. Dantas and C. Huttenhower (2015). "High-specificity targeted functional profiling in microbial communities with ShortBRED." PLoS computational biology 11(12): e1004557.

Kassambara, A. (2020). ggpubr: 'ggplot2' based publication ready plots. R package version 0.4.0.

Li, N., Y. Wang, C. You, J. Ren, W. Chen, H. Zheng and Z. Liu (2018). "Variation in raw milk microbiota throughout 12 months and the impact of weather conditions." Scientific reports 8(1): 1-10.

Liu, J., Y. Zhu, M. Jay-Russell, D. G. Lemay and D. A. Mills (2020). "Reservoirs of antimicrobial resistance genes in retail raw milk." Microbiome 8(1): 1-15.

McCarthy, W. P., T. F. O'Callaghan, M. Danahar, D. Gleeson, C. O'Connor, M. A. Fenelon and J. T. Tobin (2018). "Chlorate and other oxychlorine contaminants within the dairy supply chain." Comprehensive Reviews in Food Science and Food Safety 17(6): 1561-1575.

McHugh, A. J., C. Feehily, M. A. Fenelon, D. Gleeson, C. Hill and P. D. Cotter (2020). "Tracking the dairy microbiota from farm bulk tank to skimmed milk powder." Msystems 5(2).

Oksanen, J., F. Blanchet, M. Friendly, R. Kindt, P. Legendre, D. McGlinn, P. Minchin, R. O'Hara, G. Simpson and P. Solymos (2019). vegan: Community Ecology Package. R package version 2.5-6. 2019.

Oliver, S. P., B. M. Jayarao and R. A. Almeida (2005). "Foodborne pathogens in milk and the dairy farm environment: food safety and public health implications." Foodborne Pathogens and Disease 2(2): 115-129.

Ostrov, I., A. Harel, S. Bernstein, D. Steinberg and M. Shemesh (2016). "Development of a method to determine the effectiveness of cleaning agents in removal of biofilm derived spores in milking system." Frontiers in microbiology 7: 1498.

Overbeek, R., T. Begley, R. M. Butler, J. V. Choudhuri, H.-Y. Chuang, M. Cohoon, V. de Crécy-Lagard, N. Diaz, T. Disz and R. Edwards (2005). "The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes." Nucleic acids research 33(17): 5691-5702.

Pandey, N., A. Kumari, A. Varma, S. Sahu and M. Akbar (2014). "Impact of applying hygienic practices at farm on bacteriological quality of raw milk." Vet World 7(9).

Parente, E., A. Ricciardi and T. Zotta (2020). "The microbiota of dairy milk: a review." International Dairy Journal: 104714.

Parks, D. H., M. Chuvochina, P.-A. Chaumeil, C. Rinke, A. J. Mussig and P. Hugenholtz (2020). "A complete domain-to-species taxonomy for Bacteria and Archaea." Nature Biotechnology: 1-8.

Parks, D. H., M. Chuvochina, D. W. Waite, C. Rinke, A. Skarshewski, P.-A. Chaumeil and P. Hugenholtz (2018). "A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life." Nature biotechnology 36(10): 996-1004.

Pereira-Marques, J., A. Hout, R. M. Ferreira, M. Weber, I. Pinto-Ribeiro, L.-J. van Doorn, C. W. Knetsch and C. Figueiredo (2019). "Impact of host DNA and sequencing depth on the taxonomic resolution of whole metagenome sequencing for microbiome analysis." Frontiers in microbiology 10: 1277.

Porcellato, D., M. Aspholm, S. B. Skeie, M. Monshaugen, J. Brendehaug and H. Mellegård (2018). "Microbial diversity of consumption milk during processing and storage." International journal of food microbiology 266: 21-30.

Quigley, L., O. O'Sullivan, C. Stanton, T. P. Beresford, R. P. Ross, G. F. Fitzgerald and P. D. Cotter (2013). "The complex microbiota of raw milk." FEMS microbiology reviews 37(5): 664-698.

R Core Team (2013). R: A language and environment for statistical computing, Vienna, Austria.

Reinemann, D. J., G. Wolters, P. Billon, O. Lind and M. D. Rasmussen (2003). "Review of practices for cleaning and sanitation of milking machines." Bulletin-International Dairy Federation: 3-18.

Rubiola, S., F. Chiesa, A. Dalmasso, P. Di Ciccio and T. Civera (2020). "Detection of Antimicrobial Resistance Genes in the Milk Production Environment: Impact of Host DNA and Sequencing Depth." Frontiers in Microbiology 11: 1983.

Shi, P., S. Jia, X.-X. Zhang, T. Zhang, S. Cheng and A. Li (2013). "Metagenomic insights into chlorination effects on microbial antibiotic resistance in drinking water." Water research 47(1): 111-120.

Silaghi, C., M. Nieder, C. Sauter-Louis, G. Knubben-Schweizer, K. Pfister and M. Pfeffer (2018). "Epidemiology, genetic variants and clinical course of natural infections with Anaplasma phagocytophilum in a dairy cattle herd." Parasites & vectors 11(1): 20.

Silva, G. G. Z., K. T. Green, B. E. Dutilh and R. A. Edwards (2016). "SUPER-FOCUS: a tool for agile functional analysis of shotgun metagenomic data." Bioinformatics 32(3): 354-361.

Sundberg, M., A. Christiansson, C. Lindahl, L. Wahlund and C. Birgersson (2011). "Cleaning effectiveness of chlorine-free detergents for use on dairy farms." Journal of dairy research 78(1): 105.

Taponen, S., D. McGuinness, H. Hiitiö, H. Simojoki, R. Zadoks and S. Pyörälä (2019). "Bovine milk microbiome: a more complex issue than expected." Veterinary research 50(1): 44.

Tiefel, P. and K. Guthy (1997). "Model tests for the formation of chloroform by chlorine containing cleaning and disinfection products." Milchwissenschaft 52(12): 686-691.

Tóth, A. G., I. Csabai, E. Krikó, D. Tőzsér, G. Maróti, Á. V. Patai, L. Makrai, G. Szita and N. Solymosi (2020). "Antimicrobial resistance genes in raw milk for human consumption." Scientific reports 10(1): 1-7.

Virto, R., P. Manas, I. Alvarez, S. Condon and J. Raso (2005). "Membrane damage and microbial inactivation by chlorine in the absence and presence of a chlorine-demanding substrate." Applied and environmental microbiology 71(9): 5022-5028.

Wickham, H., M. Averick, J. Bryan, W. Chang, L. D. A. McGowan, R. François, G. Grolemund, A. Hayes, L. Henry and J. Hester (2019). "Welcome to the Tidyverse." Journal of Open Source Software 4(43): 1686.

Wood, D. E., J. Lu and B. Langmead (2019). "Improved metagenomic analysis with Kraken 2." Genome biology 20(1): 257.

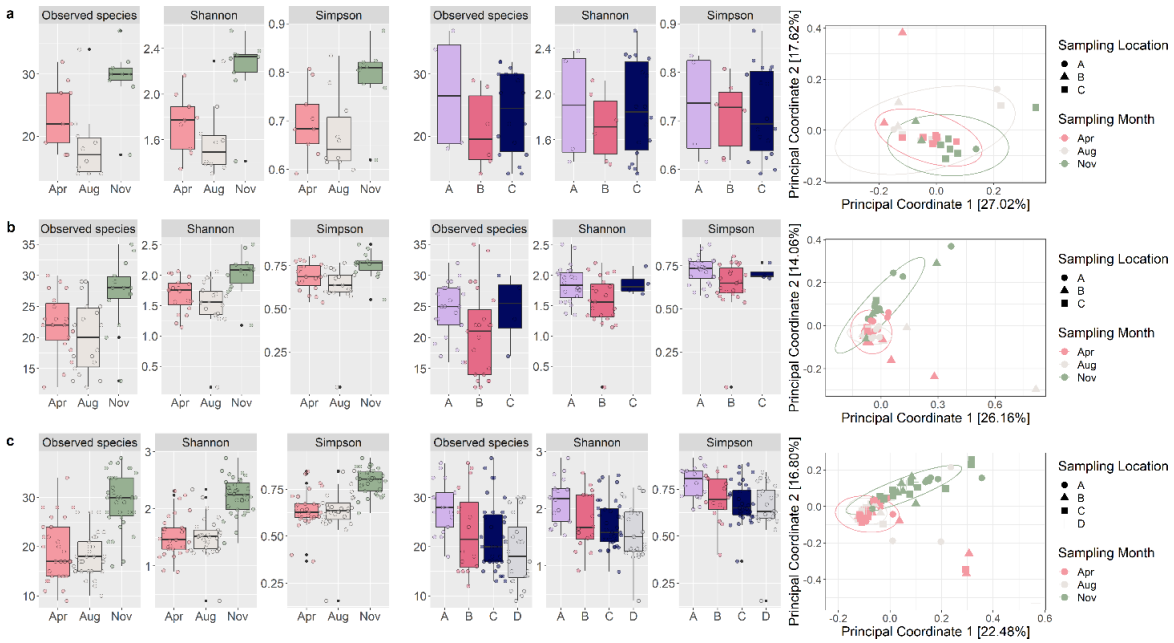
Yap, M., C. Feehily, C. J. Walsh, M. A. Fenelon, E. F. Murphy, F. M. McAuliffe, D. van Sinderen, P. W. O'Toole, O. O'Sullivan and P. D. Cotter (2020). "Evaluation of methods for the reduction of contaminating host reads when performing shotgun metagenomic sequencing of the milk microbiome." Scientific Reports 10.

Zhang, Y., Y. Lv, Y. Cui, J. Wang, S. Cao, F. Jian, R. Wang, L. Zhang and C. Ning (2016). "First molecular evidence for the presence of Anaplasma DNA in milk from sheep and goats in China." Parasitology research 115(7): 2789-2795.

2.8 Supplemental Material

Supplemental Table 2.1. Sample numbers for each cleaning method according to sampling location (A to D), where samples are from farms that used chlorine-free agents for their bulk tank alone (BTCF), farms that used chlorine-based agents in their cleaning routine (C) and farms that used a chlorine-free cleaning routine (CF).

Cleaning Methods	Sampling Location				
	A	B	C	D	Total
C	24	23	4		51
CF	17	16	35	24	92
BTCF	4	6	18		28
Total	45	45	57	24	171

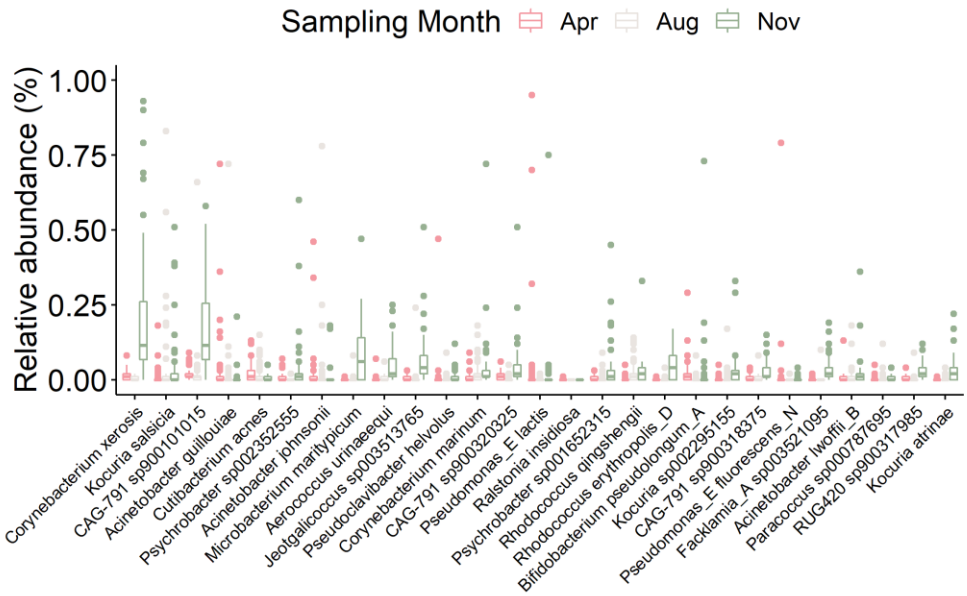


Supplemental Figure 2.1. Alpha and beta diversity of raw milk stratified by cleaning method. Samples are from (a) farms that used chlorine-free agents for their bulk tank alone (BTCF), (b) farms that used chlorine-based agents in their cleaning routine (C) and (c) farms that used a chlorine-free cleaning routine (CF). Samples are from farms from 3 different sampling months – April (Apr), August (Aug) and November (Nov) from 4 different sampling locations (A to D).

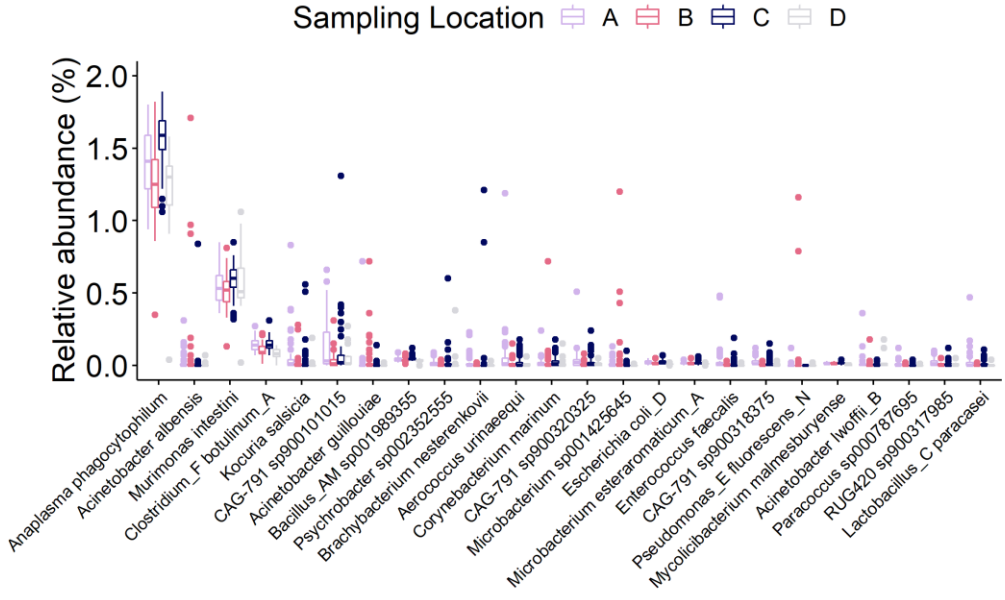
Supplemental Table 2.2. PERMANOVA analysis of taxonomic composition and functional properties (based on results from SUPER-FOCUS) of the bulk tank milk microbiota stratified by cleaning method. Samples are from farms that used chlorine-free agents for their bulk tank alone (BTCF), farms that used chlorine-based agents in their cleaning routine (C) and farms that used a chlorine-free cleaning routine (CF).

Cleaning Method		**R^2**	**P value**
BTCF	Sampling month	0.167	0.001
	Sampling location	0.150	0.005
	Residuals	0.684	
C	Sampling month	0.099	0.001
	Sampling location	0.047	0.201
	Residuals	0.853	
CF	Sampling month	0.124	0.001
	Sampling location	0.099	0.001
	Residuals	0.777	

a



b



Supplemental Figure 2.2. Significantly differential relative abundances of species that differed between (a) sampling month – April (Apr), August (Aug) and November (Nov), and (b) sampling location (locations A to D).

Supplemental Table 2.3. Relative abundances of species that differed by sampling month – April (Apr), August (Aug) and November (Nov).

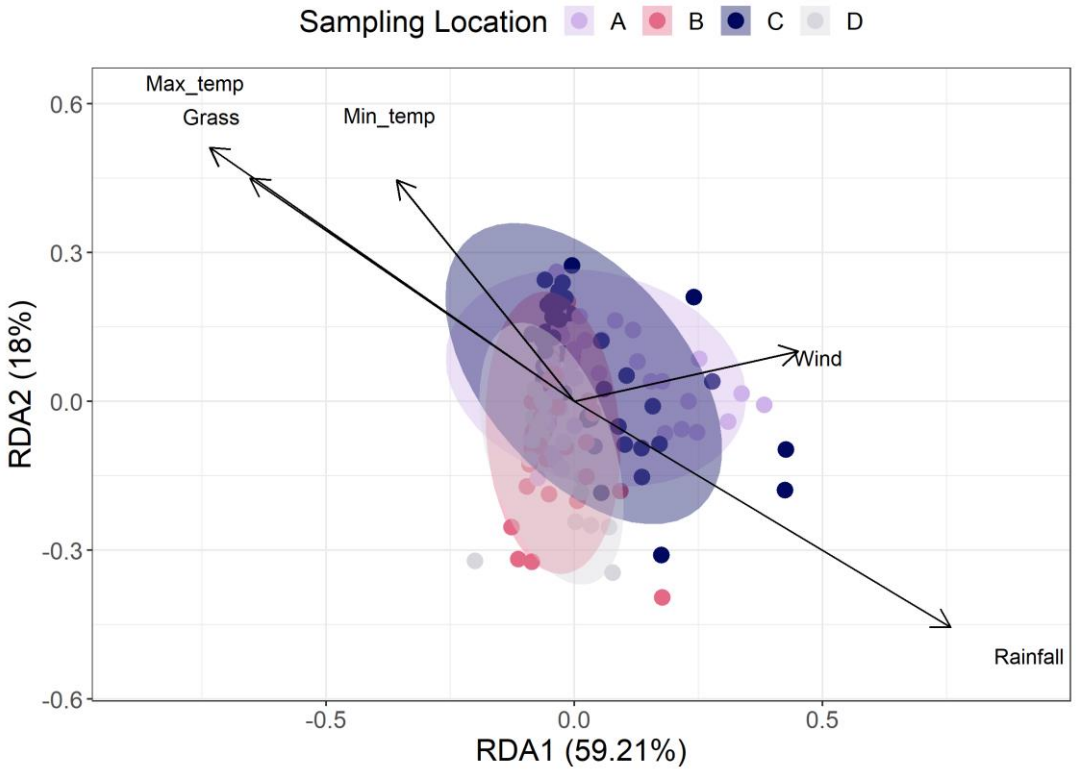
Species	Sampling Month		
	Apr	Aug	Nov
Acinetobacter guillouiae	0.188	0.020	0.005
Acinetobacter johnsonii	0.020	0.071	0.008
Acinetobacter lwoffii_B	0.005	0.012	0.022
Aerococcus urinaeequi	0.024	0.004	0.054
Bifidobacterium pseudolongum_A	0.021	0.002	0.021
CAG-791 sp900101015	0.021	0.023	0.195
CAG-791 sp900318375	0.006	0.005	0.030
CAG-791 sp900320325	0.009	0.003	0.048
Corynebacterium marinum	0.009	0.022	0.036
Corynebacterium xerosis	0.013	0.023	0.228
Cutibacterium acnes	0.020	0.017	0.005
Facklamia_A sp003521095	0.001	0.002	0.035
Jeotgalicoccus sp003513765	0.003	0.005	0.066
Kocuria atrinae	0.001	0.001	0.030
Kocuria salsicia	0.008	0.108	0.123
Kocuria sp002295155	0.002	0.006	0.031
Microbacterium maritypicum	0.000	0.002	0.084
Paracoccus sp000787695	0.002	0.004	0.004
Pseudoclavibacter helvolus	0.011	0.003	0.010
Pseudomonas_E fluorescens_N	0.038	0.000	0.004
Pseudomonas_E lactis	0.038	0.001	0.016
Psychrobacter sp001652315	0.004	0.004	0.038
Psychrobacter sp002352555	0.009	0.000	0.088
Ralstonia insidiosa	0.001	0.000	0.000
Rhodococcus erythropolis_D	0.000	0.001	0.046
Rhodococcus qingshengii	0.002	0.021	0.025
RUG420 sp900317985	0.004	0.002	0.029

Supplemental Table 2.4. Relative abundances of species that differed by sampling location (locations A to D).

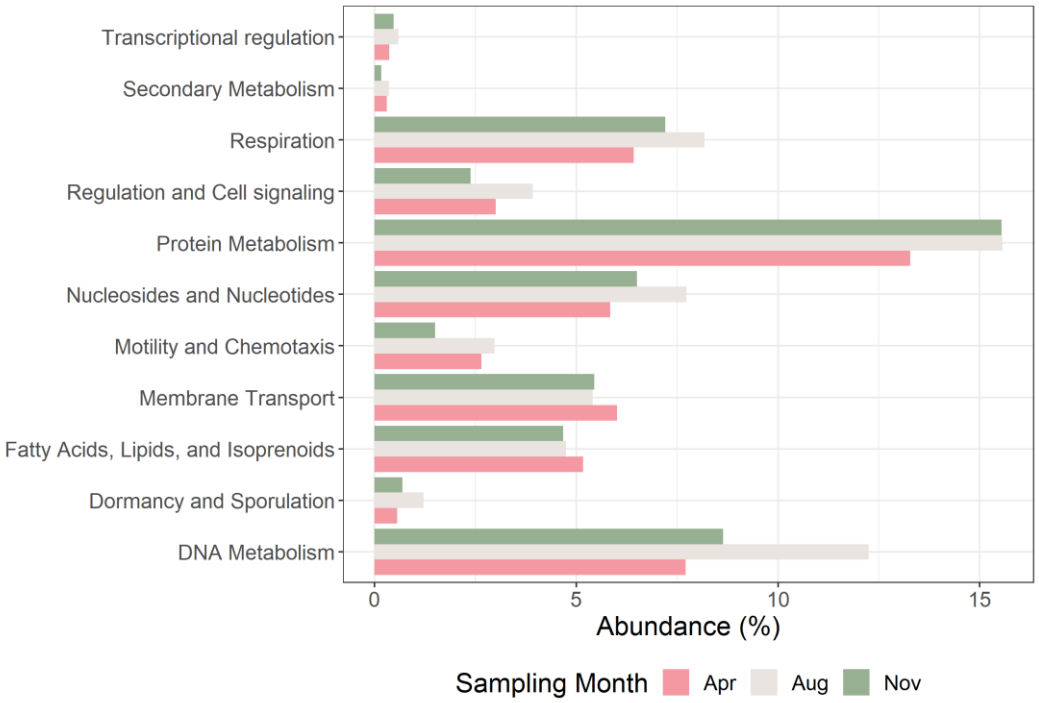
Species	Sampling Location			
	A	**B**	**C**	**D**
Acinetobacter albensis	0.025	0.270	0.016	0.213
Acinetobacter guillouiae	0.020	0.242	0.005	0.001
Acinetobacter lwoffii_B	0.020	0.008	0.009	0.018
Aerococcus urinaeequi	0.062	0.012	0.021	0.008
Anaplasma phagocytophilum	1.414	1.252	1.561	1.208
Bacillus_AM sp001989355	0.042	0.034	0.048	0.032
Brachybacterium nesterenkovii	0.074	0.002	0.038	0.001
CAG-791 sp900101015	0.139	0.032	0.086	0.045
CAG-791 sp900318375	0.024	0.005	0.016	0.005
CAG-791 sp900320325	0.037	0.008	0.021	0.010
Clostridium_F botulinum_A	0.145	0.106	0.147	0.079
Corynebacterium marinum	0.018	0.028	0.025	0.012
Enterococcus faecalis	0.036	0.004	0.008	0.001
Escherichia coli_D	0.021	0.015	0.020	0.009
Kocuria salsicia	0.243	0.018	0.029	0.010
Lactobacillus_C paracasei	0.031	0.001	0.008	0.002
Microbacterium esteraromaticum_A	0.017	0.015	0.021	0.011
Microbacterium sp001425645	0.010	0.058	0.004	0.001
Murimonas intestini	0.544	0.517	0.591	0.585
Mycolicibacterium malmesburyense	0.014	0.011	0.016	0.007
Paracoccus sp000787695	0.008	0.001	0.003	0.001
Pseudomonas_E fluorescens_N	0.004	0.047	0.000	0.002
Psychrobacter sp002352555	0.013	0.073	0.021	0.019
RUG420 sp900317985	0.022	0.004	0.012	0.006

Supplemental Table 2.5. Climate data retrieved from the Irish Meteorological Service from the sampling days at the four sampling locations. Samples are from farms from 3 different sampling months – April (Apr), August (Aug) and November (Nov) from 4 different sampling locations (A to D).

Sampling Month	Location	Rainfall (mm)	Max Temp (^{o}C)	Min Temp (^{o}C)	Grass Min Temp (^{o}C)	Mean Wind Speed (knots)
Apr	A	2.5	13.1	9.9	8.1	11
	B	3.9	12.6	7.3	5	10.6
	C	7.2	14.2	7.5	4.3	6
	D	3.7	13.4	7.6	6.5	7.2
Aug	A	0.1	17.4	10.7	4.8	6.8
	B	0	18.6	10.5	6.4	6.1
	C	0	21.9	8.5	3.9	4.4
	D	0	20.3	7.8	5	3.5
Nov	A	8.2	9.6	3.7	3.3	18.6
	B	15.2	6.0	2.2	1	9.8
	C	18.4	5.7	0.1	-1.6	4.4
	D	15.8	5.2	0.6	-2.5	6.5



Supplemental Figure 2.3. Redundancy analysis of the relationship between climate factors and sampling locations of the bulk tank milk microbiota. The variables are described in Supplemental Table 2.5. Ellipses represent clustering by the 4 different sampling locations (A to D).



Supplemental Figure 2.4. Significantly differential abundances of functions that differed between the sampling months, April (Apr), August (Aug) and November (Nov).

Supplemental Table 2.6. Relative abundances of genes assigned to significant functional classifications of SUPER-FOCUS subsystem levels 1-3 between sampling months – April (Apr), August (Aug) and November (Nov).

Subsystem Level 1	Subsystem Level 2	Subsystem Level 3	Sampling Month		
			Apr	Aug	Nov
Amino Acids and Derivatives	Alanine, serine, and glycine	-	0.658	0.764	0.814
Amino Acids and Derivatives	Alanine, serine, and glycine	Glycine and Serine Utilisation	0.258	0.336	0.338
Amino Acids and Derivatives	Alanine, serine, and glycine	Glycine cleavage system	0.085	0.090	0.130
Amino Acids and Derivatives	Arginine; urea cycle, polyamines	Urea decomposition	0.155	0.214	0.097
Amino Acids and Derivatives	Branched-chain amino acids	-	1.537	1.461	1.242
Amino Acids and Derivatives	Lysine, threonine, methionine, and cysteine	Methionine Degradation	0.335	0.322	0.443
Carbohydrates	-	-	20.429	19.438	19.236
Carbohydrates	-	Putative sugar ABC transporter (ytf cluster)	0.049	0.017	0.019
Carbohydrates	Central carbohydrate metabolism	Dehydrogenase complexes	0.323	0.263	0.232
Carbohydrates	Central carbohydrate metabolism	Glycolysis and Gluconeogenesis	0.417	0.463	0.603
Carbohydrates	Central carbohydrate metabolism	Pyruvate:ferredoxin oxidoreductase	0.110	0.131	0.175
Carbohydrates	CO2 fixation	-	0.473	0.407	0.331

Carbohydrates	CO2 fixation	Photorespiration (oxidative C2 cycle)	0.104	0.086	0.049
Carbohydrates	Di- and oligosaccharides	Fructooligosaccharides (FOS) and Raffinose Utilisation	0.237	0.209	0.297
Carbohydrates	Di- and oligosaccharides	Lactose utilisation	0.134	0.156	0.190
Carbohydrates	Fermentation	Acetone Butanol Ethanol Synthesis	0.002	0.001	0.003
Carbohydrates	Monosaccharides	Deoxyribose and Deoxynucleoside Catabolism	0.213	0.194	0.276
Carbohydrates	One-carbon Metabolism	millsd methanogensis	0.007	0.009	0.022
Carbohydrates	Organic acids	Glycerate metabolism	0.111	0.083	0.109
Carbohydrates	Polysaccharides	-	0.645	0.794	0.833
Carbohydrates	Sugar alcohols	Glycerol fermentation to 1,3-propanediol	0.011	0.009	0.032
Carbohydrates	Sugar alcohols	Propanediol utilisation	0.019	0.011	0.024
Cell Division and Cell Cycle	-	Control of cell elongation - division cycle in Bacilli	0.057	0.031	0.036
Cell Wall and Capsule	Capsular and extracellular polysacchrides	CMP-N-acetylneuraminate Biosynthesis	0.027	0.021	0.037
Cell Wall and Capsule	Capsular and extracellular polysacchrides	Phosphorylcholine incorporation in LPS	0.006	0.007	0.012
Cell Wall and Capsule	Capsular and extracellular polysacchrides	Polysaccharide deacetylases	0.016	0.021	0.035

Cell Wall and Capsule	Gram-Positive cell wall components	D-Alanyl Lipoteichoic Acid Biosynthesis	0.012	0.004	0.004
Cell Wall and Capsule	Gram-Positive cell wall components	Teichuronic acid biosynthesis	0.017	0.018	0.041
Clustering-based subsystems	-	CBSS-262719.3.peg.410	0.075	0.049	0.104
Clustering-based subsystems	-	LMPTP YwlE cluster	0.080	0.080	0.116
Clustering-based subsystems	-	Sporulation-related Hypotheticals	0.014	0.013	0.022
Clustering-based subsystems	contains Thr-tRNA-syn, pyridoxine biosyn, lipid A biosyn, 3 hypos	-	0.007	0.020	0.038
Clustering-based subsystems	contains Thr-tRNA-syn, pyridoxine biosyn, lipid A biosyn, 3 hypos	CBSS-1806.1.peg.1285	0.007	0.020	0.038
Clustering-based subsystems	Protein export?	-	0.253	0.221	0.334
Clustering-based subsystems	Protein export?	CBSS-393121.3.peg.2760	0.253	0.221	0.334
Clustering-based subsystems	Putative Isoquinoline 1-oxidoreductase subunit	-	0.037	0.014	0.017
Clustering-based subsystems	Putative Isoquinoline 1-oxidoreductase subunit	CBSS-314267.3.peg.390	0.037	0.014	0.017
Clustering-based subsystems	Putrescine/GABA utilisation cluster-temporal, to add to SSs	-	0.050	0.037	0.019

Clustering-based subsystems	Putrescine/GABA utilisation cluster-temporal, to add to SSs	GABA and putrescine metabolism from clusters	0.050	0.037	0.019
Clustering-based subsystems	recX and regulatory cluster	-	0.012	0.003	0.013
Clustering-based subsystems	Sulfatases and sulfatase modifying factor 1 (and a hypothetical)	-	0.110	0.049	0.056
Clustering-based subsystems	Sulfatases and sulfatase modifying factor 1 (and a hypothetical)	Sulfatases and sulfatase modifying factor 1	0.110	0.049	0.056
Cofactors, Vitamins, Prosthetic Groups, Pigments	Coenzyme M	-	0.009	0.003	0.012
Cofactors, Vitamins, Prosthetic Groups, Pigments	Folate and pterines	p-Aminobenzoyl-Glutamate Utilisation	0.027	0.013	0.038
Cofactors, Vitamins, Prosthetic Groups, Pigments	Riboflavin, FMN, FAD	Flavodoxin	0.035	0.013	0.034
Cofactors, Vitamins, Prosthetic Groups, Pigments	Tetrapyrroles	Cobalamin	0.107	0.077	0.129
DNA Metabolism	-	-	5.762	6.259	6.506
DNA Metabolism	DNA repair	-	3.087	3.426	3.502
DNA Metabolism	DNA repair	DNA repair, UvrABC system	0.663	0.695	0.853

DNA Metabolism	DNA replication	-	1.561	1.672	1.805
DNA Metabolism	DNA replication	DNA replication, archaeal	0.050	0.046	0.067
DNA Metabolism	DNA replication	DNA topoisomerases, Type II, ATP-dependent	0.292	0.355	0.408
Dormancy and Sporulation	-	-	0.212	0.241	0.279
Dormancy and Sporulation	-	Spore Core Dehydration	0.015	0.008	0.026
Dormancy and Sporulation	-	Sporulation Cluster	0.026	0.029	0.037
Fatty Acids, Lipids, and Isoprenoids	-	-	3.329	3.158	2.930
Fatty Acids, Lipids, and Isoprenoids	Fatty acids	-	1.744	1.536	1.356
Fatty Acids, Lipids, and Isoprenoids	Fatty acids	Fatty Acid Biosynthesis FASII	1.203	1.057	0.936
Fatty Acids, Lipids, and Isoprenoids	Isoprenoids	Isoprenoid Biosynthesis	0.176	0.203	0.230
Fatty Acids, Lipids, and Isoprenoids	Isoprenoids	Isoprenoid Biosynthesis: Interconversions	0.018	0.010	0.030
Iron acquisition and metabolism	-	Heme, hemin uptake and utilisation systems in Gram Positives	0.113	0.094	0.154
Iron acquisition and metabolism	Siderophores	Petrobactin-mediated iron uptake system	0.018	0.018	0.052
Iron acquisition and metabolism	Siderophores	Siderophore Pyoverdine	0.123	0.057	0.042

Membrane Transport	-	-	3.665	3.410	3.105
Membrane Transport	Sugar Phosphotransferase Systems, PTS	-	0.033	0.032	0.025
Membrane Transport	-	Transport of Manganese	0.106	0.071	0.035
Metabolism of Aromatic Compounds	Metabolism of central aromatic intermediates	N-heterocyclic aromatic compound degradation	0.035	0.014	0.016
Miscellaneous	-	EC 2.7.1.- Phosphotransferases with an alcohol group as acceptor	0.043	0.024	0.030
Miscellaneous	-	EC 6.3.4.- Ligases that form carbon-nitrogen bonds	0.265	0.216	0.275
Miscellaneous	-	Nicotinate catabolism, anaerobic	0.005	0.002	0.012
Miscellaneous	Plant-Prokaryote comparative genomics	COG0277	0.039	0.020	0.021
Miscellaneous	Plant-Prokaryote comparative genomics	Omega-amidase KE2	0.608	0.609	0.743
Motility and Chemotaxis	-	-	0.867	0.762	0.632
Motility and Chemotaxis	Flagellar motility in Prokaryota	-	0.442	0.389	0.303
Motility and Chemotaxis	Flagellar motility in Prokaryota	Flagellum	0.255	0.218	0.170
Nitrogen Metabolism	-	Allantoin Utilisation	0.039	0.023	0.032

Nitrogen Metabolism	-	Ammonia assimilation	0.442	0.424	0.344
Nitrogen Metabolism	-	Nitrosative stress	0.041	0.053	0.068
Nucleosides and Nucleotides	-	-	3.820	4.248	4.509
Nucleosides and Nucleotides	-	Pseudouridine Metabolism	0.013	0.012	0.030
Nucleosides and Nucleotides	Purines	-	2.042	2.304	2.469
Nucleosides and Nucleotides	Purines	De Novo Purine Biosynthesis	0.793	0.904	1.055
Nucleosides and Nucleotides	Pyrimidines	-	1.113	1.252	1.281
Nucleosides and Nucleotides	Pyrimidines	De Novo Pyrimidine Synthesis	0.644	0.768	0.824
Phages, Prophages, Transposable elements, Plasmids	Plasmid related functions	-	0.022	0.015	0.024
Phages, Prophages, Transposable elements, Plasmids	Plasmid related functions	Rolling-circle replication	0.021	0.014	0.023
Phosphorus Metabolism	-	Phosphonate metabolism	0.020	0.013	0.014
Protein Metabolism	-	-	9.280	9.651	10.547
Protein Metabolism	Protein biosynthesis	-	6.541	6.849	7.776
Protein Metabolism	Protein biosynthesis	Ribosome LSU bacterial	0.569	0.671	0.754
Protein Metabolism	Protein biosynthesis	Ribosome SSU bacterial	0.349	0.407	0.511
Protein Metabolism	Protein biosynthesis	Single-copy ribosomal proteins	0.579	0.584	0.706

Protein Metabolism	Protein biosynthesis	Translation elongation factor G family	0.155	0.135	0.218
Protein Metabolism	Protein biosynthesis	Translation elongation factors bacterial	0.211	0.213	0.295
Protein Metabolism	Protein biosynthesis	tRNA aminoacylation, Asp and Asn	0.196	0.230	0.261
Protein Metabolism	Protein biosynthesis	Universal GTPases	0.697	0.754	0.900
Protein Metabolism	Protein processing and modification	FeFe hydrogenase maturation	0.017	0.025	0.048
Protein Metabolism	Protein processing and modification	N-linked Glycosylation in Bacteria	0.152	0.166	0.217
Protein Metabolism	Protein processing and modification	SCIFF peptide maturase system	0.010	0.020	0.027
Regulation and Cell signalling	-	-	2.016	1.816	1.628
Regulation and Cell signalling	-	cAMP signalling in bacteria	0.546	0.501	0.355
Regulation and Cell signalling	-	HPr catabolite repression system	0.042	0.039	0.043
Regulation and Cell signalling	-	Sex pheromones in *Enterococcus faecalis* and other Firmicutes	0.107	0.079	0.132
Respiration	-	-	4.500	4.582	4.195
Respiration	-	Biogenesis of cbb3-type cytochrome c oxidases	0.032	0.020	0.031
Respiration	-	Quinone oxidoreductase family	0.092	0.058	0.032

Respiration	ATP synthases	-	0.506	0.577	0.701
Respiration	ATP synthases	V-Type ATP synthase	0.070	0.104	0.190
Respiration	Electron donating reactions	-	1.785	1.836	1.510
Respiration	Electron donating reactions	CO Dehydrogenase	0.213	0.118	0.074
Respiration	Electron donating reactions	Na (+)-translocating NADH-quinone oxidoreductase and rnf-like group of electron transport complexes	0.143	0.166	0.266
Respiration	Electron donating reactions	Respiratory Complex I	0.731	0.793	0.402
Respiration	Reverse electron transport	-	0.010	0.014	0.015
Respiration	Reverse electron transport	Archaeal membrane bound hydrogenases	0.005	0.007	0.008
Respiration	Reverse electron transport	Energy conserving hydrogenase b, Methanococcales-Methanobacteriales-Methanopyrales	0.002	0.001	0.005
Respiration	Sodium Ion-Coupled Energetics	-	0.034	0.075	0.058
Respiration	Sodium Ion-Coupled Energetics	Na+ translocating decarboxylases and related biotin-dependent enzymes	0.027	0.075	0.057
RNA Metabolism	RNA processing and modification	Polyadenylation bacterial	0.113	0.205	0.193
RNA Metabolism	RNA processing and modification	t (6)A synthesis in bacteria	0.014	0.014	0.025

RNA Metabolism	RNA processing and modification	Threonylcarbamoyladenosine	0.017	0.021	0.026
RNA Metabolism	Transcription	RNA polymerase bacterial	0.485	0.572	0.657
Secondary Metabolism	-	-	0.095	0.056	0.052
Secondary Metabolism	Biologically active compounds in metazoan cell defence and differentiation	-	0.045	0.017	0.020
Secondary Metabolism	Biologically active compounds in metazoan cell defence and differentiation	Steroid sulfates	0.045	0.017	0.020
Stress Response	Acid stress	-	0.013	0.000	0.009
Stress Response	Acid stress	Glutamate transporter involved in acid tolerance in Streptococcus	0.013	0.000	0.009
Stress Response	Oxidative stress	Oxygen stress response / Human gut microbiome	0.021	0.029	0.037
Stress Response	Oxidative stress	Redox-dependent regulation of nucleus processes	0.078	0.093	0.120
Sulphur Metabolism	Organic sulphur assimilation	Alkanesulfonate assimilation	0.236	0.162	0.145
Transcriptional regulation	-	-	0.161	0.133	0.188
Transcriptional regulation	-	CarD	0.161	0.133	0.188

Virulence	-	Mycobacterium virulence operon involved in an unknown function with a Jag Protein and YidC and YidD	0.075	0.045	0.077
Virulence	-	Mycobacterium virulence operon involved in protein synthesis (SSU ribosomal proteins)	0.206	0.204	0.280
Virulence	Resistance to antibiotics and toxic compounds	-	3.059	2.906	2.555
Virulence	Resistance to antibiotics and toxic compounds	Cadmium resistance	0.014	0.018	0.031
Virulence	Resistance to antibiotics and toxic compounds	Mycobacterial MmpL3 membrane protein cluster	0.005	0.008	0.013
Virulence	Resistance to antibiotics and toxic compounds	Resistance to fluoroquinolones	0.290	0.350	0.400

Chapter 3. Evaluation of methods for the reduction of contaminating host reads when performing shotgun metagenomic sequencing of the milk microbiome

Included as published in Scientific Reports

(doi: 10.1038/s41598-020-78773-6)

Authors: Min Yap, Conor Feehily, Calum J. Walsh, Mark Fenelon, Eileen F. Murphy, Fionnuala M. McAuliffe, Douwe van Sinderen, Paul W. O'Toole, Orla O'Sullivan and Paul D. Cotter

Contributions:

Candidate performed sample preparation, DNA extraction, sequencing library preparation, bioinformatic and statistical analysis.

Candidate drafted and edited chapter.

CF performed some DNA extractions, sequencing library preparation, bioinformatic and statistical analysis and edited chapter.

CJW performed some bioinformatic and statistical analysis.

All authors read, revised, and approved manuscript.

3.1 Abstract

Shotgun metagenomic sequencing is a valuable tool for the taxonomic and functional profiling of microbial communities. However, this approach is challenging in samples, such as milk, where a low microbial abundance, combined with high levels of host DNA, result in inefficient and uneconomical sequencing. Here we evaluate approaches to deplete host DNA or enrich microbial DNA prior to sequencing using three commercially available kits. We compared the percentage of microbial reads obtained from each kit after shotgun metagenomic sequencing. Using bovine and human milk samples, we determined that host depletion with the MolYsis Complete5 kit significantly improved microbial sequencing depth compared to other approaches tested. Importantly, no biases were introduced. Additionally, the increased microbial sequencing depth allowed for further characterisation of the microbiome through the generation of metagenome-assembled genomes (MAGs). Furthermore, with the use of a mock community, we compared three common classifiers and determined that Kraken2 was the optimal classifier for these samples. This evaluation shows that microbiome analysis can be performed on both bovine and human milk samples at a much greater resolution without the need for more expensive deep-sequencing approaches.

3.2 Introduction

Milk is an important source of nutrition for both humans and animals. Human breast milk contains bioactive compounds that contribute to a child's development and growth (Ward, Hosid et al. 2013). With the development of culture-independent methods for studying the microbiome, increasing interest has been devoted to understanding the interplay between breast milk and the infant gut microbiome (Pannaraj, Li et al. 2017). Indeed, infants primarily fed human breast milk within the first few months of life have been shown to harbour a distinct microbiome from that of non-breast-fed infants (O'Sullivan, Farver et al. 2015). 16S rRNA gene amplicon-based analysis of the human milk microbiome has also improved the understanding of the bacterial community related to mastitis. Notwithstanding these advances, the application of shotgun metagenomic approaches has the potential to provide even greater insights into the milk microbiome to help with prevention or treatment of lactational mastitis and further promote infant and maternal health (Patel, Vaidya et al. 2017).

Besides human milk, animal milk and the resultant dairy products from cows, goats, sheep and buffalo, have proven to be versatile and valuable foods commonly consumed by humans (Haug, Høstmark et al. 2007, Getaneh, Mebrat et al. 2016, Boro, Debnath et al. 2018). The bovine milk microbiome has been intensely studied, with a view to assessing and improving animal health and ensuring product quality and safety for human consumption (Jayarao, Pillai et al. 2004, Quigley, O'Sullivan et al. 2013, Rodrigues, Lima et al. 2017). As with human mastitis, culture-based microbiology methods have had variable success in the diagnosis of bovine mastitis

due to their inability to detect low abundance pathogens and because of challenges associated with subclinical infection of unknown aetiology (Kuehn, Gorden et al. 2013). Sequencing-based approaches have also addressed this issue to some degree (Hoque, Istiaq et al. 2019). In addition to animal health, research on the bovine milk microbiome has been performed motivated by the near-universal use of cows milk as an ingredient in many dairy products worldwide. Although certain bacteria, when present in milk products, can improve the resulting sensory and flavour properties (Walsh, Crispie et al. 2016), other microorganisms can negatively influence milk by deteriorating its quality, causing spoilage, or making it unsafe for consumption (Júnior, De Oliveira et al. 2018).

High-throughput sequencing technologies, specifically targeted amplicon sequencing e.g., 16S rRNA gene, and shotgun metagenomics have both been widely adopted for the study of microbiomes, including relatively highly diverse and species-rich microbiomes such as the human gut (Schmidt, Raes et al. 2018, Subramanian, Balakrishnan et al. 2018, Almeida, Mitchell et al. 2019), human oral cavity (Verma, Garg et al. 2018) and soil (Fierer 2017, Jansson and Hofmockel 2018). For environments with a low microbial load and diversity such as milk, amplicon sequencing has been the primary tool used. For bovine milk analysis, this has provided insights into the compositional changes in raw milk due to environmental factors, processing, or the production of cheese (Ward, Hosid et al. 2013, Kable, Srisengfa et al. 2016, Rodrigues, Lima et al. 2017, Li, Wang et al. 2018, Marino, de Wittenau et al. 2019). Notwithstanding the value of data already generated, extraction methodologies have been identified as having an impact on downstream

sequencing data from 16S rRNA sequencing studies (Fouhy, Clooney et al. 2016, Douglas, Ivey et al. 2020). Furthermore, although these amplicon sequencing-based approaches have provided valuable insights, shotgun metagenomic sequencing can provide greater taxonomic insights, particularly with regard to elucidating the functional potential of the microbes present and the ability to generate metagenome-assembled genomes (MAGs) (Pasolli, Asnicar et al. 2019). These MAGs provide essentially complete genome information for key taxa present in the sample, revealing their functional and safety related properties (Bowers, Kyrpides et al. 2017). However, shotgun sequencing, as an untargeted approach, sequences all the DNA in the sample, including that from the human or animal host, which in the case of milk represents a considerable majority (up to 95%) of the DNA present (McHugh, Feehily et al. 2020). To address this challenge, previous shotgun metagenomic-based studies of the human milk microbiome (Ward, Hosid et al. 2013, Asnicar, Manara et al. 2017) carried out sequencing at great depth, followed by post-processing of data to remove non-microbial reads using bioinformatic approaches. In addition to generating large numbers of reads that need to be discarded, the limited number of microbial reads generated prevents the detection of low abundance species (Pereira-Marques, Anne et al. 2019).

Many recent studies involving clinical samples with high levels of host DNA, such as saliva, sputum, and joint fluid, have investigated depleting host DNA through commercially available kits or other chemical methods (Thoendel, Jeraldo et al. 2016, Marotz, Sanders et al. 2018, Charalampous, Kay et al. 2019, Nelson, Pope et al. 2019). However, such approaches have not been applied to milk. Often due to

logistics or sampling circumstances, samples are frozen and stored before DNA is extracted. To address this technological caveat, the aim of this study was to evaluate methods of host DNA depletion/microbial DNA enrichment to optimise shotgun sequencing-based analysis of the milk microbiome.

3.3 Materials and Methods

3.3.1 Collection, storage and initial treatment of milk samples

Five bovine raw milk samples and five human milk samples that were previously frozen at -80^{o}C were thawed for use in this study. Bovine raw milk samples were obtained from farms across Ireland, taken from bulk milk tanks and transported to the laboratory on ice where a subsample was stored at -80^{o}C. Human milk samples were collected as part of the Microbe Mom study, following informed approved consent of mothers. These samples were collected within the first month after delivery. The maximum possible volume of available samples was used in the study, which was 30 ml of bovine milk and 4 ml of human milk. For one bovine and one human milk sample, an additional aliquot was taken for use as positive controls through spiking with a mock community. Milk samples were prepared as follows: bovine milk and human milk samples were centrifuged at 4,500 x g for 20 minutes at 4^{o}C. After centrifugation, the cream and supernatant were discarded and the remaining pellets were subjected to two washing steps, where the pellets were resuspended in sterile PBS and centrifuged at 13,000 x g for 1 minute, after which the supernatant was discarded. After the washing steps, the pellets were resuspended in 2 ml sterile PBS (Fig. 1). For the positive control milk samples, the pellets were resuspended in 1.8 ml of sterile PBS and then spiked with 200 µl of a

mock community (see below for further details) and mixed thoroughly to ensure homogenization. All 12 samples were divided into two 1.5 ml tubes with 1 ml each for use in the subsequent DNA extractions (Fig. 1).

3.3.2 Description of mock community (positive control) and negative controls utilised

The 10 Strain Even Mix Whole Cell Material (ATCC® MSA-2003™; ATCC-LGC Partnership, Middlesex, United Kingdom) was used in this study as a positive control. This mock community is composed of approximately 2 x 10^7 ± 1 log cells of 10 strains, i.e., *Bacillus cereus*, *Bifidobacterium adolescentis*, *Clostridium beijerinckii*, *Deinococcus radiodurans*, *Enterococcus faecalis*, *Escherichia coli*, *Lactobacillus gasseri*, *Rhodobacter sphaeroides*, *Staphylococcus epidermidis* and *Streptococcus mutans*, at equal proportions). After reconstitution following the supplier's instructions, the mock community was spiked into each of one human and one bovine milk sample as positive controls prior to DNA extraction (BMS12 and HMS12). Sterile PBS was used as a negative control for each of the methods.

3.3.3 Host depletion, microbial enrichment, and DNA extraction

For use of the DNeasy PowerSoil Pro kit (PS), samples (250 µl) were processed according to the manufacturer's instructions (Qiagen, West Sussex, United Kingdom) with the exception that the lysis step was performed using a TissueLyser II (Qiagen) for 10 minutes at 30 Hz and DNA was eluted at 50 µl and stored at -20^{o}C. In the case of the MolYsis Complete5 kit (ML), samples (450 µl) were processed according to manufacturer's instructions (Molzym GmBH & Co.KG, Bremen,

Germany). Briefly, host cells were lysed through the addition of a chaotropic buffer, and the released nucleic acids were degraded by an enzyme, MolDNase. Microbial cells were then sedimented and lysed using kit reagents and proteinase K. Microbial DNA was then isolated and extracted using spin columns and 100 µl of DNA was eluted and stored at -20^{o}C. For the NEBNext Microbiome DNA Enrichment kit (NEB), 25 µl of input material was obtained from DNA extracted and eluted using the DNeasy PowerSoil Pro kit. Based on the DNA concentrations (ng/µl), the corresponding amount of beads needed was calculated and used according to the manufacturer's instructions (New England Biolabs, Massachusetts, United States). DNA was eventually eluted to 25 µl and stored at -20^{o}C.

3.3.4 DNA library preparation

DNA was quantified using the Qubit double-stranded DNA (dsDNA) high sensitivity assay kit (Invitrogen, Dublin, Ireland). In samples where no detectable DNA was observed from extraction, a full 5 µl of neat sample was used, otherwise 5 µl of 0.2 ng/µl DNA was used. All samples were prepared for shotgun metagenomic sequencing according to Illumina Nextera XT library preparation kit guidelines, with the use of unique dual indexes for multiplexing with the Nextera XT index kit (Illumina). Following indexing and clean up, libraries were pooled to an equimolar concentration of 1 nM and sequenced on an Illumina NextSeq 550 sequencing platform with V2 kit. DNA from samples extracted using PS underwent additional 16S rRNA amplicon sequencing according to the 16S metagenomic sequencing library preparation protocol (Illumina). DNA was amplified with primers specific to the V3-V4 variable region of the 16S rRNA gene and libraries were prepared from

the extracted DNA. Samples were multiplexed with the Nextera XT index kit (Illumina), pooled to an equimolar concentration of 20 nM and sequenced on an Illumina MiSeq sequencing platform with a V2 kit. All sequencing was performed at the Teagasc Sequencing Facility, in accordance with standard Illumina protocols.

3.3.5 Bioinformatic and statistical analysis of shotgun metagenomic data

Raw metagenomic shotgun reads were quality checked and trimmed with Cutadapt (v. 1.18) and FastQC (v. 0.11.8). Reads were then aligned to the bovine and human genome to determine the number of host reads using Bowtie2 (v. 2.3.4), all non-host reads were assumed to be microbial. For sub-sampling, paired non-host reads were rarefied to 50,000 reads five times, with seeds 3, 29, 50, 64, and 87, using seqtk (version 1.3). MetaPhlAn2, Kaiju, and Kraken2 were used for taxonomic classification and were compared against each other (Wood and Salzberg 2014, Truong, Franzosa et al. 2015, Menzel, Ng et al. 2016, Wood, Lu et al. 2019). The NCBI nr/nt database was used for Kaiju and Kraken2. Bracken was used on the Kraken2 output to determine and compute taxonomic abundances (Lu, Breitwieser et al. 2017). Functional profiling was performed using HUMAnN2, with genes classified based on Gene Ontology domains (Franzosa, McIver et al. 2018). Metagenome-assembled genomes (MAGs) were assembled by firstly using MEGAHIT (1.1.3) to assemble contigs, followed by binning using MetaBAT2 (2.12.1) and assessment of completeness and contamination using checkM (1.0.18) (Parks, Imelfort et al. 2015). Kaiju was used in the taxonomic classification of MAGs. High-quality bins were classified as >90% complete with <5% contamination as suggested by Bowers et al. (2017). For 16S rRNA amplicon sequencing, data was processed as

previously described (Fouhy, Deane et al. 2015) and taxonomy was assigned using BLAST against the SILVA SSURef database release 132. All further bioinformatics analyses, data visualisation and statistics were done in R using vegan, ggplot2 and pheatmap packages.

3.4 Results

3.4.1 Efficiency of methods for improving the proportion of microbial reads

Microbial DNA was extracted from each of six bovine and six human milk samples using the three different methods (Figure 3.1). Seven approaches were provisionally tested to assess their suitability for extracting DNA from low volumes of material (Supplemental Figure 3.1). From these, only the DNeasy PowerSoil Pro (PS) and MolYsis Complete5 (ML) kits yielded sufficient DNA concentrations. Therefore, the three methods evaluated were (i) the PS kit, which is commonly used in microbiome studies and is not specifically designed to enrich microbial DNA or deplete host DNA, (ii) the PS approach but used in combination with the NEBNext Microbiome Enrichment kit (NEB) and (iii) the ML kit, designed to decrease the amount of host DNA present. The extracted DNA was subjected to shotgun metagenomic sequencing, generating an average of 3,972,773 (± 383,371) high-quality paired end reads per sample (Supplemental Table 3.1), Although the DNA extraction approach taken did not significantly impact on the total number of reads per sample, the proportions of host to microbial reads differed considerably (Figure 3.2a). More specifically, the ML kit yielded a significantly higher ($P < 0.05$) percentage of microbial reads (Figure 3.2b), i.e., 38.31% average (range: 2.01 –

93.12%), relative to the other two methods, i.e., NEB kit 12.45% average, (range 1.03 – 41.63%) and PS kit 8.54% average (range: 1.22 – 30.28%).

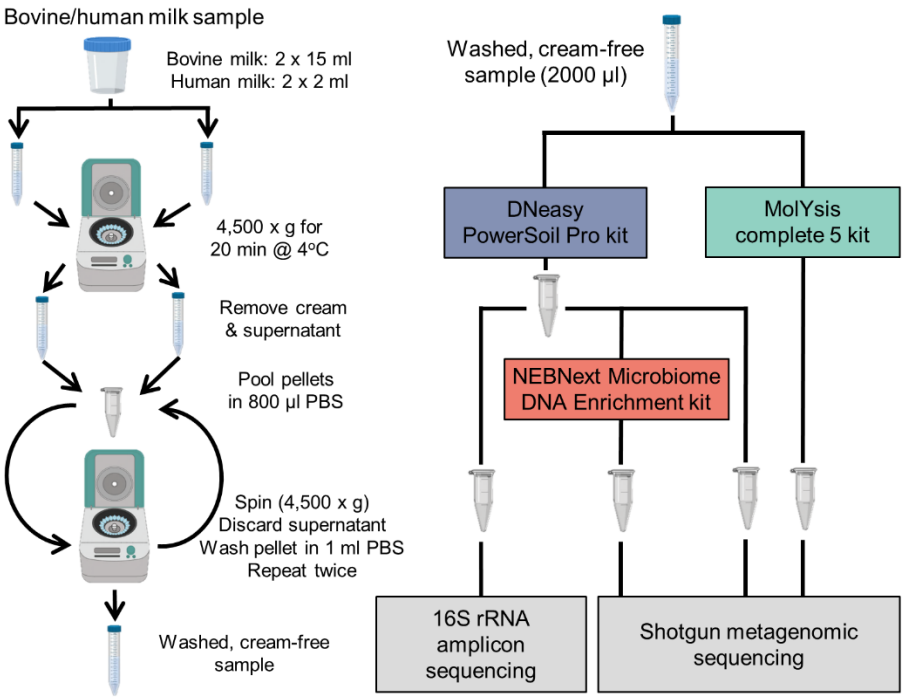


Figure 3.1. Experimental procedure used for the study. Process for skimming both bovine and human milk samples is indicated on the left. Direct extraction, host depletion, or microbial enrichment approaches are indicated on the right.

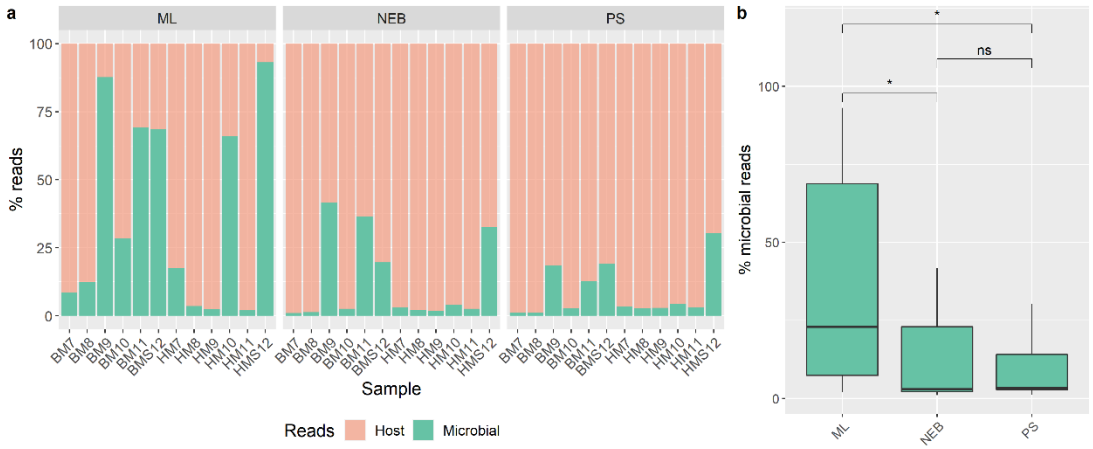


Figure 3.2. ML kit produced significantly higher proportions of microbial reads. (a) Host and microbial reads of bovine (BM) and human (HM) milk samples processed using three different methods; MolYsis Complete5 kit (ML), NEBNext Microbiome Enrichment kit (NEB) and QIAGEN DNeasy PowerSoil Pro kit (PS). (b) Comparison of the mean microbial reads of all samples within one extraction method to another. Asterisk denotes significant differences of p value < 0.05 as determined using Student's t-test.

3.4.2 Differential performance of taxonomic classification algorithms on spiked samples

Bovine and human milk samples (BMS12 and HMS12) were spiked with a mock community of ten strains, present at equal proportions, to investigate the potential for the introduction of bias due to differences in extraction efficiencies from different bacteria and to inform the choice of bioinformatic classification tools. Negative extraction controls were also included to investigate the potential for kit contamination.

After sequencing and quality control of the sequences, taxonomic classification was performed using three different tools, Kaiju, Kraken2, and MetaPhlAn2. Overall, the top 20 species detected accounted for 48.7% of the species identified by Kaiju, 98.3% of the species identified by Kraken2, and 61.3% of the species identified by MetaPhlAn2 (Figure 3.3a). The mock community accounted for an average relative abundance of 29.2% of the species identified by Kaiju, 96.6% of the species classified by Kraken2 and 58.1% of the species identified by MetaPhlAn2. All classifiers detected the 10 mock community species, although at varying relative abundances. In general, all three tools showed an under-representation of *Bifidobacterium adolescentis* (0.03 – 0.49%) and *Clostridium beijerinckii* (0.08 – 1.04%) in samples and an over-representation of *Deinococcus radiodurans* (12.29 – 38.16%) (Figure 3.3a, 3.3b). Both Kraken2 and MetaPhlAn2 produced similar mock community composition profiles, albeit with Kraken2 assigning a much higher % of reads. Both Kaiju and Kraken2 assigned reads within the negative control samples, while MetaPhlAn2 only assigned reads from one of the three negative control samples. The taxa identified were distinct from those detected in milk samples (without spiked controls) (Supplemental Table 3.2).

When the abundances of the 10 mock community strains were investigated in greater detail, Kraken2 and MetaPhlAn2 were found to assign the strains at relative abundances closer to that expected for the mock community than Kaiju. Kraken2 identified *Escherichia coli*, *Enterococcus faecalis*, *D. radiodurans*, and *Bacillus cereus* at a higher abundance than expected, while MetaPhlAn2 appeared to over-represent both *E. faecalis* and *D. radiodurans* (Figure 3.3b). When the Bray-Curtis

dissimilarity of all mock community samples were plotted in individual PCoA plots by taxonomic rank, the overall dissimilarity increased further down the taxonomic ranks, but communities classified with Kraken2 consistently clustered closer to the mock community from phylum to species (Figure 3.3c). At the highest rank (phylum), Kraken2 clustered closest to the mock community and significant differences were observed between the Kraken2 and MetaPhlAn2 communities (P = 0.012). Ultimately, Kraken2 was selected for further downstream analysis as it performed best in terms of overall correct assignment of mock community DNA.

The spiked mock community samples were additionally compared using Kraken2-classified metagenomic reads with 16S rRNA amplicon reads. The determined community composition, using these two different sequencing approaches, was similar, with small differences in relative abundances (Figure 3.3d). Additional genera, including *Acinetobacter* and *Lactococcus* were consistently detected in bovine milk samples regardless of sequencing method.

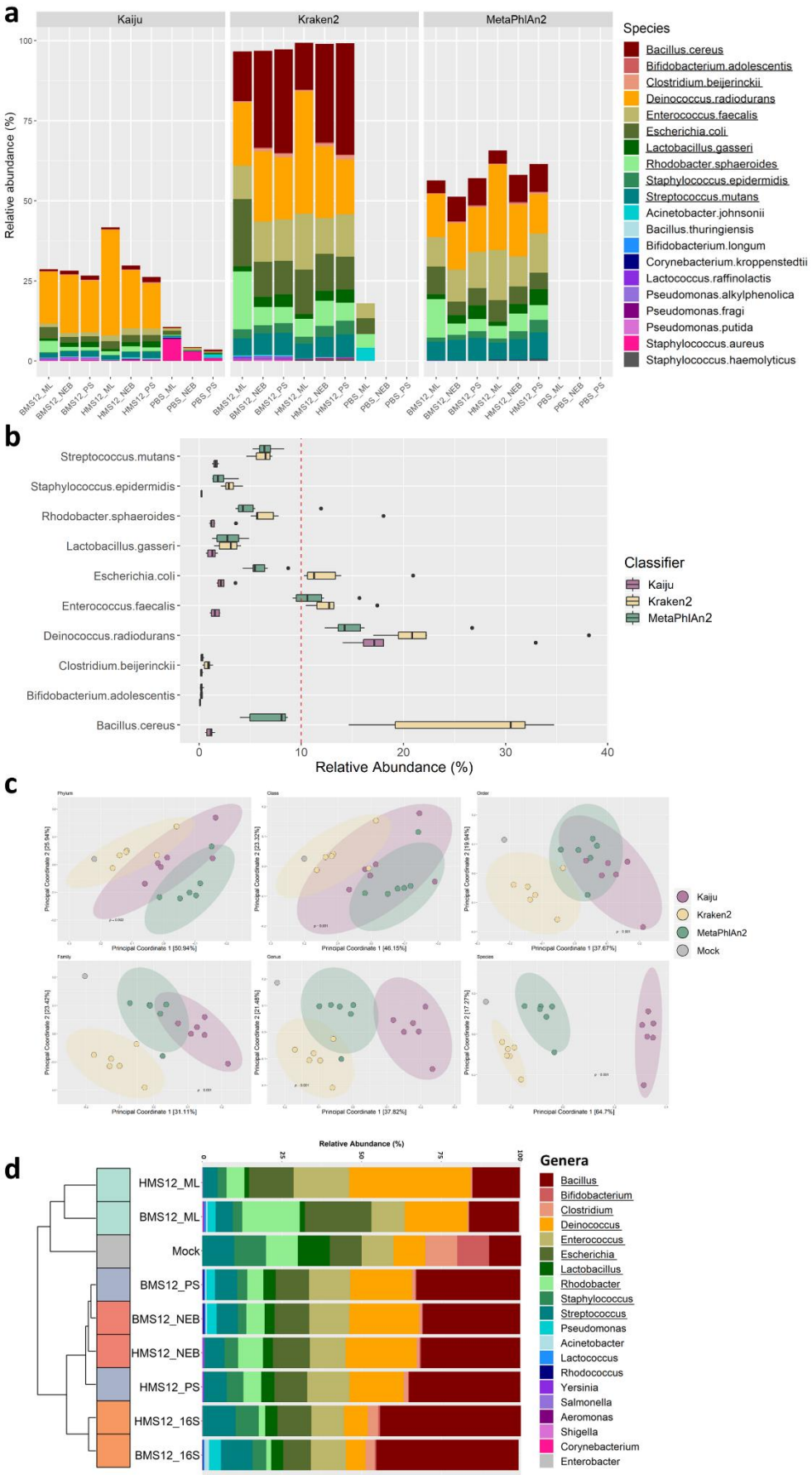

a
Kaiju
Kraken2
MetaPhlAn2
Relative abundance (%)
Species
Bacillus.cereus
Bifidobacterium.adolescentis
Clostridium.beijerinckii
Deinococcus.radiodurans
Enterococcus.faecalis
Escherichia.coli
Lactobacillus.gasseri
Rhodobacter.sphaeroides
Staphylococcus.epidermidis
Streptococcus.mutans
Acinetobacter.johnsonii
Bacillus.thuringiensis
Bifidobacterium.longum
Corynebacterium.kroppenstedtii
Lactococcus.raffinolactis
Pseudomonas.alkylphenolica
Pseudomonas.fragi
Pseudomonas.putida
Staphylococcus.aureus
Staphylococcus.haemolyticus
BMS12_ML
BMS12_NEB
BMS12_PS
HMS12_ML
HMS12_NEB
HMS12_PS
PBS_ML
PBS_NEB
PBS_PS
b
Streptococcus.mutans
Staphylococcus.epidermidis
Rhodobacter.sphaeroides
Lactobacillus.gasseri
Escherichia.coli
Enterococcus.faecalis
Deinococcus.radiodurans
Clostridium.beijerinckii
Bifidobacterium.adolescentis
Bacillus.cereus
Relative Abundance (%)
Classifier
Kaiju
Kraken2
MetaPhlAn2
c
Phylum
Class
Order
Family
Genus
Species
Principal Coordinate 1 [50.94%]
Principal Coordinate 2 [25.94%]
Principal Coordinate 1 [46.15%]
Principal Coordinate 2 [23.32%]
Principal Coordinate 1 [37.67%]
Principal Coordinate 2 [19.94%]
Principal Coordinate 1 [31.11%]
Principal Coordinate 2 [23.42%]
Principal Coordinate 1 [37.82%]
Principal Coordinate 2 [21.48%]
Principal Coordinate 1 [64.7%]
Principal Coordinate 2 [17.27%]
Kaiju
Kraken2
MetaPhlAn2
Mock
d
Relative Abundance (%)
HMS12_ML
BMS12_ML
Mock
BMS12_PS
BMS12_NEB
HMS12_NEB
HMS12_PS
HMS12_16S
BMS12_16S
Genera
Bacillus
Bifidobacterium
Clostridium
Deinococcus
Enterococcus
Escherichia
Lactobacillus
Rhodobacter
Staphylococcus
Streptococcus
Pseudomonas
Acinetobacter
Lactococcus
Rhodococcus
Yersinia
Salmonella
Aeromonas
Shigella
Corynebacterium
Enterobacter

Figure 3.3. Mock community samples were impacted by choice of bioinformatics classifier rather than method. (a) Top 20 species by relative abundance of non-host reads are displayed as determined by three different taxonomic classifiers. Unclassified and low abundance species account for the non-displayed reads. The species underlined are those present in the spiked mock community. (b) Comparison of the relative abundances of the 10 species in the mock community in the spiked milk samples according to each classification tool. Red dotted line indicates the expected 10% mark. (c) Bray-Curtis dissimilarity plots of all 6 spiked mock community samples, coloured by classifier, for each taxonomic rank. The expected mock community composition is in grey. Significant differences are noted with their p value as calculated using PERMANOVA. (d) Top 20 genus level microbial composition of each spiked milk community as determined by both shotgun metagenomics as classified using Kraken2 and 16S rRNA amplicon sequencing. Genera underlined are those present in the spiked mock community. Samples are clustered based on Bray-Curtis dissimilarity with expected mock community included.

3.4.3 Analysis of bovine and human milk samples – community structure and taxonomic and functional profiles

The species-level composition of bovine and human milk microbiome samples as assigned by Kraken2 was compared across the different extraction methods used (Figure 3.4a). In bovine milk, aliquots of the same samples that were extracted using different methods in general had generated similar profiles, albeit with slightly varying relative abundances. However, ML extraction from BM7 led to the

additional assignment of reads to *Lactococcus lactis*, *Pseudomonas alkylphenolica* and *Acinetobacter baumannii* at low abundances (<0.1%). For sample BM8, *L. lactis* was detected at higher relative abundances in the sample extracted with the ML kit (compared to the PS and NEB kit).

With regard to the human milk samples (Fig. 4b), apart from HM7, the relative taxonomy of each sample varied between by extraction methods. HM8_ML was found to be more similar in species composition to HM10_NEB and HM11_ML, reflecting the detection of *Staphylococcus epidermidis* in these samples. Similarly, a high relative abundance of *Yersinia enterocolitica* was detected in the HM9_PS, HM8_NEB, HM8_PS, HM11_NEB and HM11_PS samples.

Despite the taxonomic compositional differences, no significant dissimilarity was observed between the methods for either bovine (PERMANOVA: $P = 0.997$, $R^2 = 0.02362$) or human (PERMANOVA: $P = 0.749$, $R^2 = 0.08218$) samples (Figure 3.4c). The overall community diversity was not impacted by any method used, implying that there was no community-level bias introduced by any of the methods.
In terms of functional profiling, the predicted microbiome function in bovine or human milk samples using the three sub-domains of gene ontology (cellular components, biological processes, and molecular functions) was not significantly impacted by extraction method used (Supplemental Figure 3.2).

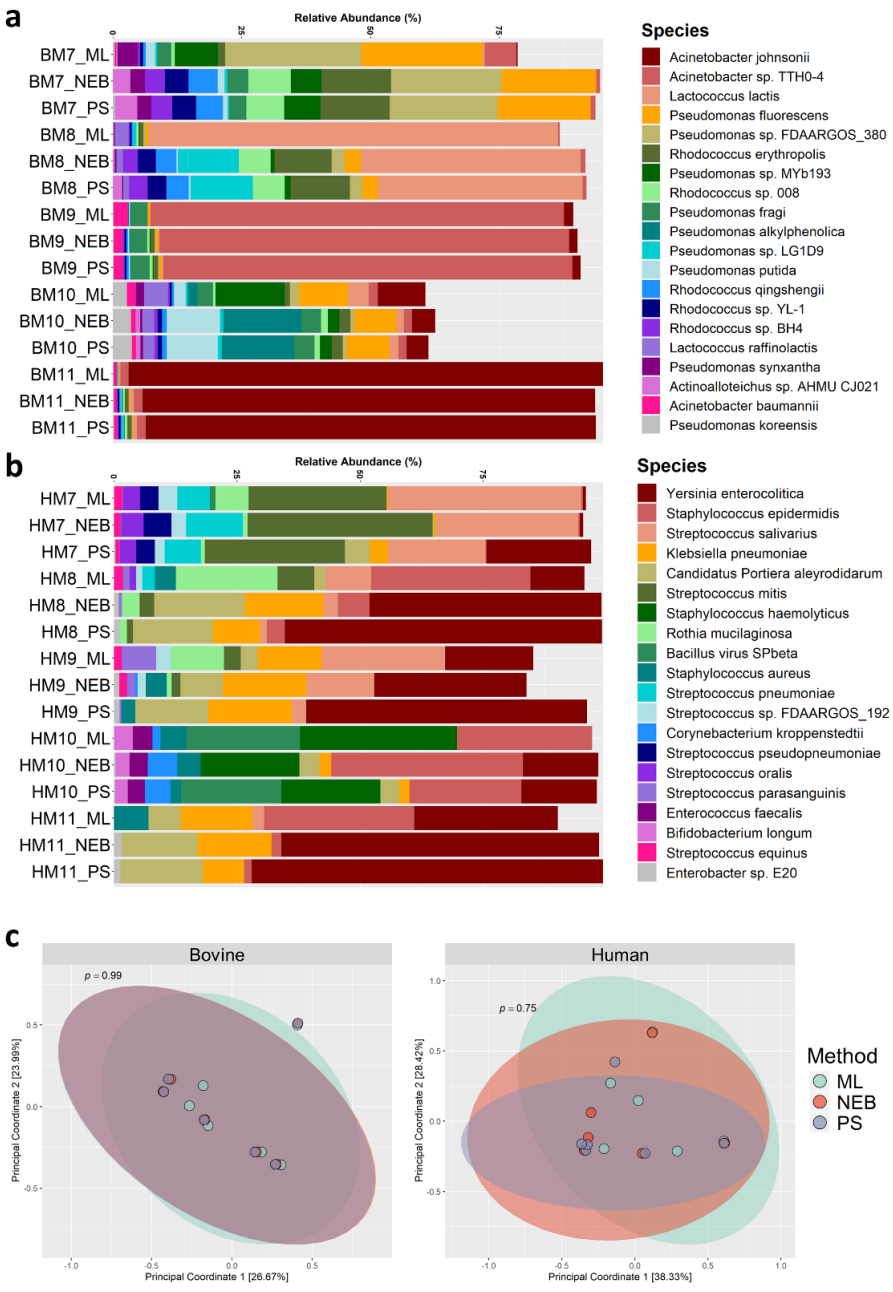


Figure 3.4. Both the bovine and human milk samples each had similar species level composition and community diversity. Top 20 species composition of all test bovine (a) and human (b) milk samples using all three extraction methods and clustered based on Bray-Curtis dissimilarity. (c) Bray-Curtis dissimilarity plots for both bovine (left) and human (right) milk samples for microbial communities determined using all three methods. Statistical community dissimilarities were calculated using PERMANOVA.

3.4.4 Characterisation of the milk microbiome

Extraction of DNA from milk samples using the ML kit resulted in greater apparent species richness from an overall higher detection of unique observed species. This increase was significant ($P < 0.05$) among bovine milk samples (Figure 3.5a). A total of 84 metagenome-assembled genomes (MAGs) were recovered across all samples, of which 19 were considered high quality, using the cut-off of > 90% completeness and < 5% contamination (Figure 3.5b, Table 3.1). From among these 19, 11 were assembled with sequence data derived from DNA extracted using the ML kit, while only 4 high-quality MAGs were recovered from DNA extracted from each of the other two methods. The majority of the high-quality MAGs recovered were from milk samples with the spiked mock community but *Corynebacterium pyruviciproducens* and *E. faecalis* MAGs were also recovered from the HM10 sample extracted using the ML kit. To evaluate if the differences between the direct sequencing (PS) and the depletion (ML) approach were due to the increased read depth or kit-introduced bias, we subsampled 50,000 non-host reads from each sample five times. Bray-Curtis analysis revealed that there was no significant difference in the community structure between both bovine (PERMANOVA: $P = 0.15$, $R^2 = 0.0260$) and human (PERMANOVA: $P = 0.07$, $R^2 = 0.0439$) samples sequenced using either approach (Supplemental Figure 3.3a, b). When comparing the dispersion within samples all related bovine samples were similar between kits, one related human sample was significantly different, and one sample could not be compared after subsampling (Supplemental Figure 3.3c, d).

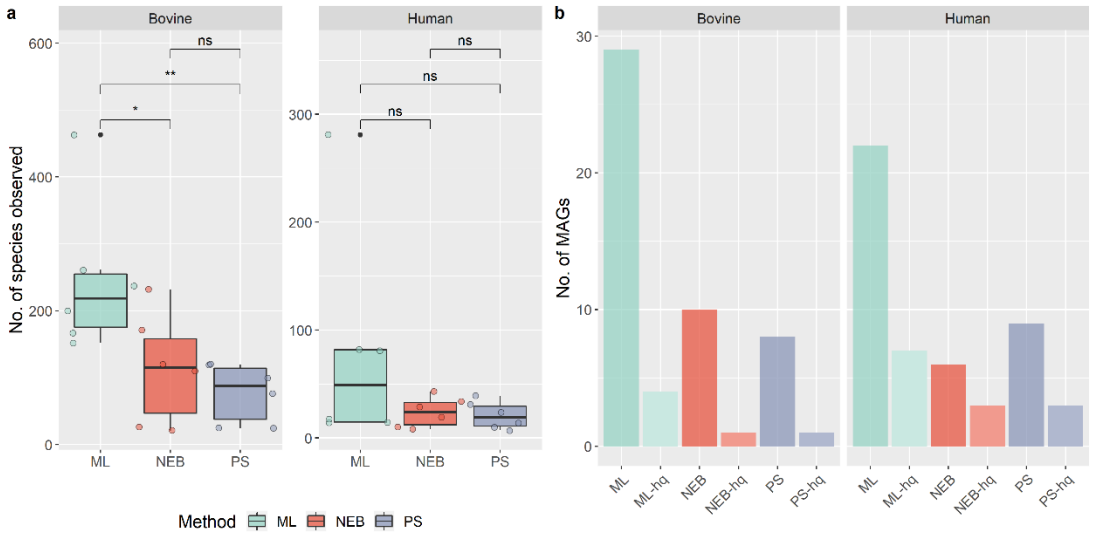


Figure 3.5. ML kit generated higher numbers of observable species and metagenome assembled genomes in milk samples. (a) Total number of species detected by each method in either bovine (left) or human (right) milk samples. A single or double asterisk indicates a significant ($P < 0.05$) or highly significant ($P < 0.01$) difference between species number as determined by ANOVA. (b) The number of total and high-quality MAGs (hq) recovered from sequenced samples using either method. High-quality MAGs were greater than 90% completeness and less than 5% contamination.

Table 3.1. High-quality metagenome-assembled genomes from shotgun sequencing analysis.

Sample	**Kit**	**Taxonomy**	**Completeness (%)[a]**	**Contamination (%)[b]**
HM10_ML	ML	*Corynebacterium pyruviciproducens*	99.76	0.1
BMS12_ML	ML	*Streptococcus mutans*	99.63	0
HMS12_ML	ML	*Enterococcus faecalis*	99.63	0

HMS12_ML	ML	*Deinococcus radiodurans*	99.15	0.21
HMS12_ML	ML	*Streptococcus mutans*	98.88	0
BMS12_ML	ML	*Deinococcus radiodurans*	98.72	0.21
HM10_ML	ML	*Enterococcus faecalis*	98.5	0
BMS12_ML	ML	*Bacillus cereus*	98.3	0.33
BMS12_ML	ML	*Rhodobacter sphaeroides*	97.58	0.15
HMS12_ML	ML	*Bacillus cereus*	96.18	0.42
HMS12_ML	ML	*Rhodobacter sphaeroides*	92.05	0.25
HMS12_NEB	NEB	*Deinococcus radiodurans*	96.17	0.21
HMS12_NEB	NEB	*Enterococcus faecalis*	94.88	0.94
HMS12_NEB	NEB	*Bacillus cereus*	94.32	0.33
BMS12_NEB	NEB	*Deinococcus radiodurans*	90.92	0.21
HMS12_PS	PS	*Bacillus cereus*	97.99	0.33
HMS12_PS	PS	*Streptococcus mutans*	96.38	1.06
HMS12_PS	PS	*Deinococcus radiodurans*	95.96	0
BMS12_PS	PS	*Deinococcus radiodurans*	91.9	0.64

[a] Completeness: ratio of observed single-copy marker genes to total single-copy marker genes in the chosen marker gene set (Parks, Imelfort et al. 2015).

[b] Contamination: ratio of observed single-copy marker genes in ≥2 copies to total single-copy marker genes in the chosen marker gene set (Parks, Imelfort et al. 2015).

3.5 Discussion

Shotgun metagenomic sequencing is a powerful method for the study of many different microbiome types. In order to maximise the potential of this method for characterising specific microbial populations, the removal of host DNA is necessary to allow for a greater sequencing depth of the microbial populations of interest. The present study evaluated three different approaches for their efficiency in host DNA depletion and, ultimately, their effectiveness with respect to studying the milk microbiome. Results showed that the ML kit gave significantly higher percentages of microbial reads when compared with the NEB kit and PS kit. The ML kit uses a chaotropic buffer that selectively lyses host cells, followed by the degradation and removal of the released host DNA using DNAse. The microbial DNA in the samples remains intact before being subjected to DNA extraction. This contrasts with the NEBNext Microbiome DNA Enrichment kit, which exploits the methylation of DNA at CpG sites in eukaryotic cells, using magnetic beads to selectively bind and remove CpG-methylated host DNA. Studies using the NEB kit for analysis of other microbiomes have shown varying results (Thoendel, Jeraldo et al. 2016, Marotz, Sanders et al. 2018, Nelson, Pope et al. 2019). The protocol recommends an input of high molecular weight gDNA (≥15 kb for optimal performance) and this was not possible to achieve in the milk samples due to the limited sample volume, and their lower relative microbial abundance when compared to samples such as human stool. Thus, the findings provided here relate to this method's performance in relation to milk samples only and should not be extrapolated to other sample types. The PS kit and similar kits have been widely used in various microbiome studies (Quigley, O'Sullivan et al. 2012, Kable, Srisengfa et al. 2016, Rodrigues, Lima et al.

2017, Falardeau, Keeney et al. 2019). The PS kit uses chemical and mechanical lysis before elution using a silica membrane spin column. Although it has no specific step for the depletion of host DNA, it has been found to be effective in terms of total DNA yield, which was evident in this study (Supplemental Table 3.1) (Quigley, O'Sullivan et al. 2012).

In this study, we found that the choice of bioinformatics tool for classification had a greater influence on community composition than the method of DNA extraction, as determined with the mock community analysis (Figures 3.3 and 4). Kraken2 and Kaiju are taxonomic binning tools, based on alignments with reference genomes. Kraken2, which was the tool that was ultimately chosen for downstream analysis, is a nucleotide sequence classifier that assigns taxonomic labels to DNA sequences. It uses an algorithm that relies on exact k-mer matches that records the species identifier for every k-mer in every genome. To classify a sequence, each k-mer in the sequence is mapped to the lowest-common ancestor (LCA) of the genomes that contain that k-mer in a database (Wood and Salzberg 2014). Kaiju works similar to Kraken2 except by classifying reads using a reference database comprising the annotated protein-coding genes of a set of microbial genomes. However, unlike Kraken2, Kaiju performed poorly in classifying the mock community samples in this study. MetaPhlAn2, unlike Kraken2 and Kaiju, is based on the alignments with species-specific marker gene sequences. Quality-controlled reads are mapped to a database of unique clade-specific marker genes with high discriminatory power, which are more optimised for human-associated gut microbial communities (Segata, Waldron et al. 2012). The relative abundances of each clade in the samples

are then estimated with species-level resolution (Truong, Franzosa et al. 2015). As MetaPhlAn2 is based on marker genes, considerable microbial sequencing depth is needed to be able to detect marker genes from low abundance organisms (McArdle and Kaforou 2020). Although MetaPhlAn2 has been found to be sensitive for low complexity food microbiomes (Walsh, Crispie et al. 2018), its performance in this study to characterise the milk microbiome was not as efficient as Kraken2, particularly in terms of total assignment, which was lower for samples classified using MetaPhlAn2 compared to Kraken2. Taken together, these results confirm that the choice of bioinformatic tool has an important influence on taxonomic composition, particularly in low abundance environments. The impact of choice of classification tools have already been reported in other studies (Walsh, Crispie et al. 2018).

The comparison of the methods used in this study showed that there were no biases in taxonomic composition and community structure when any method was used. Although the increased microbial sequencing depth of the ML kit allowed for the identification of more unique species from the milk samples compared to the other two methods, there were no significant differences in beta diversity (Figure 3.4c). Similar results were found in the analysis of functional data of the samples, where no particular method generated functional profiles that were distinct or significantly different (Supplemental Figure 3.2). In addition, we showed that metagenomic sequencing of milk samples is not limited in identification of taxa compared to the standard method of amplicon sequencing (Figure 3.3d).

The ability to track microbes at strain level is of great importance and interest both in the food industry and in uncovering mother-to-infant microbial transfer (Kable, Srisengfa et al. 2016, Asnicar, Manara et al. 2017). The ability to generate high-quality MAGs from the ML processed samples is promising when considering this challenge. Apart from recovery of mock community strains, the increased microbial sequencing depth allowed for the assembly of high-quality *Corynebacterium pyruviciproducens* and *Enterococcus faecalis* genomes from the metagenome of a human milk sample. This suggests that with greater microbial sequencing depth, the recovery of high-quality MAGs can uncover species not previously identified in the milk microbiome, particularly for species that are difficult to culture.

This study aimed to evaluate three different methods for improved microbial read sequencing depth, however certain limitations remain. The study was performed using a small number of samples, yet the current sample size was sufficient to allow for the identification of significant differences between methods. Additionally, the microbial or host DNA concentration was not absolutely quantified, which could provide a greater insight into the ratios of host to microbial DNA found in the milk samples. With respect to the mock community, it is possible that the discrepancy between observed and expected species ratios was due to a similar bias across methods, however this could not be determined in this study.

Overall, this evaluation has addressed two important problems in metagenomic sequencing of low diversity environments, specifically poor sequencing economics and poor microbial sequence depth. The results show that the host depletion

approach of the ML kit performed better than the enrichment or direct sequencing alternatives by providing the potential for deeper strain-level analysis without an observable bias. Our study additionally highlighted the importance of the correct downstream bioinformatic tool selection to accurately characterise the milk microbiome.

3.6 Declarations

3.6.1 Ethics approval and consent to participate

Ethics for collection of human breast milk samples was granted by the Research Ethics Committee of the National Maternity Hospital as part of the Microbe Mom clinical trial, registration number ISRCTN53023014. Samples were collected following informed written consent.

3.6.2 Data availability

The dataset generated and analysed during the current study are available in the ENA repository under accession number PRJEB38099.

3.6.3 Acknowledgements including funding source

Research was conducted with the financial support of the Irish Dairy Levy and Science Foundation Ireland (SFI) under Grant Numbers SFI/12/RC/2273 and 16/SP/3827 with co-funding by PrecisionBiotics Ltd. The authors thank Elaine Lawton and Fiona Crispie for their help in library preparation and DNA sequencing and David Gleeson and Lizandra Paludetti for coordinating the collection of bovine

milk samples. Additional thanks to Radka Saldova and other members of the Microbe Mom team.

3.7 References

Almeida, A., A. L. Mitchell, M. Boland, S. C. Forster, G. B. Gloor, A. Tarkowska, T. D. Lawley and R. D. Finn (2019). "A new genomic blueprint of the human gut microbiota." Nature 568(7753): 499-504.

Asnicar, F., S. Manara, M. Zolfo, D. T. Truong, M. Scholz, F. Armanini, P. Ferretti, V. Gorfer, A. Pedrotti and A. Tett (2017). "Studying vertical microbiome transmission from mothers to infants by strain-level metagenomic profiling." MSystems 2(1): e00164-00116.

Boro, P., J. Debnath, T. Kumar Das, B. C. Naha, N. Debarma, P. Deabbarma, C. Debbarma, L. Devi and T. Devi (2018). "Milk composition and factors affecting it in dairy buffaloes: a review." Journal of Entomology and Zoology Studies JEZS 340: 340-343.

Bowers, R. M., N. C. Kyrpides, R. Stepanauskas, M. Harmon-Smith, D. Doud, T. Reddy, F. Schulz, J. Jarett, A. R. Rivers and E. A. Eloe-Fadrosh (2017). "Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea." Nature biotechnology 35(8): 725-731.

Charalampous, T., G. L. Kay, H. Richardson, A. Aydin, R. Baldan, C. Jeanes, D. Rae, S. Grundy, D. J. Turner and J. Wain (2019). "Nanopore metagenomics enables rapid clinical diagnosis of bacterial lower respiratory infection." Nature Biotechnology 37(7): 783-792.

Douglas, C. A., K. L. Ivey, L. E. Papanicolas, K. P. Best, B. S. Muhlhausler and G. B. Rogers (2020). "DNA extraction approaches substantially influence the assessment of the human breast milk microbiome." Scientific reports 10(1): 1-10.

Falardeau, J., K. Keeney, A. Trmčić, D. Kitts and S. Wang (2019). "Farm-to-fork profiling of bacterial communities associated with an artisan cheese production facility." Food microbiology 83: 48-58.

Fierer, N. (2017). "Embracing the unknown: disentangling the complexities of the soil microbiome." Nature Reviews Microbiology 15(10): 579-590.

Fouhy, F., A. G. Clooney, C. Stanton, M. J. Claesson and P. D. Cotter (2016). "16S rRNA gene sequencing of mock microbial populations-impact of DNA extraction method, primer choice and sequencing platform." BMC microbiology 16(1): 123.

Fouhy, F., J. Deane, M. C. Rea, Ó. O'Sullivan, R. P. Ross, G. O'Callaghan, B. J. Plant and C. Stanton (2015). "The effects of freezing on faecal microbiota as determined using MiSeq sequencing and culture-based investigations." PloS one 10(3): e0119355.

Franzosa, E. A., L. J. McIver, G. Rahnavard, L. R. Thompson, M. Schirmer, G. Weingart, K. S. Lipson, R. Knight, J. G. Caporaso and N. Segata (2018). "Species-level functional profiling of metagenomes and metatranscriptomes." Nature methods 15(11): 962-968.

Getaneh, G., A. Mebrat, A. Wubie and H. Kendie (2016). "Review on goat milk composition and its nutritive value." Journal of Nutrition and Health Sciences 3(4): 1-10.

Haug, A., A. T. Høstmark and O. M. Harstad (2007). "Bovine milk in human nutrition–a review." Lipids in health and disease 6(1): 25.

Hoque, M. N., A. Istiaq, R. A. Clement, M. Sultana, K. A. Crandall, A. Z. Siddiki and M. A. Hossain (2019). "Metagenomic deep sequencing reveals association of microbiome signature with functional biases in bovine mastitis." Scientific reports 9(1): 1-14.

Jansson, J. K. and K. S. Hofmockel (2018). "The soil microbiome—from metagenomics to metaphenomics." Current opinion in microbiology 43: 162-168.

Jayarao, B. M., S. Pillai, A. Sawant, D. Wolfgang and N. Hegde (2004). "Guidelines for monitoring bulk tank milk somatic cell and bacterial counts." Journal of Dairy Science 87(10): 3561-3573.

Júnior, J. R., A. De Oliveira, F. d. G. Silva, R. Tamanini, A. De Oliveira and V. Beloti (2018). "The main spoilage-related psychrotrophic bacteria in refrigerated raw milk." Journal of dairy science 101(1): 75-83.

Kable, M. E., Y. Srisengfa, M. Laird, J. Zaragoza, J. McLeod, J. Heidenreich and M. L. Marco (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(4).

Kuehn, J. S., P. J. Gorden, D. Munro, R. Rong, Q. Dong, P. J. Plummer, C. Wang and G. J. Phillips (2013). "Bacterial community profiling of milk samples as a means to understand culture-negative bovine clinical mastitis." PloS one 8(4): e61959.

Li, N., Y. Wang, C. You, J. Ren, W. Chen, H. Zheng and Z. Liu (2018). "Variation in raw milk microbiota throughout 12 months and the impact of weather conditions." Scientific reports 8(1): 1-10.

Lu, J., F. P. Breitwieser, P. Thielen and S. L. Salzberg (2017). "Bracken: estimating species abundance in metagenomics data." PeerJ Computer Science 3: e104.

Marino, M., G. D. de Wittenau, E. Saccà, F. Cattonaro, A. Spadotto, N. Innocente, S. Radovic, E. Piasentier and F. Marroni (2019). "Metagenomic profiles of different types of italian high-moisture mozzarella cheese." Food microbiology 79: 123-131.

Marotz, C. A., J. G. Sanders, C. Zuniga, L. S. Zaramela, R. Knight and K. Zengler (2018). "Improving saliva shotgun metagenomics by chemical host DNA depletion." Microbiome 6(1): 1-9.

McArdle, A. J. and M. Kaforou (2020). "Sensitivity of shotgun metagenomics to host DNA: abundance estimates depend on bioinformatic tools and contamination is the main issue." Access Microbiology 2(4).

McHugh, A. J., C. Feehily, M. A. Fenelon, D. Gleeson, C. Hill and P. D. Cotter (2020). "Tracking the Dairy Microbiota from Farm Bulk Tank to Skimmed Milk Powder." Msystems 5(2).

Menzel, P., K. L. Ng and A. Krogh (2016). "Fast and sensitive taxonomic classification for metagenomics with Kaiju." Nat Commun 7: 11257.

Nelson, M. T., C. E. Pope, R. L. Marsh, D. J. Wolter, E. J. Weiss, K. R. Hager, A. T. Vo, M. J. Brittnacher, M. C. Radey and H. S. Hayden (2019). "Human and extracellular DNA depletion for metagenomic analysis of complex clinical infection samples yields optimized viable microbiome profiles." Cell reports 26(8): 2227-2240. e2225.

O'Sullivan, A., M. Farver and J. T. Smilowitz (2015). "The Influence of Early Infant-Feeding Practices on the Intestinal Microbiome and Body Composition in Infants (vol 8, pg 1, 2015)." Nutrition and Metabolic Insights 8: 87-87.

Pannaraj, P. S., F. Li, C. Cerini, J. M. Bender, S. Yang, A. Rollie, H. Adisetiyo, S. Zabih, P. J. Lincez and K. Bittinger (2017). "Association between breast milk bacterial communities and establishment and development of the infant gut microbiome." JAMA pediatrics 171(7): 647-654.

Parks, D. H., M. Imelfort, C. T. Skennerton, P. Hugenholtz and G. W. Tyson (2015). "CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes." Genome research 25(7): 1043-1055.

Pasolli, E., F. Asnicar, S. Manara, M. Zolfo, N. Karcher, F. Armanini, F. Beghini, P. Manghi, A. Tett and P. Ghensi (2019). "Extensive unexplored human microbiome diversity revealed by over 150,000 genomes from metagenomes spanning age, geography, and lifestyle." Cell 176(3): 649-662. e620.

Patel, S. H., Y. H. Vaidya, R. J. Patel, R. J. Pandit, C. G. Joshi and A. P. Kunjadiya (2017). "Culture independent assessment of human milk microbial community in lactational mastitis." Scientific reports 7(1): 1-11.

Pereira-Marques, J., H. Anne, R. M. Ferreira, M. Weber, I. Pinto-Ribeiro, L.-J. van Doorn, C. W. Knetsch and C. Figueiredo (2019). "Impact of host DNA and sequencing depth on the taxonomic resolution of whole metagenome sequencing for microbiome analysis." Frontiers in microbiology 10: 1277.

Quigley, L., O. O'Sullivan, C. Stanton, T. P. Beresford, R. P. Ross, G. F. Fitzgerald and P. D. Cotter (2013). "The complex microbiota of raw milk." FEMS microbiology reviews 37(5): 664-698.

Quigley, L., O. O'Sullivan, T. Beresford, R. Paul Ross, G. Fitzgerald and P. Cotter (2012). "A comparison of methods used to extract bacterial DNA from raw milk and raw milk cheese." Journal of Applied Microbiology 113(1): 96-105.

Rodrigues, M., S. Lima, S. Canniatti-Brazaca and R. Bicalho (2017). "The microbiome of bulk tank milk: characterization and associations with somatic cell count and bacterial count." Journal of Dairy Science 100(4): 2536-2552.

Schmidt, T. S., J. Raes and P. Bork (2018). "The human gut microbiome: from association to modulation." Cell 172(6): 1198-1215.

Segata, N., L. Waldron, A. Ballarini, V. Narasimhan, O. Jousson and C. Huttenhower (2012). "Metagenomic microbial community profiling using unique clade-specific marker genes." Nature methods 9(8): 811-814.

Subramanian, B., S. Balakrishnan, K. G. Seshadri and F. A. Valeriote (2018). "Insights into the human gut microbiome—a review." SBV J Basic Clin App Health Sci 2: 103-110.

Thoendel, M., P. R. Jeraldo, K. E. Greenwood-Quaintance, J. Z. Yao, N. Chia, A. D. Hanssen, M. P. Abdel and R. Patel (2016). "Comparison of microbial DNA enrichment tools for metagenomic whole genome sequencing." Journal of microbiological methods 127: 141-145.

Truong, D. T., E. A. Franzosa, T. L. Tickle, M. Scholz, G. Weingart, E. Pasolli, A. Tett, C. Huttenhower and N. Segata (2015). "MetaPhlAn2 for enhanced metagenomic taxonomic profiling." Nature methods 12(10): 902-903.

Verma, D., P. K. Garg and A. K. Dubey (2018). "Insights into the human oral microbiome." Archives of microbiology 200(4): 525-540.

Walsh, A. M., F. Crispie, K. Kilcawley, O. O'Sullivan, M. G. O'Sullivan, M. J. Claesson and P. D. Cotter (2016). "Microbial succession and flavor production in the fermented dairy beverage kefir." Msystems 1(5): e00052-00016.

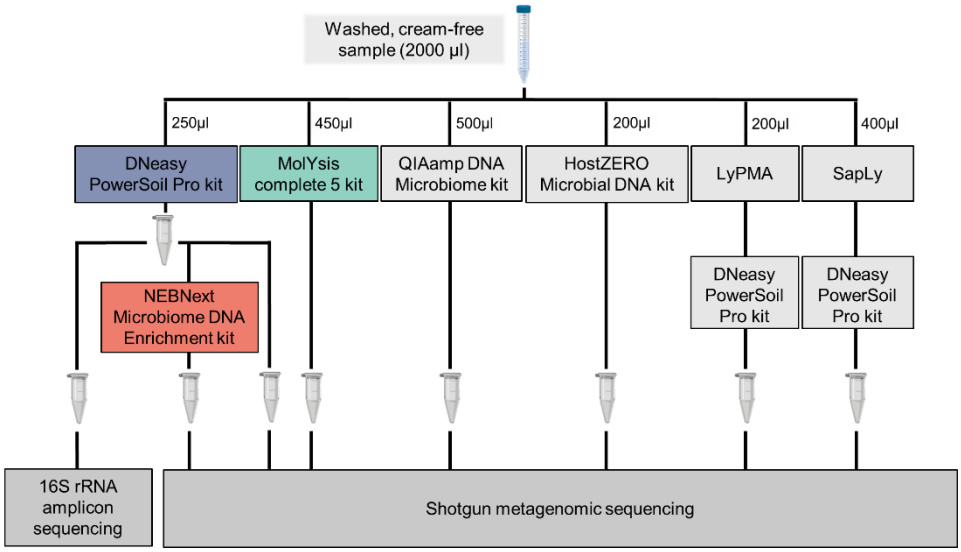
Walsh, A. M., F. Crispie, O. O'Sullivan, L. Finnegan, M. J. Claesson and P. D. Cotter (2018). "Species classifier choice is a key consideration when analysing low-complexity food microbiome data." Microbiome 6(1): 50.

Ward, T. L., S. Hosid, I. Ioshikhes and I. Altosaar (2013). "Human milk metagenome: a functional capacity analysis." BMC microbiology 13(1): 116.

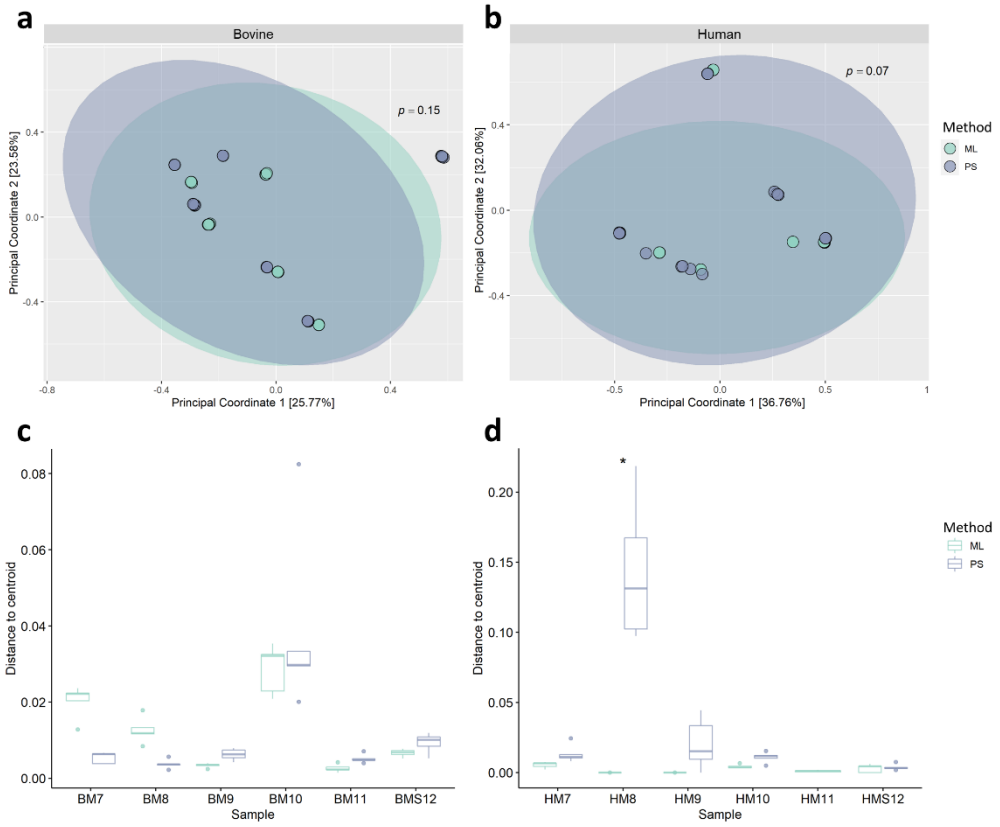
Wood, D. E., J. Lu and B. Langmead (2019). "Improved metagenomic analysis with Kraken 2." Genome Biol 20(1): 257.

Wood, D. E. and S. L. Salzberg (2014). "Kraken: ultrafast metagenomic sequence classification using exact alignments." Genome biology 15(3): 1-12.

3.8 Supplemental Material



Supplemental Figure 3.1. Initial extraction plan. Two millilitres of the cream-free milk sample were to be divided into varying volumes depending on sample volume input recommended by the different kits and methods. The QIAamp DNA Microbiome kit and HostZERO Microbial DNA kit were both available commercially (Qiagen and Zymo Research). Both kits selectively lyse host cells prior to extraction of DNA from bacterial cells. Host depletion and subsequent DNA extractions were done according to the manufacturer's instructions for both kits. The LyPMA method, according to Marotz et al. (2018), involves osmotic lysis of host cells followed by PMA treatment to remove non-microbial cells in the sample. The SapLy method, according to Charalampous et al. (2019), uses a detergent to lyse host cells followed by enzymatic removal before further extraction of DNA from microbial cells. Further evaluation of the methods in grey were discontinued due to the limited starting sample volumes. This figure was created in part with http://biorender.com.



Supplemental Figure 3.2. Functional analysis of bovine and human milk samples as based on gene ontology domains. (a) Gene families related to cellular components for bovine and (b) human milk samples. (c) Gene families related to biological processes for bovine and (d) human milk samples. (e) Gene families related to molecular function for bovine and (f) human milk samples. Figures were produced using R.

Supplementary Table 3.1. Summary of sequencing quality control data for each sample.

Sample	DNA yield (ng/µl)	Total number of raw reads	Total number of quality-filtered reads	Total number of non-host reads
BM7_ML	1.22	4,803,678	4,799,422	414,123
BM8_ML	2.02	4,218,166	4,213,627	521,433
BM9_ML	0.418	5,831,258	5,826,847	5,108,499
BM10_ML	0.764	5,503,695	5,489,219	1,559,305
BM11_ML	1.87	3,752,539	3,749,575	2,597,792
BMS12_ML	12	3,961,530	3,956,586	2,714,533
HM7_ML	0.134	4,577,558	4,572,083	801,292
HM8_ML	0.304	367,496	366,997	13,288
HM9_ML	< 0.001	899,823	898,330	22,347
HM10_ML	0.155	14,301,289	14,289,863	9,427,921
HM11_ML	< 0.001	2,558,808	2,555,194	51,285
HMS12_ML	4.01	3,135,085	3,128,637	2,913,481
PBS_ML	< 0.001	8,145	8,136	8,136
BM7_NEB	3.74	4,634,478	4,630,852	47,889
BM8_NEB	0.389	4,623,058	4,619,434	61,189
BM9_NEB	2.92	5,145,897	5,142,366	2,141,081
BM10_NEB	2.43	4,384,900	4,378,312	114,853
BM11_NEB	0.468	5,620,335	5,616,578	2,050,519
BMS12_NEB	2.4	2,990,318	2,987,713	588,814
HM7_NEB	2.17	3,714,043	3,711,382	118,695
HM8_NEB	2.86	3,505,245	3,502,584	76,561
HM9_NEB	< 0.001	7,449,515	7,444,889	137,544
HM10_NEB	1.54	4,046,734	4,043,758	166,690
HM11_NEB	< 0.001	4,773,599	4,770,318	124,328
HMS12_NEB	1.73	3,251,920	3,248,935	1,059,091
PBS_NEB	< 0.001	4,804	4,791	4,791

BM7_PS	16.6	4,306,883	4,298,577	53,485
BM8_PS	1.4	4,466,519	4,462,271	54,412
BM9_PS	12.2	949,504	948,705	174,822
BM10_PS	5.32	5,057,336	5,045,050	140,571
BM11_PS	1.93	4,039,654	4,036,626	511,785
BMS12_PS	5.83	3,734,020	3,730,408	712,548
HM7_PS	4.49	4,779,699	4,775,683	166,226
HM8_PS	5.45	3,434,585	3,431,627	97,069
HM9_PS	0.409	4,432,677	4,429,588	132,241
HM10_PS	4.08	4,606,083	4,598,639	201,712
HM11_PS	0.207	3,626,184	3,623,049	111,156
HMS12_PS	4.03	3,604,484	3,600,859	1,090,248
PBS_PS	< 0.001	659	654	654

Supplemental Table 3.2. List of species (> 1% relative abundance) assigned to negative control samples by each classifier

Classifier	**PBS_ML**	**PBS_NEB**	**PBS_PS**
MetaPhlAn2	*Lactobacillus crispatus* *Aerococcus viridans*	No species assigned	No species assigned
Kraken2	*Aerococcus viridans* *Cutibacterium acnes* *Lactobacillus amylovorus* *Actinoalloteichus* sp. AHMU CJ021 *Klebsiella pneumoniae* *Lactobacillus johnsonii* *Aerococcus urinaeequi* *Negativicoccus massiliensis* *Escherichia coli* *Enterococcus faecalis* *Rhodobacter sphaeroides* *Acinetobacter johnsonii*	*Cutibacterium acnes* *Propionibacterium* sp. oral taxon 193 *Pseudomonas pseudoalcaligenes* *Micrococcus luteus*	*Pseudomonas alcaliphila* *Negativicoccus massiliensis*

	Moraxella osloensis		
Kaiju	*Staphylococcus aureus* *Aerococcus viridans* *Lactobacillus crispatus* *Escherichia coli*	*Staphylococcus aureus*	*Ralstonia solanacearum* *Pseudomonas* sp. 286 *Acinetobacter nosocomialis* *Enterobacter cloacae* *Clostridioides difficile* *Acinetobacter johnsonii*

Chapter 4. Seasonal and geographical impact on the Irish bulk tank milk microbiota correlates with chemical composition and climatic variables.

Authors: Min Yap, Paul W. O'Toole, Orla O'Sullivan and Paul D. Cotter

Contributions: Candidate performed sample preparation, DNA extraction, sequencing library preparation, bioinformatic and statistical analysis.

Candidate drafted and edited chapter.

All authors contributed, revised, and approved manuscript.

4.1 Abstract

Season and location have previously been shown to be associated with differences in the microbiota of raw milk, especially in milk from pasture-based systems. Here we further advance research in this area by examining differences in the raw milk microbiota from several different locations across Ireland over a 12-month period, and by investigating microbiota associations with climatic variables and chemical composition. Shotgun metagenomic sequencing was used to investigate the microbiota of bulk tank raw milk collected from 9 locations (n=241) over this time. Concurrent chemical analysis of the protein, fat, lactose, total solids, non-protein nitrogen (NPN) contents and titratable acidity (TA) of the same bulk tank milk were performed. Although the raw milk microbiota was highly diverse, a core microbiota was found, with *Pseudomonas_E*, *Lactococcus, Acinetobacter* and *Leuconostoc* present in all samples. Microbial diversity significantly differed by season and location, with differences in seasonality and geography corresponding to 11.8% and 10.5% of the variation in the microbiota. Functional and antibiotic resistance profiles also varied across season and location. The analysis of other metadata revealed additional interactions, such as an association between mean daily air and grass temperatures with the abundance of spoilage taxa like *Pseudomonas* species. Correlations were identified between pathogenic, mastitis-related species, fat content and number of sun hours, suggesting a seasonal effect. Ultimately, this study expands our understanding of the interconnected nature of the microbiota, environment/climate variables and chemical composition of raw milk and provides evidence of a season- and location-specific microbiota.

4.2 Introduction

The composition of the microbiota of unprocessed milk is of great importance as it can have a major effect upon the quality of downstream dairy products. In order to produce high-quality dairy products, consistently high-quality raw milk is required. This can be challenging as there are many factors that influence the raw milk microbiota. Climate, environment, feed, season, lactation stage, storage conditions, farm system, among others, have previously been identified as contributory factors (Kable, Srisengfa et al. 2016, Doyle, Gleeson et al. 2017, Porcellato, Aspholm et al. 2018). A greater understanding of the patterns and influences of these variables on the raw milk microbiota can inform strategic interventions or practices to consistently produce high-quality raw milk ready for further processing.

High-throughput sequencing has been of considerable value when characterising diverse milk microbiomes, with the use of multiple omics methods coupled with other tools, such as modelling, generating translatable results for implementation by the industry (Ferrocino, Rantsiou et al. 2022). While culture-based approaches tend to target certain bacterial taxa, the use of high-throughput sequencing techniques can provide a more comprehensive view of the entire microbial community in milk samples. Culture-based approaches also fail to capture the viable but non-culturable bacteria or the non-readily culturable bacteria, an issue which can be overcome by the use of sequencing techniques (McHugh, Feehily et al. 2020). In addition to generating compositional data, shotgun metagenomic sequencing also provides functional profiles including information on the resistome (antibiotic resistance gene content) which has additional quality control value.

Seasonal changes can impact the raw milk microbiota, with weather conditions and temperature particularly affecting the microbial composition of raw milk (Li, Wang et al. 2018, Picinin, Bordignon-Luiz et al. 2019). Such seasonal variations have also been found to impact the microbiota of downstream dairy products, such as cheese (Sánchez-Gamboa, Hicks-Pérez et al. 2018, Ruvalcaba-Gómez, Delgado-Macuil et al. 2020). Several studies have reported differing levels of species richness and diversity between seasons, and across geographical locations (Kable, Srisengfa et al. 2016, Guo, Yu et al. 2021, Yap, Gleeson et al. 2021, Celano, Calasso et al. 2022, Liang, Wang et al. 2022). Separately, changes in the chemical or nutritional composition of milk across seasons have also been reported, with various factors including seasonality, lactation stage and farm management systems impacting the physicochemical composition of milk (Chen, Lewis et al. 2014). Protein and fat contents have reportedly displayed similar seasonal trends corresponding to stage of lactation in Ireland, Australia, and New Zealand, with fat and protein contents decreasing in spring to summer, before increasing in autumn till the end of lactation in winter (Li, Ye et al. 2019). However, no attempts have been made to correlate chemical composition with climatic factors or the bulk tank milk microbiota. The aim of this study was therefore to investigate in greater depth the fluctuations in the microbiota of raw milk collected from farms operating a pasture-based system across different regions of Ireland across a 12-month period, and to test for interactions between the milk microbiota and climatic data and milk chemical composition data.

4.3 Materials and Methods

4.3.1 Sample collection and preparation.

Raw bovine milk samples (100 ml) were collected from bulk milk tanks from farms across Ireland bi-weekly from March 2021 to March 2022. The samples were transported in a cooler box and stored at 4^{o}C for a maximum of 48h before sample processing. Samples were prepared as follows: 30 ml of the bovine milk sample was centrifuged at 4,500 x g for 20 min at 4^{o}C. After centrifugation, the cream and supernatant were discarded and the pellets were subjected to two washing steps, whereby the pellets were resuspended in sterile PBS and centrifuged at 13,000 x g for 1 minute, after which the supernatant was discarded, and the pellet was stored at -20^{o}C before DNA extraction. The chemical composition of the milk samples was determined by the Technical Services lab at the Teagasc Food Research Centre. The levels of protein, fat, lactose, total solids, non-protein nitrogen (NPN) contents and titratable acidity (TA) in bulk tank milk were determined. Monthly climate data for the sampling locations relating to mean temperature (^{o}C), total rainfall (mm), grass minimum temperature (^{o}C), mean wind speed (knots) and sunshine (hours of sun) was retrieved from the Irish Meteorological Service website (www.met.ie) (Table S2). The months of Mar, Apr and May were classified as Spring, Jun, Jul and Aug as Summer, Sep, Oct and Nov as Autumn and Dec, Jan and Feb as Winter.

4.3.2 DNA extraction.

Samples were subjected to DNA extraction using the MolYsis™ Complete5 kit (Molzym GmBH & Co. KG, Bremen, Germany), with 50 µL of DNA eluted for downstream sequencing. The MolYsis kit was used as it had been found to improve

microbiota characterisation through significantly improving microbial sequencing depth of milk samples (Yap, Feehily et al. 2020). gDNA was quantified using the Qubit dsDNA HS assay kit (Invitrogen) and stored at -20^{o}C before library preparation.

4.3.3 DNA library preparation and shotgun metagenomic sequencing.

248 samples (241 milk samples and 7 controls) were prepared for shotgun metagenomic sequencing according to Illumina Nextera XT library preparation kit guidelines, with the use of unique dual indexes for multiplexing with the Nextera XT index kit (Illumina). Following indexing and clean up, samples were pooled to equimolar concentration of 1 nM. Samples were sequenced in two pools, the first pool containing 98 samples on an Illumina NextSeq 550 sequencing platform with a V2 kit and the second containing 150 samples on an Illumina NextSeq 2000 sequencing platform with a P3 chip, at the Teagasc DNA Sequencing Facility, using standard Illumina sequencing protocols. Inter-run controls were used, and no inter-run variations were identified.

4.3.4 Bioinformatic analysis of shotgun metagenomic data.

Quality checks and adapter trimming were performed with FastQC (0.11.8) and cutadapt (2.6) and host reads were aligned to the bovine genome (*Bos taurus*) and removed with Bowtie2 (2.4.4). Taxonomic classification was performed with Kraken2 (2.0.7) (Wood, Lu et al. 2019) using the Genome Taxonomy Database (release 89) which contains Bacteria and Archaea (Parks, Chuvochina et al. 2018, Parks, Chuvochina et al. 2020). SUPER-FOCUS (Silva, Green et al. 2016) was used to

predict the microbiological functional potential of shotgun reads, through the alignment of reads against a reduced SEED database (Overbeek, Begley et al. 2005) using DIAMOND (Buchfink, Xie et al. 2015), with results classified into subsystems (sets of protein families with similar function). Resistome analysis was done using Resistance Gene Identifier (RGI 4.2.2), with the strict cut-off (Alcock, Raphenya et al. 2020). Assembly of Metagenome Assembled Genomes (MAGs) was done using metaSPAdes (3.13) (Nurk, Meleshko et al. 2017), followed by binning with MetaBAT2 (2.12.1) (Kang, Li et al. 2019) and quality assessment with checkM (1.0.18) (Parks, Imelfort et al. 2015). High-quality MAGs, of at least 90% completeness and less than 5% contamination, were assigned taxonomy with GTDB-tk (2.1.1) (Chaumeil, Mussig et al. 2022).

4.3.5 Statistical analysis and data visualisation.

Statistical analysis and data visualisation was performed in R (4.1.2) (Team 2013). All data was cleaned, analysed, and visualised in R with ggplot2, tidyverse and ggpubr packages (Wickham, Averick et al. 2019, Kassambara 2020). Kruskal-Wallis and pairwise Wilcoxon rank sum tests with Benjamini-Hochberg P-value correction were used for comparison between sampling seasons and locations. Microbiota diversity analysis was performed with the vegan package (Oksanen 2007) and beta diversity was calculated as Bray-Curtis metrics, visualised in a principal coordinate analysis plot. The “adonis” function from the vegan package was used to calculate the permutational analysis of variance (PERMANOVA) to determine differences in composition of the community between groups of samples (number of permutations = 999). Redundancy analysis was also done with vegan and visualised

using the ggord package (Beck 2017). The “multiplatt” function from the indicspecies package was used to identify taxa that were significantly associated with particular seasons and sampling locations, by calculating Pearson’s phi coefficient of association and correcting for unequal group sizes using the parameter “r.g” (Cáceres and Legendre 2009). Pearson’s correlation was measured with the R base function, cor, and visualised using ggcorrplot (Kassambara 2019).

4.4 Results

4.4.1 The Irish bulk tank milk microbiota is highly diverse, but there is a core microbiota.

241 raw milk samples collected from 9 different locations across Ireland were sampled over the course of a year. An average of 3,890,014 (± 2,258,584) high-quality paired-end reads corresponding to 0.52 Gbp were generated per sample. An average of 68.6% of metagenomic reads were bovine genome reads (range 0.4 to 98.7%). This resulted in an average of 1,121,982 (± 875,900) reads that were assigned as microbial reads (mean,31.4%, range 1.35 – 99.61%) and used for taxonomic and functional characterisation. A total of 1,112 genera and 3,332 species were detected in samples, with 26 genera and 49 species detected in samples at an average of at least 1% relative abundance and, thus, a high proportion of genera and species were detected at levels lower than 1%. At species level, species identified were classified into 5 categories (environment, host, pathogen, spoilage, technologically relevant), based on previous literature (Supplemental Table 4.1). A total of 694 MAGs were assembled, of which 153 were high-quality MAGs that were classified into 23 different genera and 35 different

species, and with 7 MAGs that could only be classified to genus level (Supplemental Table 4.2). Among the high-quality MAGs recovered, we identified 19 unique taxa that were not detected based on the classification of short reads (Supplemental Table 4.2).

Despite considerable inter-sample variation, a core microbiota was detectable, with the genera *Pseudomonas_E*, *Lactococcus, Acinetobacter* and *Leuconostoc* present in all samples, contributing to a total of 34.2% (17% *Pseudomonas_E*, 7.5% *Lactococcus*, 7.2% *Acinetobacter*, 2.5% *Leuconostoc*) of the relative abundance of the bulk milk microbiota. The genus *Lactococcus* had the greatest number of species (n=4) present in at least 90% of the samples, and these 4 species (*L. lactis, L. lactis_E, L. piscium_C* and *L. raffinolactis*) together accounted for an average of 5% of the milk microbiota (Figure 4.1). As well as being the most abundant at genus level, there were also more species of *Pseudomonas* detected than for any other genus, i.e., 12 species of *Pseudomonas* were detected in 73% of the samples (178/244 samples), with *Pseudomonas_E lundensis*, *P. lactis* and *P. fragi_B* being the three most commonly detected species from the genus *Pseudomonas*. Four species of *Acinetobacter* were identified, with *Acinetobacter albensis* present in 97% of the samples (237/244), accounting for 3.9% relative abundance. Two species of *Leuconostoc* were present, *Leuconostoc lactis_A* and *L. mesenteroides*, in 95% and 98% of the samples, respectively, accounting for 1.9% of relative abundance. It was also noted that *Bifidobacterium pseudolongum_A* and *Staphylococcus aureus* were present in 83% (202/244 samples) and 96% of all samples (234/244 samples) and contributed an average relative abundance of 0.21% and 0.86%, respectively.

Pararhizobium sp.001426685 was only present in 17 samples (7% of all samples) but was one of the top 10 most abundant species detected, accounting for an average of 1.46% relative abundance (Figure 4.1). Similar patterns were evident when MAGs were analysed, in that *Pseudomonas_E*, *Lactococcus, Acinetobacter* and *Leuconostoc* had the greatest numbers of high-quality MAGs recovered (Table S2). Interestingly, 4 additional species of *Lactococcus* were identified compared to the short read classification reported above. Twelve MAGs of *L. cremoris*, 9 MAGs of *L. laudensis*, and 2 MAGs of *L. carnosus* and 1 MAG of *L. petauri* were recovered in addition to the species mentioned above (*L. lactis* and *L. raffinolactis*). For the MAGs classified as *Pseudomonas*, 3 additional species were recovered (*P. bubulae*, *P. paracarnis* and *P. saxonica*). Conversely, more species were classified as *Acinetobacter* based on the short read data compared to the recovered *Acinetobacter* MAGs, in that only *A. albensis* (25 MAGs) and *Acinetobacter* sp002135415 (1 MAG) were recovered. Lastly, MAGs corresponding to *Leuconostoc mesenteroides* (16 MAGs) and *L. rapi* (1 MAG) were also recovered.

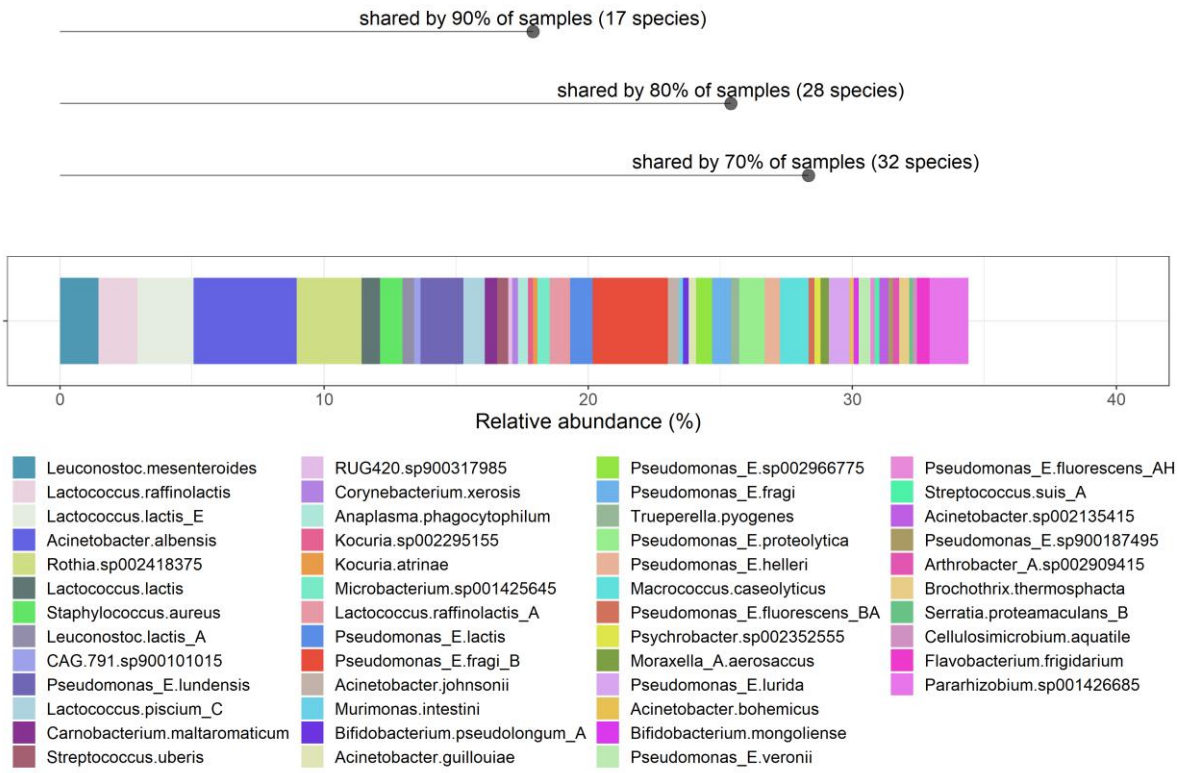


Figure 4.1. Species level composition of the bulk milk microbiota, comprising species with an average relative abundance of greater than 1%. This accounts for 34.4% of the total average relative abundance.

4.4.2 Both season and location influence the bulk tank milk microbiota.

The alpha diversity (observed taxa, Shannon, and Simpson metrics) of the bulk tank milk microbiota differed significantly across seasons, with significantly greater numbers of taxa observed in the summer months and significantly fewer taxa detected in the winter months ($P < 0.05$) (Figure 4.2a). According to the Shannon and Simpson metrics, the diversity during the summer and autumn months was significantly higher than for spring and winter ($P < 0.05$). Across the 9 sampling locations, there were a lower number of taxa observed at locations H and I, while Shannon and Simpson diversity indices were generally lower across locations A, G,

H compared to the other locations (Figure 4.2b). The beta diversity analysis of the bulk tank milk microbiota revealed significant separation of samples across seasons, with each season distinct from the others (PERMANOVA, $P < 0.05$) (Figure 4.2c, Table 4.1). The beta diversity difference between sampling locations was also statistically significant (PERMANOVA, $P < 0.05$) (Table 4.1), with sampling locations A and H being the most distinct (Figure 4.2d). Both season and location interacted significantly with the variation in microbiota composition ($P < 0.05$), and accounted for 11.8% and 10.5% of the variation, respectively.

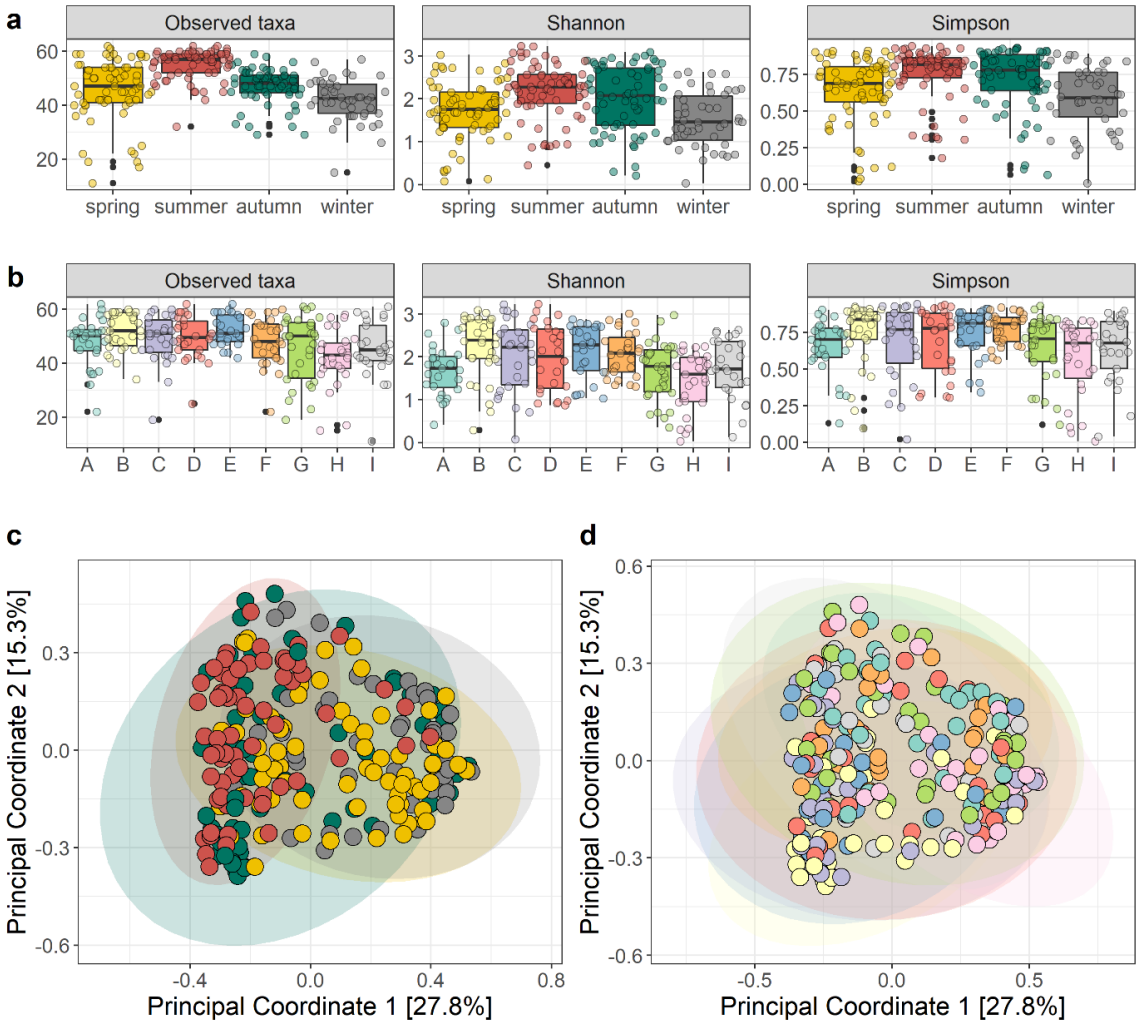


Figure 4.2. Diversity of bulk tank milk microbiota, with the observed genera, Shannon, and Simpson alpha diversity metrics by (a) season and (b) sampling location across the 12-month sampling period. Bray-Curtis principal-coordinate analysis (PCoA plots), by (c) season and (d) sampling location with ellipsis representing clustering by season and sampling location, respectively.

Table 4.1. Sample metadata and the degree to which their variation explains differences in the taxonomic composition and functional profiles of the bulk tank milk microbiota, calculated using PERMANOVA (adonis, vegan).

	Variation explained by	
Model or variable	**(R^2)**	***P***
Taxonomic composition		
Location	10.523	0.001 ***
Season	11.802	0.001 ***
Mean temperature	1.089	0.001 ***
Total rainfall	0.812	0.01 **
Grass temperature	1.068	0.006 **
Mean wind speed	0.336	0.328
Sun hours	0.635	0.042 *
Functional profile		
Location	9.76	0.001 ***
Season	12.4	0.001 ***
Mean temperature	0.867	0.052
Total rainfall	0.695	0.106
Grass temperature	0.431	0.256
Mean wind speed	0.231	0.506
Sun hours	1.02	0.04 *

Significant differences in the relative abundances of 16 of the 20 most abundant genera were found between seasons (Figure 4.3a, Supplemental Table 4.3). Specifically with respect to the core genera, the relative abundance of *Acinetobacter* was higher in summer compared to spring, *Lactococcus* was significantly more abundant in summer than the other seasons and *Pseudomonas_E* was significantly more abundant in winter and spring compared to summer and autumn ($P < 0.05$), while no seasonal differences in the relative abundances of *Leuconostoc* were observed (Figure 4.3b). For other taxa, samples collected in the summer contained significantly higher relative abundances of *Anaplasma*, *Flavobacterium*, *Macrococcus*, *Moraxella_A*, *Rothia*, *Staphylococcus* and *Streptococcus*, and significantly lower relative abundances of CAG-791 compared to the other three seasons ($P < 0.05$) (Figure 4.3a). Spring samples had significantly higher relative abundances of *Pararhizobium* and *Psychrobacter* compared to the other seasons, while relative abundances of *Microbacterium* and *Rothia*, and *Brochothrix* were significantly lower in winter and autumn samples, respectively. Both winter and spring samples had significantly higher abundances of *Bifidobacterium* compared to autumn and summer, but significantly lower abundances of *Macrococcus* and *Morexella_A* compared to autumn samples. *Carnobacterium* was detected in significantly higher relative abundances in summer and winter samples than those collected in spring and autumn.

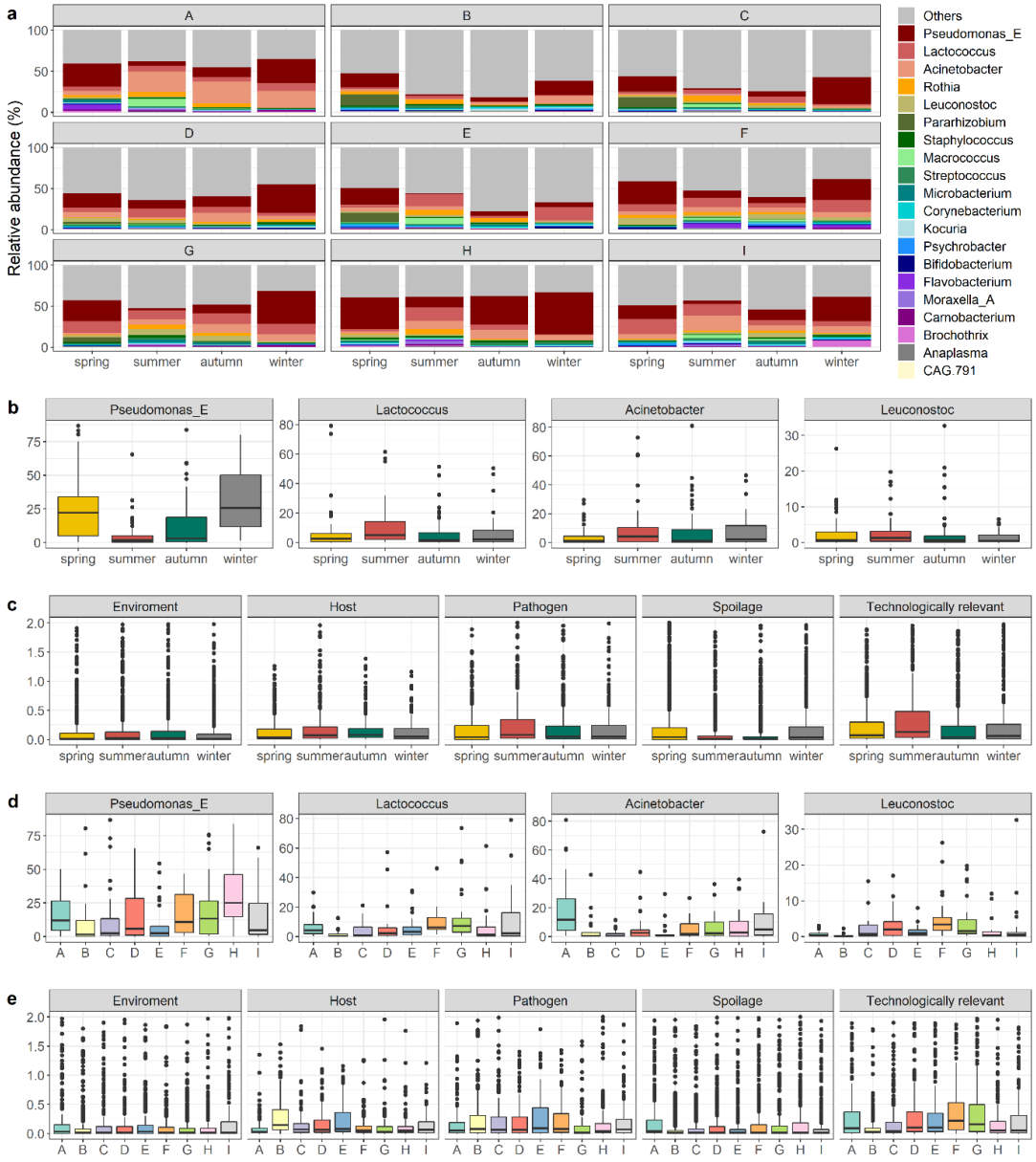


Figure 4.3. Taxonomic profiles and the core microbiota of the bulk tank milk microbiota. (a) Genus level taxonomic composition of the top 20 most abundant genera of the bulk tank milk microbiota across seasons by sampling location, and the changes in the core genera (b) and in species grouped into 5 categories (c) by season, and the changes in the core genera (d) and in species grouped into 5 categories (e). The five categories are environment-related, host-related, pathogenic, spoilage and technologically relevant species and the species in each category is in Table S1.

At the species level, species identified were assigned into 5 categories (environment, host, pathogen, spoilage, technologically relevant), based on previous literature (Supplemental Table 4.1). Species related to the environment (soil, hay) had significantly lower abundances in winter and significantly higher abundances in summer, while species related to spoilage were significantly higher in spring and winter compared to summer and autumn ($P < 0.05$) (Figure 4.3c) , which mostly consisted of *Pseudomonas* species (Supplemental Table 4.1). Summer milk samples had significantly higher abundances of environment-related, host-related, pathogenic and technologically relevant species ($P < 0.05$) (Supplemental Table 4.1). Additional analysis was performed to determine if any particular species was significantly associated with a particular season and, while no particular species was specifically associated with autumn samples, those collected in spring and summer were each significantly associated with 6 unique species and winter with 7 species ($P < 0.05$). Spring was associated with the presence of *Pararhizobium* sp001426685, *Pseudomonas_E veronii*, *Psychrobacter* sp002352555, *Bifidobacterium pseudolongum_A*, *Pseudomonas_E* sp900187495 and *Cellulosimicrobium aquatile*, summer samples were associated with *Anaplasma phagocytophilum, Lactococcus lactis*, *Macrococcus caseolyticus*, *Moraxella_A aerosaccus*, *Murimonas intestini* and *Rothia* sp002418375 and winter samples were associated with CAG-791 sp900101015, *Kocuria atrinae*, *Pseudomonas_E fragi*, *Pseudomonas_E fragi_B*, *Pseudomonas_E lactis*, *Pseudomonas_E lurida* and RUG420 sp900317985.

Significant differences in the relative abundances of 16 of the top 20 genera were also identified across sampling locations (Figure 4.3a, Supplemental Table 4.4). Among the core genera, the relative abundances differed across sampling locations, with relative abundances of *Acinetobacter* being highest in samples from location A (19.34%), followed by samples from location I (10.06%) (Figure 4.3d). *Lactococcus* abundances were highest in samples from G (12.77%), followed by those from I (10.97%) and F (10.19%). Location F had the highest relative abundance of *Leuconostoc* (5.22%), while H had the highest for *Pseudomonas_E* (32.59%) (Figure 4.3d). In particular, samples from location A had significantly higher abundances of *Acinetobacter*, *Brochothrix* and *Carnobacterium* and significantly lower *Bifidobacterium* compared to the other locations ($P < 0.05$). Location B samples had the lowest abundance of *Lactococcus* (1.85%), which was significantly lower compared to all other locations except for location C (4.2%) ($P < 0.05$), and had the highest abundances of *Kocuria*, though these levels were only significantly higher than the values from samples from A, F and G ($P < 0.05$). Location F had the highest relative abundance of *Leuconostoc*, which was significantly higher than all locations except D and G ($P < 0.05$). Relative abundances of *Pararhizobium* were lowest in location H (0.001%) compared to other locations, with B (4.54%), C (3.13%) and E (4.38%) having significantly higher relative abundances ($P < 0.05$). *Pseudomonas_E* was higher in H (32.59%) compared to all locations, and was significantly higher than B (10.85%), C (14.29%), E (10.29%) and I (15.02%) ($P < 0.05$). When the species identified were classified into categories, significantly higher abundances of environment-related species were detected in milk from A ($P < 0.05$), while B and E had significantly higher abundances of host-related species ($P < 0.05$) (Figure 4.3e,

Supplemental Table 4.1). Abundances of pathogens were significantly lower in milk from sampling location G compared to the other locations ($P < 0.05$). For species related to spoilage, significantly higher abundances were found in milk from locations A and H, and significantly lower abundances in milk from B ($P < 0.05$) and for technologically relevant species (mainly from the genera *Bifidobacterium*, *Lactococcus* and *Leuconostoc*), significantly higher abundances were found in milk from F and G and significantly lower abundances were found in milk from B ($P < 0.05$) (Figure 4.3e). Further analysis found significant associations between 6 species and 4 sampling locations ($P < 0.05$). *Acinetobacter albensis* was significantly associated with A, while CAG-791 sp900101015 and *Kocuria atrinae* were associated with B, *Leuconostoc lactis_A* was associated with D and *Leuconostoc mesenteroides*, *Lactococcus raffinolactis_A* and *Bifidobacterium mongoliense* were associated with F. No particular species were correlated with sampling locations C, E, G, H and I.

4.4.3 Functional and antibiotic resistance profiles differed by season and sampling location.

Season and location also impacted the predicted function of the bulk tank milk microbiota, with differences found in subsystem level 1 functions. Similar to taxonomic characteristics, the beta diversity of functional profiles was significantly different across seasons, with each being season distinct from the others, with the exception of spring and winter (PERMANOVA, $P < 0.05$) (Figure 4.4a, Table 4.1). With respect to sampling locations, significant differences were observed between locations (PERMANOVA, $P < 0.05$) (Table 4.1), with A and H differing from B, C, E

and F, and H and I clustering separately from I (Figure 4.4b). Differences were also observed between abundances of subsystem level 1 functions, with 14 and 19 functions significantly different by season and location, respectively (Supplemental Table 4.3 and 4.4, Figure 4.4c).

The resistome of the raw milk microbiota was predicted using RGI (Alcock, Raphenya et al. 2020) and the number of contigs mapping to antibiotic resistance genes were categorised into different drug classes (Figure 4.5). Genes from a total of 27 drug class resistance mechanisms were detected. The greatest proportion of genes belonged to the multidrug resistance category (31.7%), followed by genes encoding resistance to aminoglycosides (18.8%) and tetracycline (17.1%). Seasonal variation was evident, with a significantly greater number of contigs of relevance in samples taken in spring compared to the other seasons ($P < 0.05$). Although not statistically significant, differences between sampling locations were seen, with location H having the highest number of mapped contigs (1761 contigs) and C had the lowest (1130 contigs).

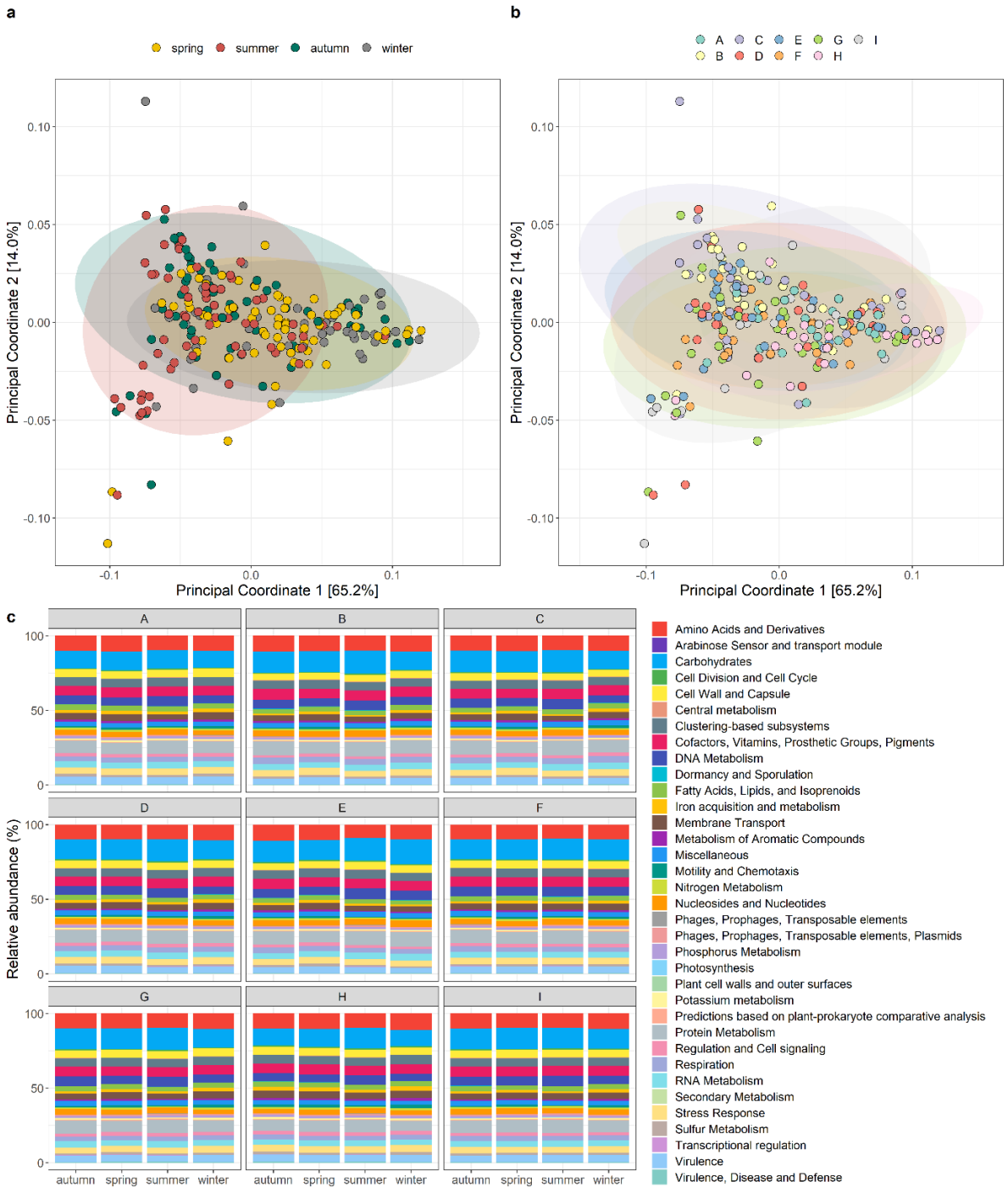


Figure 4.4. Functional characteristics of the bulk tank milk microbiota based on SUPERFOCUS analysis. Beta diversity Bray-Curtis PCoA plots by (a) season and (b) location, and (c) the subsystem level 1 functions by season across the sampling locations.

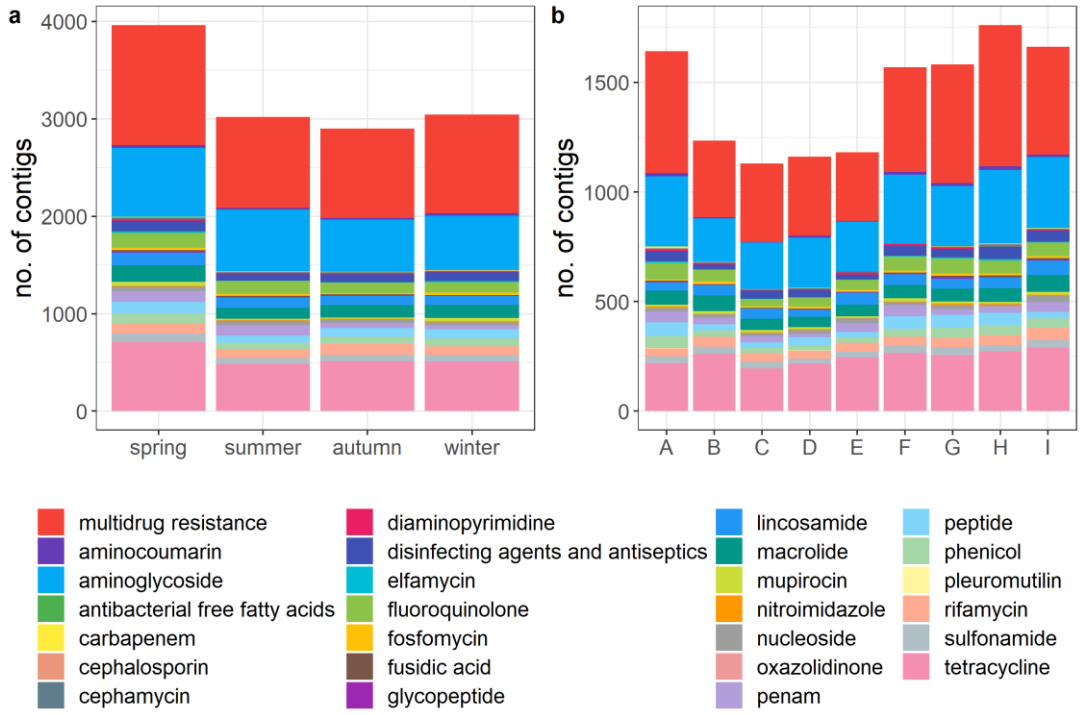


Figure 4.5. Composition of the number of contigs mapped to genes from different drug resistance classes of the bulk tank milk microbiota, by (a) season and (b) sampling location.

4.4.4 The bulk tank milk microbiota correlated with climatic factors and chemical composition.

Data from the analysis of the chemical composition of raw milk and climatic data across the sampling period are summarized in Tables 4.2 and 4.3. Redundancy analysis (RDA) was performed to investigate correlations between several climatic factors (mean temperature, total rainfall, grass temperature, mean wind speed and number of hours of sunshine) and the raw milk microbiota as well as the relationship between the microbiota and chemical composition (protein content, fat content, lactose content, total solid content, titratable acidity, non-protein nitrogen content) (Figure 4.6a and b). The Pearson's correlation coefficient was

calculated, and statistical significance was tested (Figure 4.6c). Redundancy analysis showed that the climatic variables explained a proportion of the variation in the bulk milk microbiota, with a constrained proportion of 14.37%, and with all variables statistically significant ($P < 0.05$). When visualised by season, logical patterns were seen, with mean wind speed and total rainfall related to spring and winter and mean temperature and grass temperature correlated to summer (Figure 4.6a). Of note, *Pararhizobium* and *Agrobacterium* abundances were found to be related to spring and sampling locations B, C, and E, but these genera were also found to be associated with the lactose content of milk (Figure 4.6a and b). *Pseudomonas_E* abundance was related to winter and spring seasons, and was significantly associated with the increase of the mean wind speed and total rainfall and the decrease of mean and grass temperature (Figure 4.6a), which was confirmed with a significant correlation found with these same variables, with *Pseudomonas* spp. under spoilage-related taxa (Figure 4.6c, Supplemental Table 4.1). No other visible strong relationships in any particular sampling location were detected (Figure 4.6b). From Pearson's correlation coefficient analysis, several associations were apparent between the climatic variables and chemical composition and the bulk tank microbiota. The individual elements of the chemical composition analysis (protein, fat, lactose, total solids, TA, NPN contents) of raw milk were highly correlated with each other as shown in Figure 4.6c, where significant correlations, both positive and negative were seen between the variables. No significant correlations were found between environment, host-related or technologically relevant taxa, and the chemical composition or climatic variables (Figure 4.6c). Spoilage-related species were negatively correlated with

mean and grass temperature and positively correlated with mean wind speed (Figure 4.6c), suggesting low temperature spoilage in raw milk. Pathogenic species related to mastitis or disease were significantly correlated with measures of fat, total solids content and number of sun hours (Figure 4.6c).

Table 4.2. Sample size, chemical composition, and climatic variables of bulk tank raw milk across season.

	Season			
	Spring	**Summer**	**Autumn**	**Winter**
N	69	66	63	46
Chemical composition				
Protein (%)	3.349 ± 0.1	3.455 ± 0.1	3.832 ± 0.2	3.391 ± 0.2
Fat (%)	4.15 ± 0.2	3.997 ± 0.3	4.578 ± 0.3	4.27 ± 0.3
Lactose (%)	4.82 ± 0.1	4.69 ± 0.1	4.52 ± 0.2	4.605 ± 0.2
Total solids (%)	13.04 ± 0.4	12.758 ± 0.4	13.565 ± 0.6	13.105 ± 0.4
TA (%)	0.16 ± 0.01	0.16 ± 0.01	0.15 ± 0.01	0.15 ± 0.01
NPN (%)	0.026 ± 0.002	0.027 ± 0.003	0.031 ± 0.004	0.026 ± 0.004
Climatic variables				
Mean temperature (°C)	6.7 ± 1.3	15.6 ± 1.2	11.8 ± 2.7	7.4 ± 0.7
Total rainfall (mm)	83.1 ± 34.1	58.3 ± 16.8	102.3 ± 41.1	96.9 ± 32.7
Grass temperature (°C)	-7.3 ± 1.4	4.2 ± 1.5	0.1 ± 2.4	-4.9 ± 1.0
Mean wind speed (knots)	6.2 ± 0.4	5.0 ± 0.4	5 ± 0.4	7.6 ± 1.8
Hours of sunshine (h)	11.1 ± 2.0	15.0 ± 1.0	9.7 ± 1.0	7.8 ± 0.9

Table 4.3. Sample size, chemical composition, and climatic variables of bulk tank raw milk across locations.

	Location								
	A	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**
N	27	30	27	26	25	27	30	27	25
Chemical composition									
Protein (%)	3.50 ± 0.2	3.46 ± 0.2	3.51 ± 0.2	3.52 ± 0.3	3.45 ± 0.3	3.34 ± 0.2	3.50 ± 0.2	3.34 ± 0.1	3.54 ± 0.3
Fat (%)	4.23 ± 0.6	4.22 ± 0.3	4.32 ± 0.3	4.24 ± 0.4	4.24 ± 0.4	4.09 ± 0.2	4.28 ± 0.4	4.15 ± 0.2	4.27 ± 0.4
Lactose (%)	4.65 ± 0.2	4.62 ± 0.2	4.64 ± 0.3	4.69 ± 0.1	4.75 ± 0.2	4.58 ± 0.2	4.67 ± 0.2	4.81 ± 0.1	4.67 ± 0.2
Total solids (%)	13.05 ± 0.6	13.09 ± 0.4	13.1 ± 0.5	13.09 ± 0.5	13.12 ± 0.5	12.72 ± 0.4	13.12 ± 0.5	13.03 ± 0.3	13.3 ± 0.8
TA (%)	0.16 ± 0.012	0.15 ± 0.012	0.15 ± 0.013	0.16 ± 0.012	0.15 ± 0.014	0.15 ± 0.01	0.15 ± 0.015	0.15 ± 0.008	0.15 ± 0.01
NPN (%)	0.028 ± 0.004	0.027 ± 0.004	0.027 ± 0.004	0.028 ± 0.004	0.027 ± 0.004	0.027 ± 0.004	0.027 ± 0.004	0.027 ± 0.002	0.028 ± 0.007
Climatic variables									
Mean temperature (°C)	9.8 ± 3.9	8.4 ± 3.8	8.4 ± 3.9	9.1 ± 3.8	9.8 ± 3.8	9.8 ± 3.9	8.4 ± 3.8	9.8 ± 3.9	8.4 ± 3.9
Total rainfall (mm)	83.1 ± 36.7	62.9 ± 36.6	62.9 ± 37.3	73 ± 36.4	62.9 ± 36.5	83.1 ± 35.8	62.9 ± 36.9	83.1 ± 36.9	83.1 ± 35.3
Grass temperature (°C)	-3.9 ± 4.3	-4.2 ± 4.3	-3.9 ± 4.3	-3.9 ± 4.3	-3.9 ± 4.4	-3.9 ± 4.4	-4.2 ± 4.2	-3.9 ± 4.3	-3.9 ± 4.4
Mean wind speed (knots)	5.2 ± 1.2	5.2 ± 1.1	5.2 ± 1.1	5.4 ± 1.2	5.2 ± 0.9	5.2 ± 1.2	5.2 ± 1.2	5.6 ± 1.2	5.2 ± 1.3
Hours of sunshine (h)	11.1 ± 2.9	11.1 ± 3.0	11.1 ± 3.1	11.1 ± 2.8	11.1 ± 2.7	11.1 ± 3.0	10.7 ± 3.1	11.1 ± 3.1	10.2 ± 2.9

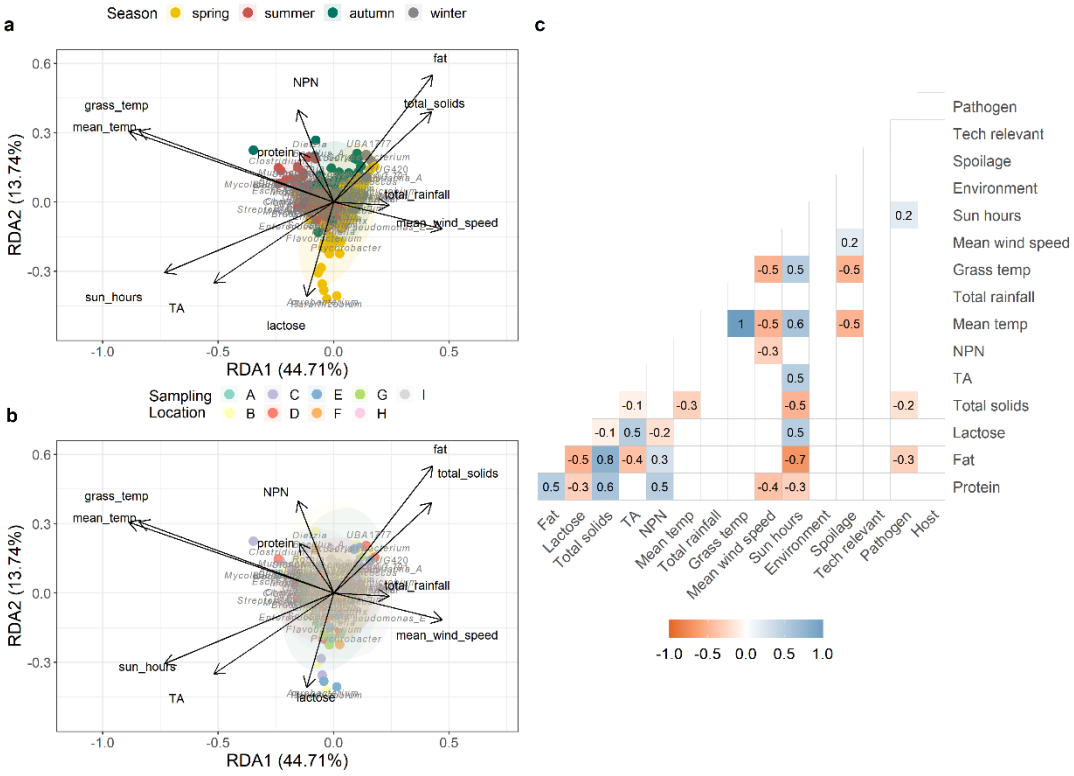


Figure 4.6. Correlation of climatic factors and chemical data and the raw milk microbiota. Redundancy analysis (RDA) of the relationship between the climatic factors and chemical data and the relative abundance of the top 20 genera of the bulk tank raw milk microbiota, as seen by (a) season and (b) location, and (c) Pearson correlation matrix of raw milk chemical composition, climatic factors with the species detected in the raw milk microbiota grouped into the 5 different categories (Table S1).

4.5 Discussion

In this study, raw milk was found to have a diverse microbiota, which differed according to season and location, as reported previously by us and by others (Guo, Yu et al. 2021, Yap, Gleeson et al. 2021, Celano, Calasso et al. 2022). At species level, a large proportion (65.6%) of taxa were present at less than 1% relative abundance, which is consistent with previous studies (Quigley, O'Sullivan et al. 2013, Kable, Srisengfa et al. 2016). In spite of this, a core microbiota was observed, with *Acinetobacter*, *Lactococcus*, *Leuconostoc* and *Pseudomonas* identified in all bulk tank raw milk samples. *Pseudomonas*, *Lactococcus* and *Acinetobacter* have been commonly detected in raw milk samples using high-throughput sequencing (Li, Wang et al. 2018, Liang, Wang et al. 2022). While less abundant, *Leuconostoc* has been found to be prevalent in Irish bovine milk, and is a potentially beneficial microbe, with some species contributing to the organoleptic properties of resultant dairy products (Quigley, O'Sullivan et al. 2013). Several factors, seasonality and geography in particular, have been found to correlate with this diversity.

The bacterial communities and the associated functional profiles and resistome of samples differed across each season. Samples from summer were observed to have the most diverse community with the highest number of taxa observed. Significant differences in the bacterial community structure between sampling seasons have also been found by Kable, Srisengfa et al. (2016) and Guo, Yu et al. (2021), with similar trends of greater evenness and richness in the spring and summer months compared to autumn or winter months. This pattern is, however, different from our previous study, that noted a higher alpha diversity in samples taken in November

(Yap, Gleeson et al. 2021), and the observations of Skeie, Håland et al. (2019) could be due to a variety of factors such as weather, animal health, feed or farm environment. In summer, abundances of environment and host-related, pathogenic and technologically relevant species were higher. This has also been reported in several studies where *Acinetobacter* (environment-related), *Lactococcus* and *Bifidobacterium* (frequently technologically relevant), *Corynebacterium* and *Streptococcus* (potentially skin/pathogenic/mastitis-related) species were more abundant in warmer months (Kable, Srisengfa et al. 2016, Li, Wang et al. 2018, Celano, Calasso et al. 2022). Host-related species like *Anaplasma phagocytophilum* and *Murimonas intestini* have been previously detected in bulk tank milk (Yap, Gleeson et al. 2021). In spring and winter, spoilage-related taxa like *Pseudomonas* spp., *Brochothrix thermosphacta* and *Flavobacterium frigidarium* were in greater abundance, which is expected due to their psychrotrophic nature (Porcellato, Aspholm et al. 2018, Skeie, Håland et al. 2019). The fluctuating abundances of taxa across the year show the influence of seasonality on the raw milk microbiota. The influence of seasonality on the chemical composition of raw milk was similar to that previously reported, with total solids higher in autumn than summer, fat content higher in autumn, protein content higher in autumn and winter and no seasonal variations found for lactose (Chen, Lewis et al. 2014). In terms of the resistome, spring samples were found to have a significantly higher number of contigs harbouring antibiotic resistance genes. While the influence of seasonality on the bulk tank milk resistome has not been investigated, the resistome of any environment will be linked to microbiota composition (Nikoloudaki, Junior et al. 2021).

This study also found significant differences in diversity and taxonomic composition across sampling locations, with certain taxa associated with specific locations. This suggests the presence of location-specific microbial communities in raw milk, where environmental factors at each sampling location influence the raw milk microbiota. Environmental and farming practices have been suggested as possible reasons for the differences in abundances across locations (Oliver, Jayarao et al. 2005, Breitenwieser, Doll et al. 2020). Variations in host and environment-related species, in addition to the association of certain species with 4 out of the 9 locations, support the hypothetical existence of a location-specific microbiota, but more data over a longer period of time is needed to establish this definitively. Regarding the resistome, in line with previous reports, the number of contigs harbouring antibiotic resistance genes differed across sampling locations (Liu, Zhu et al. 2020).

When comparing the influence of both season and geography, similar findings to those reported by Guo, Yu et al. (2021) were noted in that seasonal variations were associated with more significant differences in the raw milk microbiota than differences in geographical location. More specifically, although both factors were associated with statistically significant differences, seasonality explained a greater proportion of variation compared to location for both the taxonomic composition and the functional profiles. There are many other variables that contribute to differences in the milk microbiota that were not examined in this study but have been reported by others. These include animal housing regime, milking technology employed, animal health, diet, storage, transport, among others (Kable, Srisengfa et

al. 2016, Doyle, Gleeson et al. 2017, Frétin, Martin et al. 2018, Porcellato, Aspholm et al. 2018, Skeie, Håland et al. 2019, Du, Meng et al. 2020, Ouamba, Gagnon et al. 2022). In this study, climatic variables were found to significantly explain the variation in the bulk tank milk microbiota through PERMANOVA and redundancy analysis, a pattern which has been previously reported (Li, Wang et al. 2018). Indeed, all variables except mean wind speed were found to significantly influence variation in the milk microbiota. This is consistent with findings where climatic variables were the most frequent factors impacting production level and comfort of grazing animals (Molina Benavides, Velez Terranova et al. 2022). Mean daily and grass temperature seemed to have had a greater impact and were seen to be correlate with taxa related spoilage like *Pseudomonas* species. Shathele (2009) previously found that temperature was related to mastitis occurrence, with different bacteria causing mastitis in warm and cold months. Temperature or heat stress has also been found to affect cows, which in turn affect the composition of milk and the occurrence of mastitis (Stürmer, Busanello et al. 2018). Relationships between fat content and hours of sun and the presence of pathogenic species were seen in our study. This suggests possible seasonal effects on both the fat content and microbiota of raw milk. Albonico, Barelli et al. (2020) found the fat composition of milk was strongly associated with milk microbiota richness. Therefore, from the current study, climatic variables were seen to interact with and apparently influence the chemical and microbiological characteristics of raw milk, similar to results reported by Molina Benavides, Velez Terranova et al. (2022), which could be useful in predicting raw milk quality and safety, though further work has to be done.

This study is one of the few studies that employed shotgun metagenomic sequencing on the bulk tank milk microbiota. Many previous studies have used amplicon sequencing or culture-based methods to understand the composition of milk across farms and seasons (Kable, Srisengfa et al. 2016, Li, Wang et al. 2018, Picinin, Bordignon-Luiz et al. 2019, Celano, Calasso et al. 2022). However, the current study had several limitations, one of which is the reliance on sequencing data alone, which meant that only relative abundances (rather than absolute abundances) of taxa could be inferred, which adds some bias as a rise in one taxa could mean the decrease in another. Furthermore, all microbial DNA present in the sample was sequenced, which included both viable and potentially some non-viable microorganisms. Moreover, sampling of raw milk from the same farms across multiple years would help to validate findings on the core microbiota and season and location-specific nature of the raw milk microbiota.

This study clearly adds to the current research on the effect of seasonality and geography on the microbiota and resistome of bulk tank raw milk. Additionally, the effects of climatic variables on the chemical properties and composition of the raw milk microbiota show how interconnected all these factors are, which has potential for use in predicting raw milk quality and safety through modelling. Despite the highly diverse milk microbiota, a core microbiota exists and potentially season and location-specific taxa, but further sampling needs to be done to determine if these patterns hold true over an even longer period.

4.6 Declarations

4.6.1 Data availability

Sequence data generated during the current study have been deposited in the European Nucleotide Archive under the accession number PRJEB55911.

4.6.2 Acknowledgements including funding source

Research was conducted with the financial support of the Irish Dairy Levy and VistaMilk. This research was conducted with the financial support of Science Foundation Ireland (SFI) together with the Irish Department of Agriculture, Food and the Marine under grant number SFI/16/RC/3835 (VistaMilk). The authors thank Dr Diarmuid Sheehan for sample coordination, Arlene McGrath for sample collection and chemical composition data, Dr Cathy Lordan for help with sample preparation and Dr Fiona Crispie for help with DNA sequencing.

4.7 References

Albonico, F., C. Barelli, D. Albanese, M. Manica, E. Partel, F. Rosso, S. Ripellino, M. Pindo, C. Donati and A. Zecconi (2020). "Raw milk and fecal microbiota of commercial Alpine dairy cows varies with herd, fat content and diet." Plos one 15(8): e0237262.

Alcock, B. P., A. R. Raphenya, T. T. Lau, K. K. Tsang, M. Bouchard, A. Edalatmand, W. Huynh, A.-L. V. Nguyen, A. A. Cheng and S. Liu (2020). "CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database." Nucleic acids research 48(D1): D517-D525.

Beck, M. (2017). "ggord: Ordination Plots with ggplot2." R package version 1(0): 588.

Breitenwieser, F., E. V. Doll, T. Clavel, S. Scherer and M. Wenning (2020). "Complementary use of cultivation and high-throughput amplicon sequencing reveals high biodiversity within raw milk microbiota." Frontiers in microbiology 11: 1557.

Buchfink, B., C. Xie and D. H. Huson (2015). "Fast and sensitive protein alignment using DIAMOND." Nature methods 12(1): 59-60.

Cáceres, M. D. and P. Legendre (2009). "Associations between species and groups of sites: indices and statistical inference." Ecology 90(12): 3566-3574.

Celano, G., M. Calasso, G. Costantino, M. Vacca, A. Ressa, O. Nikoloudaki, P. De Palo, F. M. Calabrese, M. Gobbetti and M. De Angelis (2022). "Effect of Seasonality on Microbiological Variability of Raw Cow Milk from Apulian Dairy Farms in Italy." Microbiology Spectrum: e00514-00522.

Chaumeil, P.-A., A. J. Mussig, P. Hugenholtz and D. H. Parks (2022). "GTDB-Tk v2: memory friendly classification with the Genome Taxonomy Database." bioRxiv.

Chen, B., M. J. Lewis and A. S. Grandison (2014). "Effect of seasonal variation on the composition and properties of raw milk destined for processing in the UK." Food chemistry 158: 216-223.

Doyle, C. J., D. Gleeson, P. W. O'Toole and P. D. Cotter (2017). "High-throughput metataxonomic characterization of the raw milk microbiota identifies changes reflecting lactation stage and storage conditions." International journal of food microbiology 255: 1-6.

Doyle, C. J., D. Gleeson, P. W. O'Toole and P. D. Cotter (2017). "Impacts of seasonal housing and teat preparation on raw milk microbiota: a high-throughput sequencing study." Applied and environmental microbiology 83(2): e02694-02616.

Du, B., L. Meng, H. Liu, N. Zheng, Y. Zhang, X. Guo, S. Zhao, F. Li and J. Wang (2020). "Impacts of milking and housing environment on milk microbiota." Animals 10(12): 2339.

Ferrocino, I., K. Rantsiou and L. Cocolin (2022). "Investigating dairy microbiome: an opportunity to ensure quality, safety and typicity." Current Opinion in Biotechnology 73: 164-170.

Frétin, M., B. Martin, E. Rifa, V.-M. Isabelle, D. Pomiès, A. Ferlay, M.-C. Montel and C. Delbès (2018). "Bacterial community assembly from cow teat skin to ripened cheeses is influenced by grazing systems." Scientific reports 8(1): 1-11.

Guo, X., Z. Yu, F. Zhao, Z. Sun, L.-Y. Kwok and S. Li (2021). "Both sampling seasonality and geographic origin contribute significantly to variations in raw milk microbiota, but sampling seasonality is the more determining factor." Journal of Dairy Science 104(10): 10609-10627.

Kable, M. E., Y. Srisengfa, M. Laird, J. Zaragoza, J. McLeod, J. Heidenreich and M. L. Marco (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(4): e00836-00816.

Kang, D. D., F. Li, E. Kirton, A. Thomas, R. Egan, H. An and Z. Wang (2019). "MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies." PeerJ 7: e7359.

Kassambara, A. (2019). "ggcorrplot: Visualization of a Correlation Matrix using'ggplot2'." R package version 0.1 3.

Kassambara, A. (2020). "ggpubr:"ggplot2" based publication ready plots."

Li, N., Y. Wang, C. You, J. Ren, W. Chen, H. Zheng and Z. Liu (2018). "Variation in raw milk microbiota throughout 12 months and the impact of weather conditions." Scientific reports 8(1): 1-10.

Li, S., A. Ye and H. Singh (2019). "Seasonal variations in composition, properties, and heat-induced changes in bovine milk in a seasonal calving system." Journal of dairy science 102(9): 7747-7759.

Liang, L., P. Wang, X. Zhao, L. He, T. Qu and Y. Chen (2022). "Single-molecule real-time sequencing reveals differences in bacterial diversity in raw milk in different regions and seasons in China." Journal of Dairy Science.

Liu, J., Y. Zhu, M. Jay-Russell, D. G. Lemay and D. A. Mills (2020). "Reservoirs of antimicrobial resistance genes in retail raw milk." Microbiome 8(1): 1-15.

McHugh, A. J., C. Feehily, M. A. Fenelon, D. Gleeson, C. Hill and P. D. Cotter (2020). "Tracking the dairy microbiota from farm bulk tank to skimmed milk powder." MSystems 5(2): e00226-00220.

Molina Benavides, R., M. Velez Terranova, S. Perilla Duque, R. Campos Gaona and H. Sanchez Guerrero (2022). "Correlations between bulk tank milk analysis with weather conditions in dairy farms under tropical environments." Italian Journal of Animal Science 21(1): 217-227.

Nikoloudaki, O., W. J. L. Junior, S. Campanaro, R. Di Cagno and M. Gobbetti (2021). "Role prediction of Gram-negative species in the resistome of raw cow's milk." International Journal of Food Microbiology 340: 109045.

Nurk, S., D. Meleshko, A. Korobeynikov and P. A. Pevzner (2017). "metaSPAdes: a new versatile metagenomic assembler." Genome research 27(5): 824-834.

Oksanen, J. (2007). "Vegan: community ecology package. R package version 1.8-5." http://www. cran. r-project. org.

Oliver, S. P., B. M. Jayarao and R. A. Almeida (2005). "Foodborne pathogens in milk and the dairy farm environment: food safety and public health implications." Foodbourne Pathogens & Disease 2(2): 115-129.

Ouamba, A. J., M. Gagnon, G. LaPointe, P. Y. Chouinard and D. Roy (2022). "Graduate Student Literature Review: Farm management practices: Potential microbial sources that determine the microbiota of raw bovine milk." Journal of Dairy Science.

Overbeek, R., T. Begley, R. M. Butler, J. V. Choudhuri, H.-Y. Chuang, M. Cohoon, V. de Crécy-Lagard, N. Diaz, T. Disz and R. Edwards (2005). "The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes." Nucleic acids research 33(17): 5691-5702.

Parks, D. H., M. Chuvochina, P.-A. Chaumeil, C. Rinke, A. J. Mussig and P. Hugenholtz (2020). "A complete domain-to-species taxonomy for Bacteria and Archaea." Nature biotechnology 38(9): 1079-1086.

Parks, D. H., M. Chuvochina, D. W. Waite, C. Rinke, A. Skarshewski, P.-A. Chaumeil and P. Hugenholtz (2018). "A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life." Nature biotechnology 36(10): 996-1004.

Parks, D. H., M. Imelfort, C. T. Skennerton, P. Hugenholtz and G. W. Tyson (2015). "CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes." Genome research 25(7): 1043-1055.

Picinin, L., M. Bordignon-Luiz, M. Cerqueira, I. Toaldo, F. Souza, M. Leite, L. Fonseca and A. Lana (2019). "Effect of seasonal conditions and milk management practices

on bulk milk quality in Minas Gerais State-Brazil." Arquivo Brasileiro de Medicina Veterinária e Zootecnia 71: 1355-1363.

Porcellato, D., M. Aspholm, S. B. Skeie, M. Monshaugen, J. Brendehaug and H. Mellegård (2018). "Microbial diversity of consumption milk during processing and storage." International journal of food microbiology 266: 21-30.

Quigley, L., O. O'Sullivan, C. Stanton, T. P. Beresford, R. P. Ross, G. F. Fitzgerald and P. D. Cotter (2013). "The complex microbiota of raw milk." FEMS microbiology reviews 37(5): 664-698.

Ruvalcaba-Gómez, J. M., R. J. Delgado-Macuil, L. X. Zelaya-Molina, O. Maya-Lucas, E. Ruesga-Gutiérrez, L. M. Anaya-Esparza, Z. Villagrán-de la Mora, D. A. López-De la Mora and R. I. Arteaga-Garibay (2020). "Bacterial succession through the artisanal process and seasonal effects defining bacterial communities of raw-milk Adobera cheese revealed by high throughput DNA sequencing." Microorganisms 9(1): 24.

Sánchez-Gamboa, C., L. Hicks-Pérez, N. Gutiérrez-Méndez, N. Heredia, S. García and G. V. Nevárez-Moorillón (2018). "Microbiological changes during ripening of Chihuahua cheese manufactured with raw milk and its seasonal variations." Foods 7(9): 153.

Shathele, M. (2009). "Weather effect on bacterial mastitis in dairy cows." International Journal of Dairy Science 4(2): 57-66.

Silva, G. G. Z., K. T. Green, B. E. Dutilh and R. A. Edwards (2016). "SUPER-FOCUS: a tool for agile functional analysis of shotgun metagenomic data." Bioinformatics 32(3): 354-361.

Skeie, S. B., M. Håland, I. M. Thorsen, J. Narvhus and D. Porcellato (2019). "Bulk tank raw milk microbiota differs within and between farms: A moving goalpost challenging quality control." Journal of dairy science 102(3): 1959-1971.

Stürmer, M., M. Busanello, J. P. Velho, V. I. Heck and I. M. P. Haygert-Velho (2018). "Relationship between climatic variables and the variation in bulk tank milk

composition using canonical correlation analysis." International Journal of Biometeorology 62(9): 1663-1674.

Team, R. C. (2013). "R: A language and environment for statistical computing."

Wickham, H., M. Averick, J. Bryan, W. Chang, L. D. A. McGowan, R. François, G. Grolemund, A. Hayes, L. Henry and J. Hester (2019). "Welcome to the Tidyverse." Journal of open source software 4(43): 1686.

Wood, D. E., J. Lu and B. Langmead (2019). "Improved metagenomic analysis with Kraken 2." Genome biology 20(1): 1-13.

Yap, M., C. Feehily, C. J. Walsh, M. Fenelon, E. F. Murphy, F. M. McAuliffe, D. van Sinderen, P. W. O'Toole, O. O'Sullivan and P. D. Cotter (2020). "Evaluation of methods for the reduction of contaminating host reads when performing shotgun metagenomic sequencing of the milk microbiome." Scientific reports 10(1): 1-11.

Yap, M., D. Gleeson, P. W. O'Toole, O. O'Sullivan and P. D. Cotter (2021). "Seasonality and geography have a greater influence than the use of chlorine-based cleaning agents on the microbiota of bulk tank raw milk." Applied and environmental microbiology 87(22): e01081-01021.

4.8 Supplemental Material

Supplemental table 4.1. Species assignment into 5 categories (environment, host, pathogen, spoilage or technologically relevant).

Species	Category	Average relative abundance (%)
Acinetobacter albensis	Environment	3.90
Acinetobacter bohemicus	Environment	0.16
Acinetobacter guillouiae	Environment	0.29
Acinetobacter johnsonii	Environment	0.42
Acinetobacter sp002135415	Environment	0.34
Arthrobacter_A sp002909415	Environment	0.38
CAG-791 sp900101015	Environment	0.24
Cellulosimicrobium aquatile	Environment	0.18
Moraxella_A aerosaccus	Environment	0.21
Pararhizobium sp001426685	Environment	0.39
Rothia sp002418375	Environment	0.24
Anaplasma phagocytophilum	Host	0.47
Murimonas intestini	Host	0.14
RUG420 sp900317985	Host	0.22
Kocuria atrinae	Host	0.16
Kocuria sp002295155	Host	0.19
Corynebacterium xerosis	Pathogen	0.49
Macrococcus caseolyticus	Pathogen	0.69
Staphylococcus aureus	Pathogen	2.13
Streptococcus suis_A	Pathogen	0.82
Streptococcus uberis	Pathogen	1.47
Trueperella pyogenes	Pathogen	0.76
Brochothrix thermosphacta	Spoilage	0.43
Flavobacterium frigidarium	Spoilage	1.46

Microbacterium sp001425645	Spoilage	1.08
Pseudomonas_E fluorescens_AH	Spoilage	0.48
Pseudomonas_E fluorescens_BA	Spoilage	0.33
Pseudomonas_E fragi	Spoilage	0.15
Pseudomonas_E fragi_B	Spoilage	1.46
Pseudomonas_E helleri	Spoilage	0.16
Pseudomonas_E lactis	Spoilage	0.23
Pseudomonas_E lundensis	Spoilage	0.71
Pseudomonas_E lurida	Spoilage	2.86
Pseudomonas_E proteolytica	Spoilage	0.58
Pseudomonas_E sp002966775	Spoilage	0.85
Pseudomonas_E sp900187495	Spoilage	1.62
Pseudomonas_E veronii	Spoilage	0.78
Psychrobacter sp002352555	Spoilage	0.96
Serratia proteamaculans_B	Spoilage	0.60
Bifidobacterium mongoliense	Technologically relevant	0.17
Bifidobacterium pseudolongum_A	Technologically relevant	0.45
Carnobacterium maltaromaticum	Technologically relevant	0.23
Lactococcus lactis	Technologically relevant	2.46
Lactococcus lactis_E	Technologically relevant	0.16
Lactococcus piscium_C	Technologically relevant	0.14
Lactococcus raffinolactis	Technologically relevant	0.86
Lactococcus raffinolactis_A	Technologically relevant	0.18
Leuconostoc lactis_A	Technologically relevant	0.42
Leuconostoc mesenteroides	Technologically relevant	0.32

Supplemental Table 4.2**.** High-quality MAGs, with taxonomy assigned using GTDB-tk with the quality determined by checkM.

Season	Location	GTDB-tk assignment*	Completeness	Contamination
spring	A	s__Brochothrix thermosphacta	99.23	0.08
spring	A	**s__Microbacterium maritypicum**	99.72	3.15
spring	A	s__Rothia sp002418375	99.16	0
spring	A	g__Acinetobacter	90.46	1.32
spring	A	s__Acinetobacter albensis	99.57	0
spring	A	s__Lactococcus_A raffinolactis	93.43	1.59
spring	B	s__Pararhizobium sp001426685	92.49	1.32
spring	B	s__Pararhizobium sp001426685	97.73	1.23
spring	B	**s__Pseudomonas_E bubulae**	95.21	0.89
spring	B	s__Rothia sp002418375	96.49	0.37
spring	B	s__Staphylococcus aureus	93.98	0.3
spring	B	**s__Lactococcus petauri**	97.92	0.31
spring	C	s__Pararhizobium sp001426685	99.49	1.01
spring	C	s__Pseudomonas_E proteolytica	96.69	0
spring	C	s__Trueperella pyogenes	95.46	0.24
spring	D	s__Leuconostoc lactis_A	98.62	3.83
spring	D	s__Pararhizobium sp001426685	96.13	0.56
spring	D	s__Acinetobacter albensis	99.31	0.11
spring	D	**s__Pseudomonas_E saxonica**	90.01	0.41
spring	D	**s__Lactococcus_A laudensis**	93.44	0.34
spring	D	s__Leuconostoc mesenteroides	92.58	0.7
spring	D	s__Lactococcus cremoris	94.82	1.3
spring	D	**s__Helcococcus ovis**	98.46	1.35
spring	D	s__Trueperella pyogenes	90.61	0.08
spring	D	s__Streptococcus uberis	97.88	0
spring	D	s__Leuconostoc mesenteroides	93.52	0.66
spring	E	s__Pararhizobium sp001426685	96.8	0.14
spring	E	s__Pararhizobium sp001426685	98.21	0.68
spring	E	s__Lactococcus_A raffinolactis	91.36	0.53

spring	E	s__Pseudomonas_E helleri_A	97.61	2.64
spring	E	s__Acinetobacter albensis	98.82	0.67
spring	F	s__Trueperella pyogenes	98.15	0
spring	F	s__Leuconostoc mesenteroides	99.63	0
spring	F	**s__Lactococcus_A laudensis**	96.91	0
spring	F	s__Leuconostoc mesenteroides	96.23	2.01
spring	F	s__Bifidobacterium mongoliense	95.71	0.14
spring	F	**s__Pseudomonas_E saxonica**	99.12	0
spring	F	s__Leuconostoc mesenteroides	94.53	0.11
spring	F	**s__Lactococcus_A laudensis**	92.43	0.38
spring	G	s__Pararhizobium sp001426685	95.22	0.27
spring	G	s__Leuconostoc mesenteroides	98.74	0
spring	G	s__Rothia sp002418375	98.11	1.01
spring	G	**s__Lactococcus_A laudensis**	92.61	0.31
spring	G	**s__Lactococcus cremoris**	94.65	0.46
spring	H	s__Rothia sp002418375	93.83	0.54
spring	H	**s__Pseudomonas_E bubulae**	92.43	0
spring	H	**s__Lactococcus_A laudensis**	95.85	1.89
spring	I	**s__Pseudomonas_E bubulae**	97.24	0
spring	I	**s__Lactococcus cremoris**	98.87	1.13
spring	I	s__Acinetobacter albensis	96.6	0.69
spring	I	s__Leuconostoc lactis_A	90.94	1.1
summer	A	s__Macrococcus_B caseolyticus	95.34	0.8
summer	A	s__Staphylococcus aureus	96.86	1.7
summer	A	s__Lactococcus_A raffinolactis	99.62	1.32
summer	A	s__Macrococcus_B caseolyticus	99.47	0.26
summer	A	s__Acinetobacter albensis	98.88	1.12
summer	C	s__Macrococcus_B caseolyticus	91.95	0.25
summer	C	**s__Lactococcus cremoris**	90.27	1.36
summer	D	**s__Lactococcus cremoris**	100	0.13
summer	D	**g__Specibacter**	96.13	0.61
summer	E	s__Macrococcus_B caseolyticus	90.18	0.4
summer	F	s__Flavobacterium frigidarium	93.38	0

summer	F	s__Macrococcus_B caseolyticus	92.68	1.64
summer	F	s__Flavobacterium frigidarium	97.56	0.41
summer	G	s__Leuconostoc mesenteroides	94.3	0.55
summer	G	s__Acinetobacter albensis	98.89	0.61
summer	G	s__Leuconostoc mesenteroides	98.82	0.53
summer	G	**s__Lactococcus cremoris**	96.92	0.45
summer	G	s__Acinetobacter albensis	94.15	0.82
summer	G	**s__Streptococcus dysgalactiae**	94.63	0.14
summer	G	**s__Lactococcus_A carnosus**	93.31	1.36
summer	H	**s__Lactococcus cremoris**	95.01	3.43
summer	H	**s__Lactococcus_A carnosus**	94.54	0.14
summer	H	s__Staphylococcus aureus	90.53	1.52
summer	H	s__Leuconostoc mesenteroides	100	0
summer	H	s__Carnobacterium maltaromaticum	100	0
summer	H	**s__Pseudomonas_E paracarnis**	90.86	1.03
summer	I	s__Macrococcus_B caseolyticus	90.03	0.11
summer	I	s__Lactococcus_A raffinolactis	90.51	0.44
summer	I	**s__Moraxella_A sp002478835**	93.38	0
summer	I	g__Lactococcus_A	96.02	0.32
summer	I	s__Acinetobacter albensis	99.62	1.7
summer	I	s__Streptococcus uberis	98.41	1.69
summer	I	s__Acinetobacter albensis	94.48	0.55
autumn	A	s__Acinetobacter albensis	94.48	0.55
autumn	C	s__Leuconostoc mesenteroides	98.69	2.85
autumn	C	**s__Lactococcus cremoris**	93.5	0.66
autumn	C	**g__Frigoribacterium**	95.6	3.08
autumn	D	**s__Lactococcus cremoris**	99.22	3.15
autumn	D	s__Acinetobacter albensis	97.79	0.18
autumn	E	s__Rothia sp002418375	99.63	1.79
autumn	F	s__Leuconostoc mesenteroides	99.36	0
autumn	F	**s__Epilithonimonas bovis**	98.87	0.38
autumn	F	s__Leuconostoc mesenteroides	96.76	0.82

autumn	F	**s__Lactococcus_A laudensis**	99.25	0.47
autumn	F	s__Bifidobacterium mongoliense	97.22	0.8
autumn	F	s__Flavobacterium frigidarium	93.12	0
autumn	F	**s__Psychrobacter immobilis_F**	99.76	3
autumn	F	**s__Pseudomonas_E bubulae**	93.07	0.67
autumn	F	s__Acinetobacter sp002135415	96.57	0
autumn	G	s__Leuconostoc mesenteroides	99.81	3.15
autumn	G	s__Leuconostoc lactis_A	99.25	0.75
autumn	G	s__Acinetobacter albensis	98.54	0
autumn	G	**s__Microbacterium maritypicum**	98.58	2.3
autumn	G	s__Acinetobacter albensis	97.03	0.55
autumn	H	**s__Lactococcus cremoris**	100	0
autumn	H	g__Acinetobacter	99.47	1.06
autumn	H	**s__Streptococcus agalactiae**	100	0.53
autumn	H	s__Acinetobacter albensis	94.36	0.61
autumn	I	**g__Porphyromonas_A**	100	0
autumn	I	s__Lactococcus lactis	97.25	0
autumn	I	s__Lactococcus_A raffinolactis	93.65	3.66
autumn	I	s__Leuconostoc mesenteroides	96.23	0.5
autumn	I	s__Macrococcus_B caseolyticus	98.11	0.19
autumn	I	**s__Streptococcus ruminantium**	97.55	0.98
autumn	I	**s__Microbacterium maritypicum**	99.44	0
autumn	I	s__Acinetobacter albensis	98.75	1.62
autumn	I	**s__Lactococcus cremoris**	91.47	0.47
autumn	I	s__Acinetobacter albensis	99.12	0.38
autumn	I	**s__Pseudomonas_E bubulae**	99.51	3
winter	A	g__Acinetobacter	96.13	1.1
winter	A	s__Acinetobacter albensis	98.94	0.53
winter	A	**s__Lactococcus cremoris**	97.44	0.53
winter	A	s__Acinetobacter albensis	95.02	0.27
winter	B	**s__Pseudomonas_E bubulae**	97.54	0.08
winter	C	**s__Pseudomonas_E bubulae**	97.36	1.19
winter	D	s__Acinetobacter albensis	92.42	0

winter	D	s__Rothia sp002418375	97.74	1.07
winter	F	s__Acinetobacter albensis	93.39	0.61
winter	F	**s__Lactococcus_A laudensis**	97.81	4.92
winter	F	s__Lactococcus_A raffinolactis	95.08	2.51
winter	F	s__Acinetobacter albensis	92.26	2.42
winter	F	**s__Leuconostoc rapi**	95.58	1.66
winter	F	s__Leuconostoc mesenteroides	99.5	0.38
winter	F	**s__Lactococcus_A laudensis**	93.07	0.58
winter	F	**s__Lactococcus_A laudensis**	95.53	0
winter	F	s__Flavobacterium frigidarium	97.34	0
winter	F	s__Leuconostoc mesenteroides	95.31	1.82
winter	G	s__Acinetobacter albensis	94.89	0.38
winter	G	s__Lactococcus_A raffinolactis	99.62	1.19
winter	G	s__Leuconostoc lactis_A	98.91	0.53
winter	G	s__Lactococcus_A raffinolactis	94.25	0.57
winter	G	s__Acinetobacter albensis	98.26	0.27
winter	H	**s__Pseudomonas_E bubulae**	98.74	1.13
winter	H	s__Streptococcus uberis	94.03	0
winter	I	s__Pseudomonas_E fragi	91.02	1.95
winter	I	s__Staphylococcus aureus	96.13	1.66
winter	I	**s__Pseudomonas_E bubulae**	95.58	0.55
winter	I	s__Acinetobacter albensis	95.64	1.26
winter	I	s__Brochothrix thermosphacta	98.31	0.36
winter	I	s__Acinetobacter albensis	95.28	0.5
winter	I	s__Rothia sp002418375	95.8	0
winter	I	s__Lactococcus_A raffinolactis	93.61	0.27

*Taxa in bold text indicates high-quality MAGs recovered that were not detected in the classification of short reads using Kraken2.

Supplemental Table 4.3. Relative abundances of genera and subsystem level 1 functions that differed significantly by sampling season.

Genera	Season			
	spring	summer	autumn	winter
Pseudomonas_E	22.77	5.41	12.02	31.90
Lactococcus	6.48	10.41	6.03	6.91
Rothia	1.70	4.63	2.57	0.51
Pararhizobium	5.93	0.01	0.00	0.00
Staphylococcus	0.90	1.59	1.18	1.13
Macrococcus	0.40	3.21	0.63	0.04
Streptococcus	0.87	1.46	1.06	1.01
Microbacterium	0.95	1.16	1.02	0.30
Psychrobacter	1.22	0.30	0.53	0.55
Bifidobacterium	0.96	0.34	0.41	0.75
Flavobacterium	0.68	1.01	0.17	0.49
Moraxella_A	0.23	1.44	0.49	0.04
Carnobacterium	0.30	0.75	0.21	0.77
Brochothrix	0.35	0.22	0.05	1.16
Anaplasma	0.25	0.86	0.26	0.05
CAG-791	0.39	0.11	0.39	0.46
Subsystem level 1 functions	Season			
	spring	summer	autumn	winter
Arabinose Sensor and transport module	0.000966	0.000779	0.000516	0.000696
Carbohydrates	13.424	14.229	13.361	12.409
Cell Wall and Capsule	4.9328	4.984	4.926	5.126
Clustering-based subsystems	5.737	5.627	5.596	5.616
Cofactors, Vitamins, Prosthetic Groups, Pigments	6.424	6.542	6.636	6.574
Dormancy and Sporulation	0.141	0.139	0.179	0.146

Fatty Acids, Lipids, and Isoprenoids	3.172	3.189	3.328	3.392
Nitrogen Metabolism	1.269	1.103	1.196	1.224
Phages, Prophages, Transposable elements	0.062	0.086	0.0628	0.0656
Potassium metabolism	1.064	0.974	1.045	1.169
Respiration	3.418	3.376	3.574	3.371
Transcriptional regulation	0.153	0.156	0.153	0.139
Virulence	4.643	4.208	4.438	4.506
Virulence, Disease and Defence	0.506	0.658	0.549	0.548

Supplemental Table 4.4. Relative abundances of genera and subsystem level 1 functions that differed significantly by sampling location.

Genera	Sampling location								
	A	B	C	D	E	F	G	H	I
Pseudomonas_E	17.43	10.85	14.29	16.24	10.29	16.09	20.05	32.59	15.02
Lactococcus	6.71	1.85	4.20	7.06	6.71	10.19	12.77	7.35	10.97
Acinetobacter	19.34	3.95	1.87	6.18	2.93	5.41	6.66	8.00	10.06
Leuconostoc	0.76	0.28	2.13	3.42	1.48	5.22	4.35	1.66	2.76
Pararhizobium	0.07	4.54	3.13	0.96	4.38	0.00	1.75	0.00	0.00
Macrococcus	2.92	0.11	1.05	0.55	2.30	1.41	0.29	0.32	1.76
Microbacterium	1.29	0.59	0.56	0.60	1.17	0.33	1.92	0.40	1.20
Corynebacterium	0.35	1.19	0.65	0.70	1.03	0.88	0.46	0.81	1.11
Kocuria	0.44	1.27	0.45	0.57	0.68	0.44	0.61	0.56	0.97
Bifidobacterium	0.26	0.73	0.36	0.74	0.53	1.66	0.29	0.40	0.53
Flavobacterium	1.77	0.03	0.02	0.02	0.02	2.49	0.06	0.49	0.56
Moraxella_A	0.84	0.16	0.37	0.47	0.78	0.64	0.42	1.22	0.47
Carnobacterium	1.08	0.10	0.17	0.25	0.38	0.42	0.92	0.64	0.40
Brochothrix	0.68	0.05	0.10	0.05	0.02	0.12	0.15	0.64	1.83
Anaplasma	0.21	0.58	0.70	0.49	0.45	0.33	0.32	0.20	0.10
CAG-791	0.14	0.66	0.31	0.28	0.46	0.22	0.22	0.27	0.39
Subsystem level 1 functions	Sampling location								
	A	B	C	D	E	F	G	H	I
Carbohydrates	12.01	13.91	13.79	13.72	14.15	13.59	13.76	12.04	13.75
Central metabolism	0.18	0.22	0.22	0.19	0.24	0.20	0.21	0.18	0.20
Clustering-based subsystems	5.67	5.72	5.60	5.55	5.82	5.54	5.60	5.74	5.62
Cofactors, Vitamins, Prosthetic Groups, Pigments	6.57	6.66	6.63	6.43	6.49	6.64	6.42	6.46	6.55
Dormancy and Sporulation	0.13	0.21	0.15	0.14	0.16	0.13	0.14	0.15	0.15

Fatty Acids, Lipids, and Isoprenoids	3.61	3.22	3.21	3.19	3.10	3.14	3.19	3.39	3.30
Metabolism of Aromatic Compounds	1.36	1.29	1.16	1.18	1.13	1.09	1.14	1.39	1.16
Miscellaneous	2.96	3.07	3.20	3.15	3.13	3.22	3.15	2.96	3.07
Nitrogen Metabolism	1.33	1.27	1.23	1.19	1.17	1.11	1.09	1.25	1.15
Nucleosides and Nucleotides	3.39	3.82	3.58	3.60	3.78	3.61	3.56	3.26	3.68
Phages, Prophages, Transposable elements	0.06	0.05	0.06	0.07	0.07	0.08	0.09	0.08	0.08
Phages, Prophages, Transposable elements, Plasmids	0.93	0.79	0.87	0.92	1.04	1.00	1.03	0.97	1.00
Photosynthesis	0.02	0.04	0.03	0.02	0.04	0.02	0.02	0.02	0.02
Protein Metabolism	8.00	8.63	8.78	8.39	8.42	8.40	8.19	7.55	8.42
Secondary Metabolism	0.04	0.06	0.06	0.05	0.05	0.04	0.04	0.05	0.04
Sulphur Metabolism	1.63	1.38	1.28	1.36	1.28	1.35	1.36	1.59	1.41
Transcriptional regulation	0.14	0.16	0.15	0.15	0.16	0.15	0.15	0.15	0.15
Virulence	4.80	4.13	4.23	4.65	4.21	4.50	4.50	4.65	4.41
Virulence, Disease and Defence	0.54	0.45	0.51	0.57	0.54	0.58	0.63	0.67	0.61

Chapter 5. Development of sequencing-based methodologies to distinguish viable from non-viable cells in a bovine milk matrix: a pilot study.

Included as published in Frontiers in Microbiology

(doi: 10.3389/fmicb.2022.1036643)

Authors: Min Yap, Paul W. O'Toole, Orla O'Sullivan and Paul D. Cotter

Contributions: Candidate performed sample preparation, DNA extraction, sequencing library preparation, bioinformatic and statistical analysis.
Candidate drafted and edited chapter.
All authors contributed, revised, and approved manuscript.

5.1 Abstract

Although high-throughput DNA sequencing-based methods have been of great value for determining the composition of microbial communities in various environments, there is the potential for inaccuracies arising from the sequencing of DNA from dead microorganisms. In this pilot study, we compared different sequencing-based methods to assess their relative accuracy with respect to distinguishing between viable and non-viable cells, using a live and heat-inactivated model community spiked into bovine milk. The methods used were shotgun metagenomics with and without propidium monoazide (PMA) treatment, RNA-based 16S rRNA sequencing and metatranscriptomics. The results showed that methods were generally accurate, though significant differences were found depending on the library types and sequencing technologies. Different molecular targets were the basis for variations in the results generated using different library types, while differences in the derived composition data from Oxford Nanopore Technologies- and Illumina-based sequencing likely reflect a combination of different sequencing depths, error rates and bioinformatics pipelines. Although PMA was successfully applied in this study, further optimisation is required before it can be applied in a more universal context for complex microbiomes. Overall, these methods show promise and represent another important step towards the ultimate establishment of approaches that can be applied to accurately identify live microorganisms in milk and other food niches.

5.1.1 Importance

High-throughput DNA sequencing has been useful in characterising microbial communities in various food-related environments. However, there is potential for misleading compositional data to be generated, because of the sequencing of DNA from dead microorganisms. The differentiation of live/dead cell states is also particularly important when making decisions on food safety or spoilage, when DNA from cells recently inactivated through food processing/preservation steps can be problematic. Overcoming this would enhance the power of DNA sequencing for food microbiology testing. In this study, 4 different sequencing-based methods were evaluated for their ability to differentiate between viable and non-viable cells spiked into a milk matrix. We identified differences between data derived by these methods due to differing molecular targets and sequencing technologies used. This study provides insight that will support the enhancement of sequencing-based approaches to better characterise live microbes in food-related environments.

5.2 Introduction

Microorganisms are ubiquitous and their presence and activity influence the niches that they inhabit. The use of high-throughput sequencing methods has expanded the understanding of microbiomes in clinical, food and environmental settings. However, the majority of the commonly used sequencing methods, such as amplicon or shotgun metagenomic sequencing, target (meta)genomic DNA (gDNA) in a manner that does not distinguish between DNA from microorganisms that are viable from those that non-viable microbes (Emerson, Adams et al. 2017). This is important as some studies have found that DNA from non-viable cells and other

extracellular DNA can persist in various environments, such as water, soil, food and the built environment, for days to weeks (Nocker and Camper 2006, Li, Tun et al. 2017). Not being able to distinguish between viable and non-viable microorganisms can lead to an overestimation of the relative abundance of particular taxa and/or the active metabolic processes that they encode (Carini, Marsden et al. 2016, Mancabelli, Milani et al. 2021). This, in turn, has potentially problematic implications when detecting spoilage or pathogenic microbes, determining bioburdens and the effectiveness of antimicrobial/cleaning treatments or processes (Carini, Marsden et al. 2016, Emerson, Adams et al. 2017). In a food setting, the accurate identification of live pathogenic or spoilage microbes is of considerable importance with respect to decision making processes relating to food safety, quality and product release. Therefore, the future widespread application of high-throughput sequencing as a tool by the food industry will be reliant on the ability to distinguish between live and dead cells in a food or food-related community.

Rapid culture-independent bacterial viability analysis is complex, and viability assays often query several different aspects such as cell membrane integrity, cellular metabolic activity or the presence of functional nucleic acids that allow for transcription/translation and DNA replication (Hammes, Berney et al. 2010, Emerson, Adams et al. 2017). Traditional culture-based techniques continue to be most extensively used in food microbiology testing and are the industry standard for detecting viable foodborne microorganisms (Foddai and Grant 2020). However, these methods can only detect a proportion of the viable microorganisms present in a sample as they rely on isolation and growth of a subset of microorganisms on

culture media, which is not usually reflective of the entire community and which can result in an underestimation of the levels of specific microorganisms (Gupta, Mortensen et al. 2019, Foddai and Grant 2020). To address issues associated with culture-based approaches, nucleic acid-based culture-independent methods have been studied and employed to detect viable microorganisms. Cell viability dyes, such as ethidium monoazide (EMA) or propidium monoazide (PMA), bind to accessible DNA that are not protected by a cell membrane after photoactivation and prevents further amplification of any non-intact DNA in downstream steps (Nocker and Camper 2006, Marotz, Morton et al. 2021, Wang, Yan et al. 2021). Methods involving the use of such dyes have been used with nucleic acid-based methods such as polymerase chain reaction (PCR), quantitative PCR (qPCR) and amplicon and metagenomic sequencing to detect viable microbes in clinical, environmental and food microbiomes (Carini, Marsden et al. 2016, Kable, Srisengfa et al. 2019, Marotz, Morton et al. 2021). Ribonucleic acid (RNA)-based techniques focused on the transcriptome have also been investigated with a view to distinguishing between live and dead cells in samples. As RNA has a shorter half-life than DNA, the use of RNA as a molecular target can more accurately represent the living/viable microbial population (Emerson, Adams et al. 2017, Gomez-Silvan, Leung et al. 2018). Indeed, the detection of messenger (mRNA) or ribosomal ribonucleic acid (rRNA) has been used to identify metabolically active cells in various samples (Mira Miralles, Maestre-Carballa et al. 2019, Yang, Cousineau et al. 2020, Zhou, Leung et al. 2020). Metatranscriptomics targets mRNA and provides information relating to the identity of the microorganisms that are metabolically active and the genes that these microbes are expressing (Chen, Chen et al. 2017).

The more targeted use of 16S rRNA has also been shown to have potential in identifying active bacteria in foods (Mira Miralles, Maestre-Carballa et al. 2019).

Here we investigate the relative success with which four sequencing-based methods distinguish between viable and non-viable bacteria in milk samples. A five-strain model community consisting of different proportions of live and dead cells was spiked into bovine milk samples before undergoing two DNA-based approaches, shotgun sequencing with and without prior PMA treatment, and two RNA-based approaches, metatranscriptomics and RNA-based 16S rRNA sequencing, on both representative Illumina (shorter read length; higher read numbers) and Oxford Nanopore Technologies (ONT) (longer read length; lower read numbers).

5.3 Materials and Methods

5.3.1 Model community and samples

Strains employed to subsequently create a model community were obtained from the Teagasc Dairy Production Centre (DPC) culture collection. These strains were *Bacillus velezensis* DPC 3313, *Escherichia coli* DPC 6912, *Lactococcus lactis* DPC 4268, *Pseudomonas proteolytica* DPC 6056 and *Staphylococcus haemolyticus* DPC 5987 and represent bacteria found in the bovine milk microbiome (Quigley, O'Sullivan et al. 2013). Strains were grown on tryptic soy agar (TSA; Thermo Fisher, Ireland) overnight at 30^{o}C (*B. velezensis* and *P. proteolytica*) or 37^{o}C (*E. coli*, *L. lactis* and *S. haemolyticus*) to generate fresh cultures. Cells were diluted in 10 ml phosphate-buffered saline (PBS; Thermo Fisher) to obtain a bacterial suspension of a density equivalent to approximately 10^7 cells/mL for each strain. Equal volumes of

strains were combined to generate a 5-strain model community for spiking into milk samples. This mixture was used as a live model community and a dead model community was generated through heat inactivation at 95^{o}C for 15 minutes (resulting in an absence of subsequent growth on TSA).

Store-bought UHT skim milk was used for milk spiking studies in order to ensure a low load of background microorganisms. 2 different viability conditions were used for spiking; live (consisting of 1 mL of the live-cell model community) and dead (consisting of 1 mL of the dead-cell model community). Cells were spiked into 10 mL of milk in triplicate and left to incubate at room temperature overnight. Unspiked milk samples and both the live and dead bacterial model communities, which did not undergo any treatment and were used for spiking, were used as controls. Milk samples were then centrifuged at 4,500 x g for 15 min at 4^{o}C, the supernatant was discarded, and the cells pellets were subjected to two washing steps whereby the pellets were resuspended in sterile PBS and centrifuged at 13,000 x g for 1 min, after which the supernatant was discarded. Samples used for Shotgun and PMA-Shotgun metagenomic analysis were resuspended in 1mL sterile PBS while, for samples used for metatranscriptomic- and 16S rRNA-based analysis, pellets were resuspended in 500 µL RNALater and stored at -80^{o}C before extraction.

5.3.2 PMA treatment and DNA extraction

Each 1 mL sample was divided into 2 x 500 µL samples, one of which was subjected to PMA (PMAxx dye, Biotium, CA, USA) treatment and the other without PMA treatment. Preliminary investigations were performed to optimise the final PMA

concentration, incubation and light exposure time with the quantification of total bacteria performed using the Femto Bacterial DNA Quantification kit (Zymo Research, CA, USA). The final parameters for PMA treatment involved the use of a final PMA concentration of 20 µM and incubation in the dark at room temperature and light exposure with the PMA-Lite™ LED photolysis device (Biotium, USA) for 30 min each. Following PMA treatment, both PMA-treated and PMA-free samples were subjected to DNA extraction using the MolYsis Complete5 kit (Molzym GmBH & Co. KG, Bremen, Germany), with 50 µL of DNA eluted for downstream sequencing. The MolYsis kit was used as it had been found to significantly improve microbial sequencing depth for milk samples (Yap, Feehily et al. 2020). gDNA was quantified using the Qubit dsDNA HS assay kit (Invitrogen) and stored at -20^{o}C before library preparation. Total bacteria levels were again quantified using the Femto Bacterial DNA Quantification kit (Zymo Research).

5.3.3 Illumina DNA library preparation and shotgun metagenomic sequencing

A total of 18 samples (6 PMA-treated, 6 non-PMA-treated, 6 controls) were prepared for shotgun metagenomic sequencing according to Illumina Nextera XT library preparation kit guidelines, with the use of unique dual indexes for multiplexing with the Nextera XT index kit (Illumina). Following indexing and clean up, samples were pooled to equimolar concentration of 1 nM. Samples were sequenced on an Illumina NextSeq 550 sequencing platform with a V2 kit, at the Teagasc DNA Sequencing Facility, using standard Illumina sequencing protocols.

5.3.4 MDA amplification and Oxford Nanopore DNA metagenomic library preparation

Whole metagenome amplification was performed on samples using multiple displacement amplification (MDA) with the REPLI-g UltraFast Mini kit (Qiagen, West Sussex, United Kingdom), according to the manufacturer's instructions. After amplification, DNA was quantified using the Qubit dsDNA BR assay kit (Invitrogen) and normalised to 400 ng in 7.5 µL. Two libraries were prepared, with each library containing 8 samples (6 samples (PMA-treated or non-PMA-treated) and 3 controls) per flowcell. The Rapid Barcoding kit (SQK-RBK004, Oxford Nanopore Technologies, United Kingdom) was used to prepare the libraries. For this, DNA was tagmented and barcodes were attached to fragments before pooling, followed by clean-up using Ampure XP beads (Beckman Coulter) to concentrate the pooled library directly prior to sequencing. DNA was sequenced using a GridION (release 19.12.6) with single flow cells for each library (R 9.4.1) that were primed prior to loading libraries with MinKNOW (core 3.6.5) and integrated basecalling by Guppy (3.2.10).

5.3.5 RNA extraction

Samples to be used for 16S rRNA and metatranscriptomic-based analysis were subjected to RNA extraction using a TRIzol chloroform protocol with on-column DNase purification with the PureLink RNA Mini kit (Thermo Fisher Scientific). Briefly, samples suspended in RNALater were first centrifuged at maximum speed (17,000 x g) for 1 min at 4^{o}C, after which the supernatant was discarded. One mL of TRIzol reagent was added to each sample and mixed through pipetting up and down to disperse the pellet, followed by incubation for 10 min at 60^{o}C. To each sample, 200

µL of chloroform was added and the tubes were shaken vigorously by hand for 15 secs before incubation at room temperature for 3 mins. The samples were then centrifuged at 12,000 x g for 15 min at 4^{o}C, to separate the mixture into a lower, red phenol-chloroform phase, an interphase, and a colourless upper aqueous phase containing the RNA. Approximately 400 µL of upper aqueous phase of RNA was transferred to a RNase-free tube where an equal volume of ice-cold 100% ethanol was added for precipitation and vortexed for even mixing. The samples were then purified through on-column DNase treatment following the manufacturer's protocol with final elution at 30 µL in RNase-free water. RNA concentration was quantified using the Qubit RNA HS assay kit (Invitrogen) and quality checked using an Agilent Bioanalyzer with the RNA 6000 pico assay kit (Agilent). RNA was stored at -80^{o}C for use in downstream library preparation steps.

5.3.6 cDNA synthesis

For 16S methods, 20 µL of RNA for each of the 9 samples (6 samples and 3 controls) were used for cDNA synthesis with the Superscript IV First Strand Synthesis System (Invitrogen) according to the manufacturer's instructions, using 2µM of the reverse primer of the V3-V4 region of the 16S rRNA gene that are commonly used in Illumina 16S sequencing. The cDNA generated was stored at -20^{o}C overnight before use in library preparation.

5.3.7 Illumina 16S-rRNA library preparation and sequencing

5 µl of the cDNA generated was used in library preparation according to Illumina 16S Metagenomic Sequencing Library Preparation guidelines, with the use of dual

indexes for multiplexing with the Nextera XT index kit (Illumina). Following indexing and clean up, samples were pooled to equimolar concentration of 20 nM. Samples were sequenced on an Illumina MiSeq sequencing platform with a V3 kit, at the Teagasc DNA Sequencing Facility, using Illumina sequencing protocols.

5.3.8 Oxford Nanopore 16S-rRNA library preparation and sequencing.

The cDNA generated was normalised to 10 ng in 10 µL for use in library preparation with the 16S Barcoding kit (SQK-RAB204, Oxford Nanopore Technologies) according to the manufacturer's instructions. Briefly, amplification of the 16S gene was performed using barcodes, followed by library clean up using Ampure XP beads (Beckman Coulter) and pooling to 10 ng/µL. Sequencing adapters were attached before the loading of libraries on a single primed flow cell (R 9.4.1) and sequenced on the GridION (release 19.12.6) with MinKNOW (core 3.6.5) and integrated basecalling by Guppy (3.2.10).

5.3.9 rRNA depletion

For metatranscriptomics, rRNA depletion was carried out using the QIAseq FastSelect – 5S/16S/23S kit (Qiagen) according to the manufacturer's instructions with 20 µL of each of the 9 samples (6 samples and 3 controls). The first step of RNA fragmentation was performed at 89^{o}C for 7 minutes. After bead clean up, rRNA-depleted RNA was stored at -80^{o}C before use in library preparation.

5.3.10 Illumina metatranscriptomics library preparation and sequencing.

First strand cDNA synthesis was performed using the NEBNext Ultra II Directional RNA Library Prep Kit (Brennan & Co., Dublin, Ireland) with 10 µL of rRNA-depleted RNA. Library preparation was performed according to the manufacturer's instructions. Seven cycles were used for PCR amplification during indexing of adapter ligated DNA. The quality and quantity were measured using a high sensitivity DNA chip on the Bioanalyzer (Agilent) and Qubit HS dsDNA kit (Invitrogen) respectively. An additional clean up step was performed due to the presence of primer and adapter dimers before the quality and quantity was measured. Samples were pooled to equimolar concentration of 4 nM before sequencing on the Illumina NextSeq 550 using a mid-output (2 x 75 bp kit) at the Teagasc DNA Sequencing Facility, using Illumina sequencing protocols.

5.3.11 Oxford Nanopore metatranscriptomics library preparation and sequencing.

rRNA-depleted RNA was normalised to 100ng in 7.5 µL for use in library preparation with the Direct cDNA Native Barcoding kit (SQK-DCS109, Oxford Nanopore Technologies) with the native barcoding expansion kit (EXP-NBD104, Oxford Nanopore Technologies) according to the manufacturer's instructions. Briefly, reverse transcription and strand switching was done to prepare full-length cDNAs followed by barcode ligation. Barcoded samples were pooled prior to adapter ligation before a final clean up with Ampure XP beads (Beckman Coulter) before loading the library onto a single primed flow cell (R 9.4.1) and sequenced on a GridION (release 19.12.6) with MinKNOW (core 3.6.5) and integrated basecalling by Guppy (3.2.10).

5.3.12 Bioinformatic analysis

Quality checks and adapter trimming for Illumina methods (shotgun, PMA-shotgun and metatranscriptomics) were done with FastQC (v. 0.11.8) (Andrews 2010) and cutadapt (v. 2.6) (Martin 2011) and host reads were aligned to the bovine genome (*Bos taurus*) and removed with Bowtie2 (v. 2.4.4) (Langmead and Salzberg 2012). Quality checks for ONT methods (shotgun, PMA-shotgun and metatranscriptomics) were done with MinIONQC (Lanfear, Schalamun et al. 2019) and Nanoplot (v. 1.28.2) (De Coster, D'Hert et al. 2018), with adapter trimming done with porechop (v. 0.2.4). Host reads were aligned with Minimap2 (v. 2.17-r974) to the bovine genome (*Bos taurus*) (Li 2018). Sortmerna (v. 2.1b) was used for rRNA removal prior to the analysis of metatranscriptomic data (Kopylova, Noé et al. 2012). Taxonomic classification for all metatranscriptomics, shotgun and PMA-shotgun methods was performed with Kraken2 (2.0.7) using the Genome Taxonomy Database (release 89) which contains Bacteria and Archaea (Parks, Chuvochina et al. 2018, Wood, Lu et al. 2019, Parks, Chuvochina et al. 2020, Parks, Chuvochina et al. 2022). For the Illumina 16S rRNA method, quality control, adapter trimming and denoising was done with Qiime2 (2021.2) with cutadapt and DADA2 (Martin 2011, Callahan, McMurdie et al. 2016, Bolyen, Rideout et al. 2019). For the ONT 16S rRNA method, analysis was performed using the q2ONT pipeline (https://github.com/DeniRibicic/q2ONT), which uses porechop (0.2.4) to demultiplex reads and trimmomatic (0.38) to discard reads shorter than 1400 bp and crop reads to that length. Qiime2 (2021.2) was used for the remaining steps, where sequences were dereplicated and chimeric sequences were filtered out with VSEARCH (2.3.4) at 85% identity before reads

were aligned with MAFFT (7.475) and highly variable positions were masked and filtered out. For both 16S rRNA methods, taxonomy was assigned with the Qiime2 trained classifier using the Silva 138 database (Quast, Pruesse et al. 2012).

5.3.13 Statistical analysis and data visualisation.

Performance metrics were calculated in R (4.1.2) (Team 2013), with the relative abundances of a model community strain correctly (TP) and incorrectly (FP) quantified by the method, along with the relative abundances of other taxa present in the sample correctly (TN) and incorrectly (FN) quantified. The following metrics were used:

(1) Accuracy is the correct quantification of model community and other taxa across all observations, A = TP + TN / TP + TN + FP + FN

(2) Precision is the fraction of correctly quantified model community strains within abundances of taxa identified as the model community, P = TP / TP + FP

(3) Sensitivity is the fraction of correctly quantified model community strains within abundances of the actual model community strains, S = TP / TP + FN

(4) F-score is the harmonic mean of precision and sensitivity, F = (2 x P x S) / (P + S)

Diversity analysis was done with the vegan package (Oksanen 2007), with beta diversity calculated as Bray-Curtis metrics, visualised in a principal coordinate analysis plot. The “adonis” function from the vegan package was used to calculate the permutational analysis of variance (PERMANOVA) to determine differences in composition of the community between groups of samples (number of permutations = 999). Hierarchical clustering was done with dendextend (Galili 2015)

using Bray-Curtis distances. Data was cleaned, analysed, and visualised in R with ggplot2, tidyverse and ggpubr packages (Wickham 2011, Wickham, Averick et al. 2019, Kassambara 2020).

5.4 Results

To determine the relative success with which sequencing-based methods could distinguish live and dead cells in milk, ultra-heat treatment (UHT) bovine milk was spiked with a 5-strain model community of representative bacteria found in milk (*Bacillus velezensis* DPC 3313, *Escherichia coli* DPC 6912, *Lactococcus lactis* DPC 428, *Pseudomonas proteolytica* DPC 6056 and *Staphylococcus haemolyticus* DPC 5987) at two different viability conditions, before undergoing different extraction and library preparation approaches and sequencing. Four methods, shotgun metagenomics with (PMA-shotgun) and without PMA treatment (Shotgun), metatranscriptomics (MetaT) and RNA-based 16S rRNA sequencing (16S rRNA), sequenced on both Illumina and Oxford Nanopore Technologies (ONT) platforms, were evaluated, and compared to determine their relative efficacy in differentiating viable and non-viable cells (Figure 5.1). Sequencing depth differed depending on the platform and the application (Supplemental Table 5.1). We did not endeavour to sequence at matching depths as the focus was applying the technologies at the depth corresponding to those at which the respective platforms are typically used.

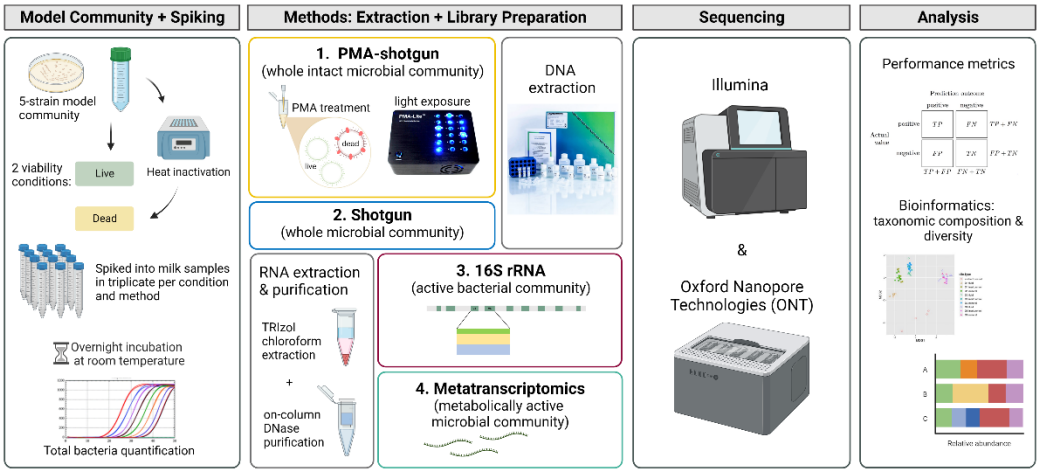


Figure 5.1 Experimental design of study (created with BioRender.com).

5.4.1 Taxonomic classification of model community strains was accurate to genus level across all methods.

In general, the model community taxonomy was determined with one classifier to ensure consistency across methods. This classifier, Kraken2, was previously found to efficiently characterise the milk microbiome when compared to other classifiers (Yap, Feehily et al. 2020). In addition, a trained classifier in Qiime2 was used for the analysis of 16S rRNA data (Illumina-sequenced 16S rRNA: i-16s and ONT-sequenced 16S rRNA: o-16s). Comparing across methods, the taxonomic identity of the model community was consistent to genus level in all cases (Table 5.1). This was also the case at species level, where assignment at species level was possible, with the exception that reads generated from Illumina sequencing were assigned as *Staphylococcus hominis*, while the same model community strain was classified as a *Staphylococcus haemolyticus* when ONT data was analysed. However, this is not unexpected as *S. hominis* is classified under the *S. haemolyticus* group of bacteria (Becker, Heilmann et al. 2014). ONT-generated metatranscriptomic data and the

short read length 16S rRNA data (regardless of sequencing platform) could not be resolved beyond genus level. Indeed, *Staphylococcus* was not detected from within the ONT-derived metatranscriptomic data, most likely due to insufficient sequencing depth combined with low levels of gene expression by the strain (Supplemental Table 5.1). For downstream analysis, genus level data was used.

Total bacteria qPCR was performed to quantify the bacteria in each sample and the log number of cells/mL of each microorganism per sample was estimated using a combination of qPCR results and relative abundance values. The abundances of cells recovered from Live samples were greater than those in the heat-treated samples (Supplemental Figure 5.1). Some reads were classified into genera other than those corresponding to the 5 spiked strains (Supplemental Table 5.2), and some of these were also detected in the unspiked milk samples (Supplemental Tabl 5.3). The reads classified as others were not prioritised in the further analysis due to the particular focus on differentiating between the members of the model community in the Live and Dead samples.

Table 5.1. The highest resolution of taxonomic identity identified for the model community strains across all methods.

Model community strain	Shotgun		PMA-shotgun		Metatranscriptomics		16S rRNA	
	Illumina	Nanopore	Illumina	Nanopore	Illumina	Nanopore	Illumina	Nanopore
Bacillus velezensis DPC 3313	*Bacillus velezensis*	*Bacillus velezensis*	*Bacillus velezensis*	*Bacillus velezensis*	*Bacillus velezensis*	*Bacillus*	*Bacillus*	*Bacillus*
Escherichia coli DPC 6912	*Escherichia coli*	*Escherichia coli*	*Escherichia coli*	*Escherichia coli*	*Escherichia coli*	*Escherichia*	*Escherichia-Shigella*	*Escherichia-Shigella*
Lactococcus lactis DPC 4268	*Lactococcus lactis*	*Lactococcus lactis*	*Lactococcus lactis*	*Lactococcus lactis*	*Lactococcus lactis*	*Lactococcus*	*Lactococcus*	*Lactococcus*
Pseudomonas proteolytica DPC 6056	*Pseudomonas_E proteolytica*	*Pseudomonas_E proteolytica*	*Pseudomonas_E proteolytica*	*Pseudomonas_E proteolytica*	*Pseudomonas_E proteolytica*	*Pseudomonas_E*	*Pseudomonas*	*Pseudomonas*
Staphylococcus haemolyticus DPC 6283	*Staphylococcus hominis*	*Staphylococcus haemolyticus*	*Staphylococcus hominis*	*Staphylococcus haemolyticus*	*Staphylococcus hominis*	-	*Staphylococcus*	*Staphylococcus*

5.4.2 Methods were accurate though the precision and sensitivity varied.

To evaluate the performance of each method, several performance metrics such as accuracy, precision, sensitivity, and F-score were calculated. The comparisons were made between the abundances of the live and dead model community controls that had been sequenced and analysed with the various methods. Overall, most approaches showed a high level of accuracy, which is the fraction of correctly identified taxa (model community or other) relative to all observations, with a median of 0.939 ± 0.046 (Figure 5.2). However, the F-score, which gives a sense of how precise and sensitive the method is, revealed that only half of the methods scored at least 0.5, indicating issues in precision or sensitivity in these instances. The two PMA-shotgun sequencing and analysis methods yielded outputs that were most precise, while the 16S rRNA methods were the most sensitive, including when separated into both live and dead spiked samples (Figure 5.2, Supplemental Figure 5.2). ONT-sequenced 16S rRNA (o-16s) was the most accurate and had the best overall F-score, while metatranscriptomics (i-metaT) and PMA-shotgun (i-pma) had the highest accuracy and F-score out of all the Illumina-based methods, respectively.

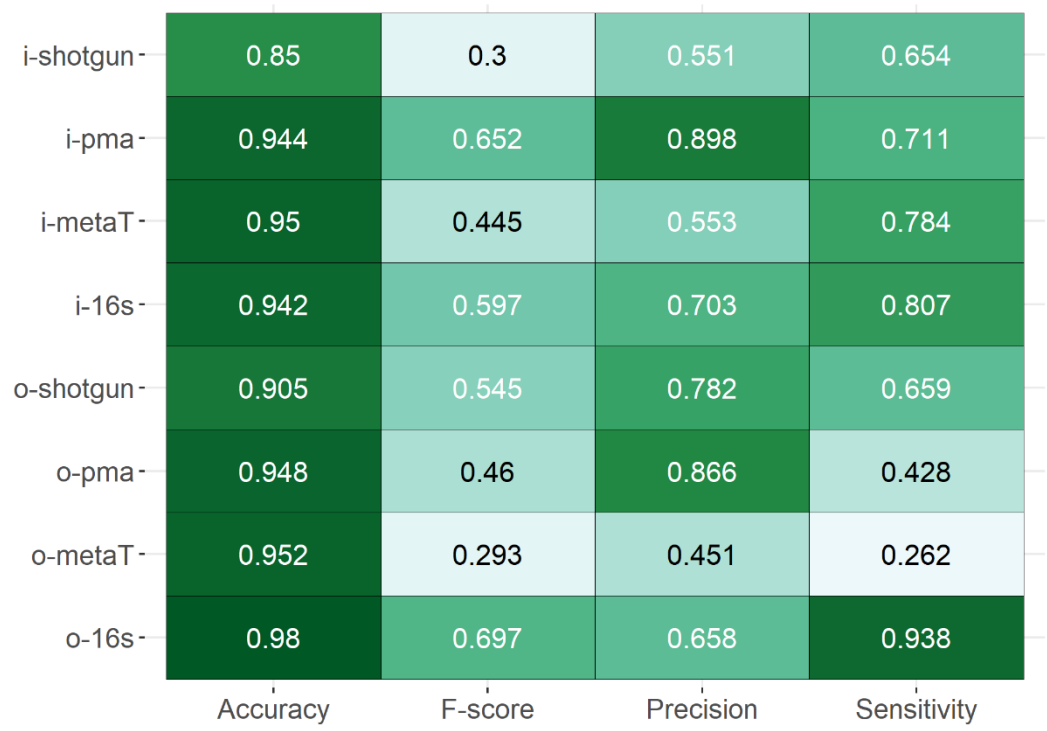


Figure 5.2. Overall performance of methods based on the accuracy, F-score, precision, and sensitivity.

5.4.3 Significant differences were observed between data depending on library types (DNA and RNA) and sequencing methods (Illumina and ONT).

Beta diversity analysis was employed to determine the differences between the outputs generated from the different approaches taken, and several sample clusters were apparent (Figure 5.3a). Samples first clustered by library type, whereby the outputs from DNA-based analysis (Shotgun and PMA-shotgun) were significantly distinct from those generated using the RNA-based approaches (metatranscriptomics and 16S rRNA) (PERMANOVA, $R^2 = 0.457$, $P < 0.01$), and also by sequencing methods, with Illumina-sequenced and ONT-sequenced outputs clustering apart (PERMANOVA, $R^2 = 0.195$, $P < 0.01$). Differences in nucleic acid

input and sequencing methods used were also apparent in the hierarchical clustering of the Bray-Curtis distances (Figure 5.3b). The composition of taxa varied across library types, with *Escherichia* and *Lactococcus* being found at greater abundance when DNA-based methods were used, while *Pseudomonas* was the most active on the basis of RNA-based methods (Figure 5.3c). In particular, the Illumina-sequenced 16S rRNA data points clustered away from those generated using other methods. This was also obvious in the taxonomic composition where abundances of other taxa besides the model community was lowest in i-16S compared to other methods (Figure 5.3c).

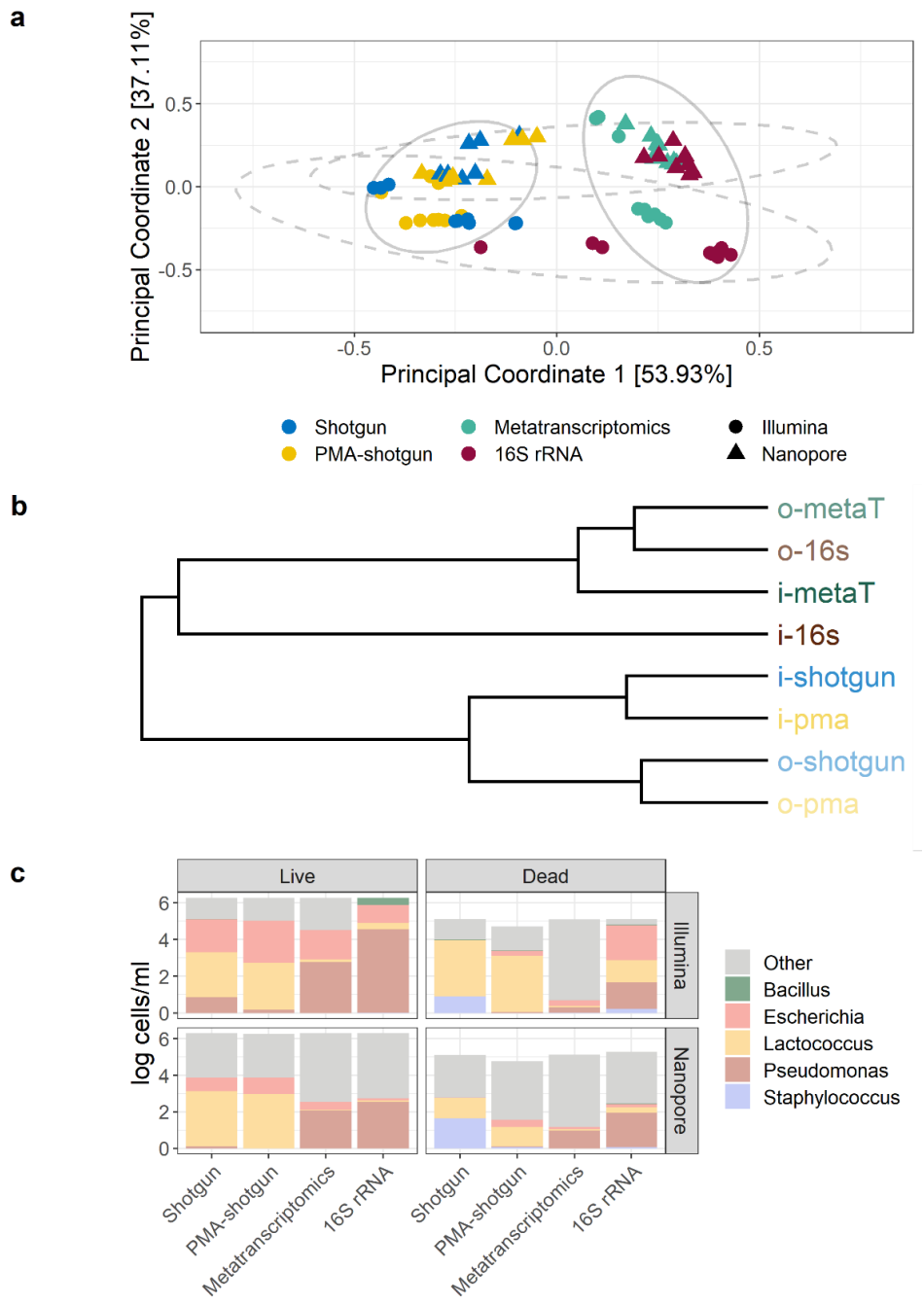


Figure 5.3. Beta diversity and taxonomic profiles of samples. (a) Bray-Curtis principal-coordinate analysis (PCoA) plot, with ellipses representing clustering by DNA- and RNA-based methods (solid) and sequencing method (dashed) and (b) hierarchical clustering of methods based on Bray-Curtis distances. (c) Composition of taxa in samples based on live or dead spiked model community, by the evaluated methods and sequencing method.

Among the DNA-generated dataset, the effect of PMA was clear, with lower DNA and overall calculated cell counts observed for samples that underwent PMA treatment (Figure 5.4). Notably, although the Shotgun methods do not exclusively select for viable cells, the accuracy of those methods were comparable to the PMA-shotgun methods (Figure 5.2). Precision was higher in PMA-shotgun methods compared to the Shotgun methods, while sensitivity was higher for i-pma and lower for o-pma when compared to i-shotgun and o-shotgun, respectively. In terms of the composition, most of the abundances of model community taxa in PMA-shotgun samples were lower in total cells/mL than those in Shotgun samples, apart from *Escherichia* cells/mL, which was higher in PMA-shotgun methods than Shotgun for both live and dead spiked samples (Figure 5.3c). The cell counts of *Bacillus* and *Pseudomonas* were generally lower, with the exception of i-shotgun where both taxa were found in abundances greater than the other DNA-based methods. *Lactococcus* was the most abundant of the model community taxa quantified, which was consistent across the two methods, with exception of o-shotgun, which detected higher abundances of *Staphylococcus*. *Staphylococcus* was clearly detected in the Shotgun samples spiked with the dead cells whereas levels were negligible in the PMA-shotgun samples. Between sequencing methods, distinct separation was found between Illumina and ONT methods (Figure 5.3a, b) and, in terms of composition, ONT methods detected higher abundances of other non-model community taxa, compared to the Illumina methods (Figure 5.3c).

Although there were no obvious differences in the accuracy of the methods across the different RNA-based approaches, the 16S rRNA method outperformed the

metatranscriptomics approaches with respect to precision and sensitivity (Figure 5.2). In general, methods were more sensitive than precise, except for o-metaT which had poor sensitivity. *Staphylococcus* reads were detected in low abundances across all RNA-based methods, while read numbers corresponding to *Pseudomonas* were consistently found at high abundance (Figure 5.3c). Similar to the DNA-based methods, clear separation between library types was seen between Illumina- and ONT-based methods (Figure 5.3). As mentioned above, Illumina-sequenced 16S rRNA was the most distinct, clustering separately from other methods and with the lowest proportion of other taxa identified compared to the other RNA-based methods (Figure 5.3a, b). It was the only method that had detectable levels of *Bacillus* in live samples and its samples spiked with the dead-cell model community had a taxonomic profile that differed from the other samples spiked with dead cell concentrations (Figure 5.3c).

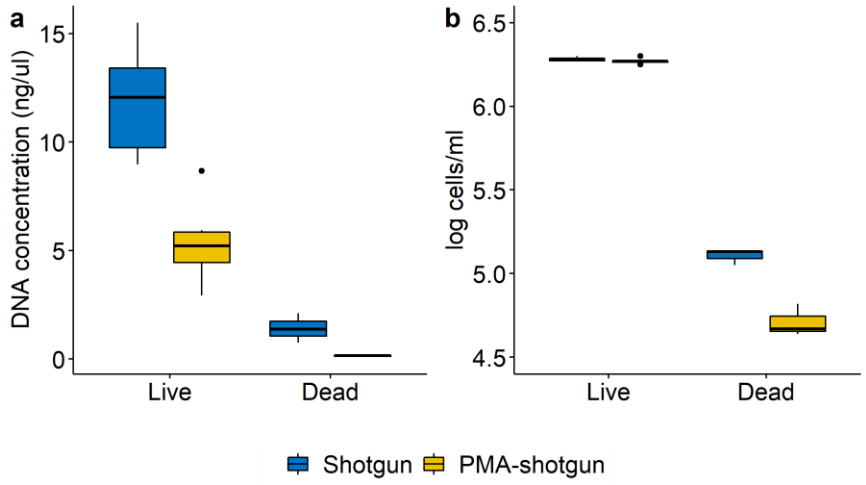


Figure 5.4. Differences in concentrations of (a) DNA quantified using Qubit and (b) cells quantified through qPCR between Shotgun and PMA-shotgun methods for live and dead spiked samples.

5.5 Discussion

While DNA sequencing-based approaches have enabled the characterisation of complex microbial communities more efficiently, they generally provide information that reflects all microbial DNA present in the community and do not specifically select for viable microorganisms (Carini, Marsden et al. 2016, Wang, Yan et al. 2021). In various settings, the ability to distinguish between viable or metabolically active and non-viable or dead microorganisms can be important and, indeed, necessary in terms of food safety and public health, and for food processing involving live microorganisms. While culture-based approaches are commonly used by industry and public health authorities to determine viability of individual microbes (Emerson, Adams et al. 2017), high-throughput sequencing-based methods have great potential as a means of characterising microbial communities in foods or food environments. In this study, we explored 4 sequencing-based approaches to assess the relative success with which they distinguish between live and dead cells of a 5-strain model community in a milk matrix.

All methods performed accurately, despite the slight compositional differences between the two library types. The cause of the difference could possibly relate to the approach of each library type, with the use of PMA or other intercalating dyes taking advantage of cell integrity to remove cells with a compromised membrane (Emerson, Adams et al. 2017, Wang, Yan et al. 2021), while RNA-based methods examine the activity of cells, as RNA has a shorter average half-life than DNA (Emerson, Adams et al. 2017, Li, Tun et al. 2017, Gomez-Silvan, Leung et al. 2018, Yang, Cousineau et al. 2020). This implies that *Pseudomonas* were most active,

while *Escherichia* and *Lactococcus* had the greatest abundance of intact cells. The different targets and the different library preparation methods (amplicon, shotgun and metatranscriptomics) make it difficult to directly compare the methods with each other. Microbial community studies with both types of libraries (DNA and RNA) often perform a single molecular method on the samples, such as 16S rRNA sequencing or qPCR, to allow for comparison of results from DNA and RNA libraries. In these studies, consistent outcomes have been reported in that distinctions have been found between library types in water, sediment, and food samples when both DNA and RNA were extracted and sequenced from the same samples (Ferrocino, Greppi et al. 2016, Li, Tun et al. 2017, Mira Miralles, Maestre-Carballa et al. 2019, Pearman, Biessy et al. 2021).

Our data also reveal clear differences between the outputs of Illumina- and ONT-based sequencing technologies, with samples clustering distinctly based on sequencing technologies used. Differences between sequencing technologies have also been reported in intestinal and nasal microbiota samples with species level classification and the clustering of the microbiome showing significant effects based on the sequencing technology utilised (Heikema, Horst-Kreft et al. 2020, Alili, Belda et al. 2021). Besides the known higher error rate of ONT sequencing compared to Illumina (Alili, Belda et al. 2021), a few other possibilities could account for this. First, in order to maintain consistency during analysis, bioinformatic tools that are more commonly applied to short read data were used for all the sequencing data generated and, thus, specific tools developed for long read sequencing data analysis were not used. It is possible that improved classification could be achieved

should the relevant tools for error correction for long reads be applied to the ONT sequencing data. Another contributory factor is the difference in sequencing depth. ONT sequencing had been previously found to produce a high proportion of unclassified reads when a shallow sequencing depth was employed, and greater sequencing depth (30-50X) has been known to enable self-correction of the error rate that is associated with Nanopore sequencing data (Alili, Belda et al. 2021, Delahaye and Nicolas 2021). These factors could account for the numerical difference in species level classification of the *Staphylococcus* strain, where Illumina sequencing afforded greater taxonomic resolution than ONT sequencing, as S. *hominis* is specifically classified within the *S. haemolyticus* group (Becker, Heilmann et al. 2014). Additionally, for shotgun sequencing with ONT, MDA amplification was needed to generate sufficient quantities of DNA required for sequencing, which has been previously found to introduce some biases (McHugh, Yap et al. 2021). Despite all these points, as improvements are made to the sequencing technology and analytical tools, coupled with its advantages in portability, time effectiveness and the possibility of real-time analysis, there is great potential for the use of ONT sequencing in characterising microbial communities in food and food-related environments.

PMA is a photoreactive DNA-binding dye that binds to non-viable cells that have compromised membranes and inhibits further amplification by PCR (Nocker and Camper 2006). Our results show evidence of the effect of PMA, with more precise results and lower abundances of live and dead cells achieved when compared to regular shotgun sequencing. The use of PMA did not bias the community structure,

which adds to their promising application. However, studies have found that the effectiveness of PMA varies significantly (Mancabelli, Milani et al. 2021, Wang, Yan et al. 2021). Biological matrix, sample biomass, microbial community diversity and experimental conditions have been known to contribute to this variation (Li, Tun et al. 2017, Mancabelli, Milani et al. 2021, Shen, McFarland et al. 2021, Wang, Yan et al. 2021). The optimisation of dye concentration, incubation and light exposure times are required depending on the samples, which prevents the application of a universal PMA treatment for all types of samples. Moreover, methods based on membrane integrity may result in an overestimation of viable cells as the lethal stress may not lead to the immediate disruption of the cell membrane and dyes used may be ineffective against cells with a hardy cell wall, such as spores (Emerson, Adams et al. 2017). Although the close clustering of PMA-shotgun to Shotgun samples could suggest the insufficient or unsuccessful PMA treatment, the low diversity samples created by spiking the simple model community into milk samples could be the reason for such results. Regardless, this further emphasises the need to optimise PMA treatment, which could differ between samples. Therefore, though promising, more work is needed to calibrate PMA concentration and incubation and light exposure times for specific sample matrices before it can be effectively applied to DNA sequencing experiments with confidence.

Within the data generated by RNA-based methods, results differed between metatranscriptomics and 16S rRNA sequencing, which is not surprising as they use different types of RNA as targets. Metatranscriptomics targets mRNA, which has a very short half-life, while 16S rRNA sequencing uses rRNA, which is generally more

stable than mRNA and is more abundant in cells (Emerson, Adams et al. 2017). In spite of these differences, both mRNA and rRNA have been found to be good markers of bacterial viability, with several studies suggesting that rRNA might provide more accurate taxonomic profiling and may be more successful for low-biomass samples compared to mRNA approaches (Emerson, Adams et al. 2017, Yang, Cousineau et al. 2020). Additionally, the cells in both model communities may not have the same physiological state after overnight incubation, which could affect their metabolic activity and the resulting differences between the two RNA-targeting methods. Futhermore, cDNA conversion and the different library preparation steps could have introduced biases that can contribute to the observed differences (Li, Tun et al. 2017, Sessegolo, Cruaud et al. 2019). The targeted nature of 16S rRNA sequencing is possibly the reason that the Illumina-sequenced 16S method is the most sensitive and distinct from all the methods (Fig 3). As amplicon sequencing is more widely established and applied, analysis tools for such data are well refined compared to other community-based methods like metatranscriptomics. In spite of this, several studies have employed both RNA- and DNA-based 16S rRNA sequencing concurrently to give a more comprehensive view of both the overall and metabolically active community, which is useful in understanding microbial community dynamics (Gomez-Silvan, Leung et al. 2018, Mira Miralles, Maestre-Carballa et al. 2019). However, from a practical perspective, RNA-based methods are generally more laborious than DNA-based methods, as RNA is more complicated to handle, which can result in RNA losses during processing (Emerson, Adams et al. 2017, Li, Tun et al. 2017, Marcelino, Irinyi et al. 2019). In addition to the need for a greater sequencing depth, the

metatranscriptomics methods require more processing that could introduce additional biases or losses, which possibly impacted the accuracy of results in our study. Therefore, metatranscriptomics, for now, might not be the most suitable method if the aim is to determine viability alone. It, however, is able to provide useful information on the active microbes and expressed genes that would expand the understanding of microbial community dynamics of foods, food processes or environments (Chen, Chen et al. 2017).

Though this study provided insight into sequencing-based methods that could be employed to distinguish between viable and non-viable microbial communities, there were some limitations. First, overnight incubation caused the growth of the model community to levels beyond the expected equal concentrations, causing greater difficulties in determining the success of methods and hindering the analysis of the potential of each method in quantifying live and dead cells. While the kits used were based on prior in-house experience, simultaneous DNA and RNA extractions could have been performed, which would have standardised the laboratory workflow. The use of MDA for ONT DNA-based samples was necessary to obtain sufficient DNA yields for sequencing but may be a confounding factor in the analysis. Lastly, DNA-based 16S rRNA sequencing could have been done to allow for better comparison with the RNA-based method and to help draw further conclusions of the differences between library types.

Despite these challenges, this study gives insight into the use of sequencing-based methods in distinguishing viable and non-viable cells in food samples. While it was

difficult to make direct comparisons between the methods due to the differences in their molecular targets, this study shows that differences exist between library types and sequencing technologies, which serves as a stepping-stone to refining these methods. Alongside general optimisation of PMA treatments, more complex matrices and more complex microbial communities can be used to test the performance of these methods since the milk samples used had a low microbial density and diversity. Thus, further research is warranted before using these methods to characterise viable microbial communities in food and food-related environments.

5.6 Declarations

5.6.1 Data availability

Sequence data generated during the current study have been deposited in the European Nucleotide Archive under the accession number PRJEB53378.

5.6.2 Acknowledgements including funding source

Research was conducted with the financial support of the Irish Dairy Levy. The authors thank Laura Finnegan for advice on working with RNA, and Fiona Crispie and Elaine Lawton for their help with DNA sequencing.

5.7 References

Alili, R., E. Belda, P. Le, T. Wirth, J.-D. Zucker, E. Prifti and K. Clément (2021). "Exploring Semi-Quantitative Metagenomic Studies Using Oxford Nanopore Sequencing: A Computational and Experimental Protocol." Genes 12(10): 1496.

Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data, Babraham Bioinformatics, Babraham Institute, Cambridge, United Kingdom.

Becker, K., C. Heilmann and G. Peters (2014). "Coagulase-negative staphylococci." Clinical microbiology reviews 27(4): 870-926.

Bolyen, E., J. R. Rideout, M. R. Dillon, N. A. Bokulich, C. C. Abnet, G. A. Al-Ghalith, H. Alexander, E. J. Alm, M. Arumugam and F. Asnicar (2019). "Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2." Nature biotechnology 37(8): 852-857.

Callahan, B. J., P. J. McMurdie, M. J. Rosen, A. W. Han, A. J. A. Johnson and S. P. Holmes (2016). "DADA2: High-resolution sample inference from Illumina amplicon data." Nature methods 13(7): 581-583.

Carini, P., P. J. Marsden, J. W. Leff, E. E. Morgan, M. S. Strickland and N. Fierer (2016). "Relic DNA is abundant in soil and obscures estimates of soil microbial diversity." Nature microbiology 2(3): 1-6.

Chen, G., C. Chen and Z. Lei (2017). "Meta-omics insights in the microbial community profiling and functional characterization of fermented foods." Trends in Food Science & Technology 65: 23-31.

De Coster, W., S. D'Hert, D. T. Schultz, M. Cruts and C. Van Broeckhoven (2018). "NanoPack: visualizing and processing long-read sequencing data." Bioinformatics 34(15): 2666-2669.

Delahaye, C. and J. Nicolas (2021). "Sequencing DNA with nanopores: Troubles and biases." PloS one 16(10): e0257521.

Emerson, J. B., R. I. Adams, C. M. B. Román, B. Brooks, D. A. Coil, K. Dahlhausen, H. H. Ganz, E. M. Hartmann, T. Hsu and N. B. Justice (2017). "Schrödinger's microbes: tools for distinguishing the living from the dead in microbial ecosystems." Microbiome 5(1): 1-23.

Ferrocino, I., A. Greppi, A. La Storia, K. Rantsiou, D. Ercolini and L. Cocolin (2016). "Impact of nisin-activated packaging on microbiota of beef burgers during storage." Applied and environmental microbiology 82(2): 549-559.

Foddai, A. C. and I. R. Grant (2020). "Methods for detection of viable foodborne pathogens: Current state-of-art and future prospects." Applied Microbiology and Biotechnology 104(10): 4281-4288.

Galili, T. (2015). "dendextend: an R package for visualizing, adjusting and comparing trees of hierarchical clustering." Bioinformatics 31(22): 3718-3720.

Gomez-Silvan, C., M. H. Leung, K. A. Grue, R. Kaur, X. Tong, P. K. Lee and G. L. Andersen (2018). "A comparison of methods used to unveil the genetic and metabolic pool in the built environment." Microbiome 6(1): 1-16.

Gupta, S., M. S. Mortensen, S. Schjørring, U. Trivedi, G. Vestergaard, J. Stokholm, H. Bisgaard, K. A. Krogfelt and S. J. Sørensen (2019). "Amplicon sequencing provides more accurate microbiome information in healthy children compared to culturing." Communications biology 2(1): 1-7.

Hammes, F., M. Berney and T. Egli (2010). "Cultivation-independent assessment of bacterial viability." High resolution microbial single cell analytics: 123-150.

Heikema, A. P., D. Horst-Kreft, S. A. Boers, R. Jansen, S. D. Hiltemann, W. de Koning, R. Kraaij, M. A. de Ridder, C. B. van Houten and L. J. Bont (2020). "Comparison of Illumina versus nanopore 16S rRNA gene sequencing of the human nasal microbiota." Genes 11(9): 1105.

Kable, M. E., Y. Srisengfa, Z. Xue, L. C. Coates and M. L. Marco (2019). "Viable and total bacterial populations undergo equipment-and time-dependent shifts during milk processing." Applied and environmental microbiology 85(13): e00270-00219.

Kassambara, A. (2020). "ggpubr:“ggplot2” based publication ready plots." R package version 0.4. 0 438.

Kopylova, E., L. Noé and H. Touzet (2012). "SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data." Bioinformatics 28(24): 3211-3217.

Lanfear, R., M. Schalamun, D. Kainer, W. Wang and B. Schwessinger (2019). "MinIONQC: fast and simple quality control for MinION sequencing data." Bioinformatics 35(3): 523-525.

Langmead, B. and S. L. Salzberg (2012). "Fast gapped-read alignment with Bowtie 2." Nature methods 9(4): 357-359.

Li, H. (2018). "Minimap2: pairwise alignment for nucleotide sequences." Bioinformatics 34(18): 3094-3100.

Li, R., H. M. Tun, M. Jahan, Z. Zhang, A. Kumar, W. Dilantha Fernando, A. Farenhorst and E. Khafipour (2017). "Comparison of DNA-, PMA-, and RNA-based 16S rRNA Illumina sequencing for detection of live bacteria in water." Scientific reports 7(1): 1-11.

Mancabelli, L., C. Milani, R. Anzalone, G. Alessandri, G. A. Lugli, C. Tarracchini, F. Fontana, F. Turroni and M. Ventura (2021). "Free DNA and Metagenomics Analyses: Evaluation of Free DNA Inactivation Protocols for Shotgun Metagenomics Analysis of Human Biological Matrices." Frontiers in Microbiology: 2905.

Marcelino, V. R., L. Irinyi, J.-S. Eden, W. Meyer, E. C. Holmes and T. C. Sorrell (2019). "Metatranscriptomics as a tool to identify fungal species and subspecies in mixed communities–a proof of concept under laboratory conditions." IMA fungus 10(1): 1-10.

Marotz, C., J. T. Morton, P. Navarro, J. Coker, P. Belda-Ferre, R. Knight and K. Zengler (2021). "Quantifying live microbial load in human saliva samples over time reveals stable composition and dynamic load." Msystems 6(1): e01182-01120.

Martin, M. (2011). "Cutadapt removes adapter sequences from high-throughput sequencing reads." EMBnet. journal 17(1): 10-12.

McHugh, A. J., M. Yap, F. Crispie, C. Feehily, C. Hill and P. D. Cotter (2021). "Microbiome-based environmental monitoring of a dairy processing facility highlights the challenges associated with low microbial-load samples." NPJ science of food 5(1): 1-13.

Mira Miralles, M., L. Maestre-Carballa, M. Lluesma-Gomez and M. Martinez-Garcia (2019). "High-throughput 16S rRNA sequencing to assess potentially active bacteria and foodborne pathogens: a case example in ready-to-eat food." Foods 8(10): 480.

Nocker, A. and A. K. Camper (2006). "Selective removal of DNA from dead cells of mixed bacterial communities by use of ethidium monoazide." Applied and environmental microbiology 72(3): 1997-2004.

Oksanen, J. (2007). "Vegan: community ecology package. R package version 1.8-5." http://www. cran. r-project. org.

Parks, D. H., M. Chuvochina, P.-A. Chaumeil, C. Rinke, A. J. Mussig and P. Hugenholtz (2020). "A complete domain-to-species taxonomy for Bacteria and Archaea." Nature biotechnology 38(9): 1079-1086.

Parks, D. H., M. Chuvochina, C. Rinke, A. J. Mussig, P.-A. Chaumeil and P. Hugenholtz (2022). "GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy." Nucleic acids research 50(D1): D785-D794.

Parks, D. H., M. Chuvochina, D. W. Waite, C. Rinke, A. Skarshewski, P.-A. Chaumeil and P. Hugenholtz (2018). "A standardized bacterial taxonomy based on genome

phylogeny substantially revises the tree of life." Nature biotechnology 36(10): 996-1004.

Pearman, J. K., L. Biessy, J. D. Howarth, M. J. Vandergoes, A. Rees and S. A. Wood (2021). "Deciphering the molecular signal from past and alive bacterial communities in aquatic sedimentary archives." Molecular ecology resources.

Quast, C., E. Pruesse, P. Yilmaz, J. Gerken, T. Schweer, P. Yarza, J. Peplies and F. O. Glöckner (2012). "The SILVA ribosomal RNA gene database project: improved data processing and web-based tools." Nucleic acids research 41(D1): D590-D596.

Quigley, L., O. O'Sullivan, C. Stanton, T. P. Beresford, R. P. Ross, G. F. Fitzgerald and P. D. Cotter (2013). "The complex microbiota of raw milk." FEMS microbiology reviews 37(5): 664-698.

Sessegolo, C., C. Cruaud, C. Da Silva, A. Cologne, M. Dubarry, T. Derrien, V. Lacroix and J.-M. Aury (2019). "Transcriptome profiling of mouse samples using nanopore sequencing of cDNA and RNA molecules." Scientific reports 9(1): 1-12.

Shen, J., A. G. McFarland, V. B. Young, M. K. Hayden and E. M. Hartmann (2021). "Toward accurate and robust environmental surveillance using metagenomics." Frontiers in Genetics 12: 151.

Team, R. C. (2013). "R: A language and environment for statistical computing."

Wang, Y., Y. Yan, K. N. Thompson, S. Bae, E. K. Accorsi, Y. Zhang, J. Shen, H. Vlamakis, E. M. Hartmann and C. Huttenhower (2021). "Whole microbial community viability is not quantitatively reflected by propidium monoazide sequencing approach." Microbiome 9(1): 1-13.

Wickham, H. (2011). "ggplot2." Wiley interdisciplinary reviews: computational statistics 3(2): 180-185.

Wickham, H., M. Averick, J. Bryan, W. Chang, L. D. A. McGowan, R. François, G. Grolemund, A. Hayes, L. Henry and J. Hester (2019). "Welcome to the Tidyverse." Journal of open source software 4(43): 1686.

Wood, D. E., J. Lu and B. Langmead (2019). "Improved metagenomic analysis with Kraken 2." Genome biology 20(1): 1-13.

Yang, M., A. Cousineau, X. Liu, Y. Luo, D. Sun, S. Li, T. Gu, L. Sun, H. Dillow and J. Lepine (2020). "Direct metatranscriptome RNA-seq and multiplex RT-PCR amplicon sequencing on Nanopore MinION–promising strategies for multiplex identification of viable pathogens in food." Frontiers in microbiology 11: 514.

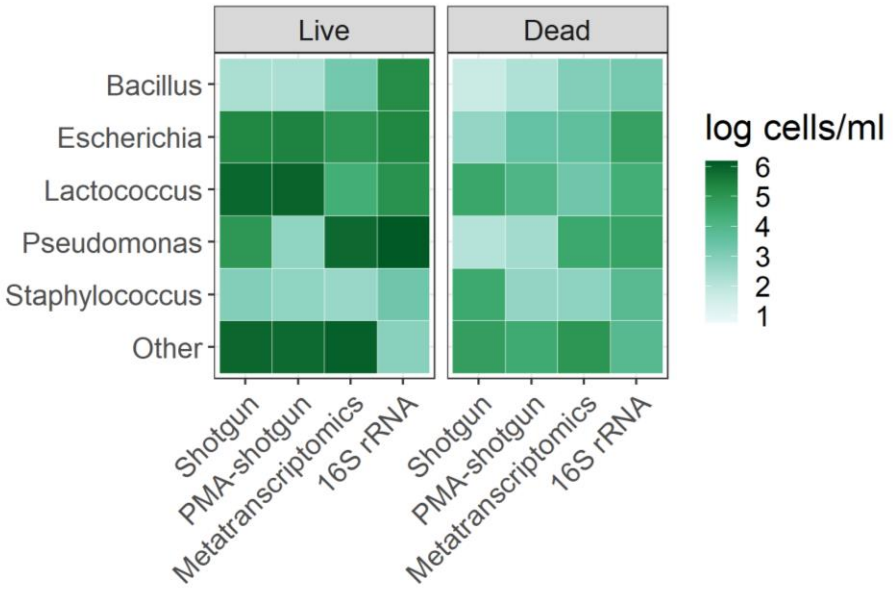
Yap, M., C. Feehily, C. J. Walsh, M. Fenelon, E. F. Murphy, F. M. McAuliffe, D. van Sinderen, P. W. O'Toole, O. O'Sullivan and P. D. Cotter (2020). "Evaluation of methods for the reduction of contaminating host reads when performing shotgun metagenomic sequencing of the milk microbiome." Scientific reports 10(1): 1-11.

Zhou, Y., M. H. Leung, X. Tong, Y. Lai, J. C. Tong, I. A. Ridley and P. K. Lee (2020). "Profiling airborne microbiota in mechanically ventilated buildings across seasons in Hong Kong reveals higher metabolic activity in low-abundance bacteria." Environmental Science & Technology 55(1): 249-259.

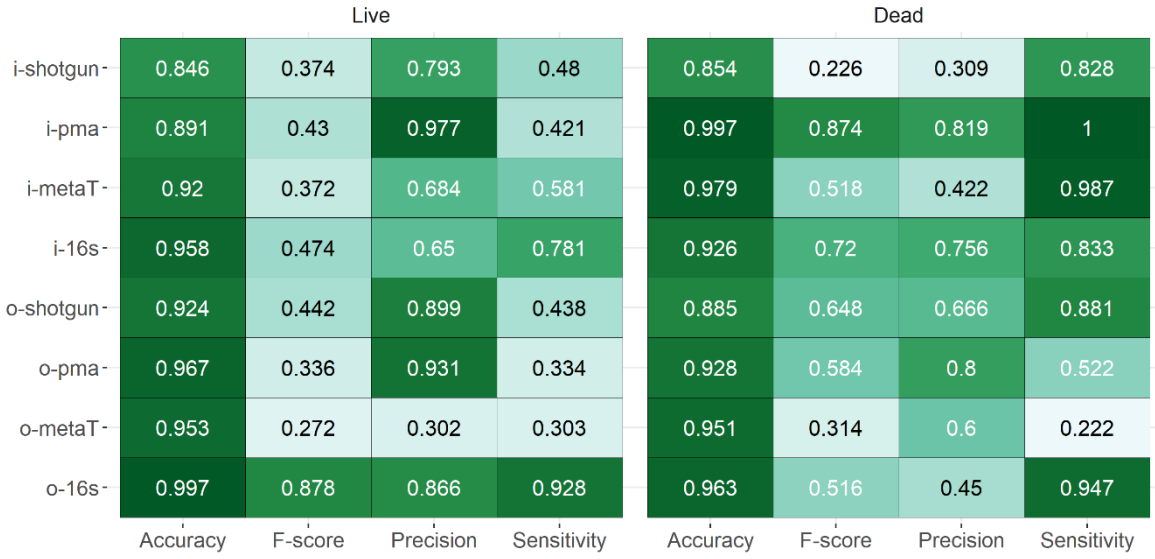
5.8 Supplemental Material

Supplemental Table 5.1. Sequencing quality metrics for each method.

Sequencing quality metrics	Shotgun		PMA-shotgun		Metatranscriptomics		16S rRNA	
	Illumina	Nanopore	Illumina	Nanopore	Illumina	Nanopore	Illumina	Nanopore
Total reads (post QC)	51,942,599	195,025	54,999,513	240,167	315,713,655	12,817	861,528	5,109,250
Total bases	8,399,308,260	414,568,905	8,817,312,962	529,322,712	20,184,088,333	6,041,520	1,003,649,400	6,555,619,361
Average reads per sample	5,771,400	16,252	6,111,057	20,014	35,079,295	1068	95,725	425,771
Median reads per sample	5,207,867 ± 1,973,157	16,679 ± 5,370	5,626,993 ± 5,458,311	25,266 ± 14,702	50,532,158 ± 26,316,834	667 ± 1246	105,232 ± 29,638	316,600 ± 521,012
Median read length	150	1107	150	910	75	378	250	1111



Supplemental Figure 5.1. Abundances of samples spiked into Live and Dead conditions, quantified using qPCR and relative abundance data, stratified into the 4 different methods studied.



Supplemental Figure 5.2. Performance metrics of samples based on comparison of abundances with the live and dead model community controls.

Supplemental Table 5.2. Other taxa found in milk samples at relative abundances above 0.1%.

	Shotgun	**PMA-Shotgun**	**Metatranscriptomics**	**16S rRNA**
Illumina	*Streptococcus*	*Streptococcus*	*Alloscardovia*	*g__Thermus*
	Anoxybacillus	*Kocuria*	*Thermus*	*g__Chroococcidiopsis*
	Thermus	*Romboutsia*	*Acinetobacter*	*_SAG_2023*
	Rothia	*Rothia*	*Anoxybacillus*	*g__Anoxybacillus*
	Kocuria	*Enterococcus_G*	*Cutibacterium*	*g__Streptococcus*
	Microbacterium	*Corynebacterium*	*Streptococcus*	*g__Enhydrobacter*
	Corynebacterium		*Bradyrhizobium*	*g__Cloacibacterium*
	Bifidobacterium		*Cloacibacterium*	*g__Leptolyngbya*
	Leuconostoc		*UBA9464*	*_Es-Yyy1000*
	CAG-791		*Bacillus_A*	*g__Acinetobacter*
	Carnobacterium		*Stenotrophomonas*	*g__Leptolyngbya*
	Mogibacterium		*Ralstonia*	*_PCC-6306*
	Enterococcus_G		*Kocuria*	*g__Anaerobacillus*
	Anaplasma		*Delftia*	*g__Kocuria*
	RUG420		*Lactobacillus*	*g__Lactobacillus*
	Lactobacillus_C		*Sphingomonas*	*g__Burkholderia-*
	Romboutsia		*Paenibacillus*	*Caballeronia-*
	Murimonas		*Vibrio*	*Paraburkholderia*
	Parageobacillus		*Methylobacterium*	
			Blastomonas	
			Psychrobacter	
			Brevundimonas	
			Sphingobacterium	
			Streptomyces	
			Enterobacter	
			Corynebacterium	
			Mycobacterium	
			Acidovorax_D	
			Moraxella_A	
			Ochrobactrum	

			Herbaspirillum *Pseudoalteromonas* *Chryseobacterium_B* *Sphingobium* *Achromobacter* *Chlamydia* *Dietzia* *Paeniglutamicibacter* *Microbacterium* *Agrobacterium* *Diaphorobacter* *Kluyvera* *Enterococcus_B*	
Oxford Nanopore Technologies	*Streptococcus* *Anoxybacillus* *Thermus* *Clostridium_F* *Parageobacillus* *Acinetobacter* *Hafnia* *Leuconostoc* *Stenotrophomonas* *Rahnella* *Turicibacter* *Yersinia*	*Streptococcus* *Photobacterium*	*Finegoldia*	*d__Bacteria* *p__Proteobacteria* *f__Enterobacteriaceae* *c__Alphaproteobacteria* *g__Acinetobacter* *c__Gammaproteobacteri* *a* *o__Rickettsiales* *g__Enhydrobacter*

Supplemental Table 5.3. Taxa found in control milk samples at relative abundances above 0.1%.

	Shotgun	**Metatranscriptomics**	**16S rRNA**
Illumina	*Bacillus*	*Escherichia*	*g__Pseudomonas*
	Enterococcus_G	*Lactococcus*	*g__Escherichia-Shigella*
	Escherichia	*Pseudomonas_E*	*g__Lactococcus*
	Lactococcus	*Staphylococcus*	*g__Bacillus*
	Pseudomonas_E	*Streptococcus*	*g__Staphylococcus*
	Staphylococcus	*Acidovorax_D*	*g__Thermus*
	Streptococcus	*Acinetobacter*	*g__Anoxybacillus*
		Agrobacterium	*g__Enhydrobacter*
		Alloscardovia	*g__Cloacibacterium*
		Anoxybacillus	*g__Streptococcus*
		Blastomonas	*g__Acinetobacter*
		Bradyrhizobium	*g__Lactobacillus*
		Brevundimonas	*g__Leptolyngbya_Es-*
		Chryseobacterium_B	*Yyy1000*
		Cloacibacterium	*g__Kocuria*
		Corynebacterium	*g__Burkholderia-*
		Cutibacterium	*Caballeronia-*
		Delftia	*Paraburkholderia*
		Diaphorobacter	
		Kluyvera	
		Lactobacillus	
		Microbacterium	
		Moraxella_A	
		Ochrobactrum	
		Ralstonia	
		Sphingobacterium	
		Sphingobium	
		Sphingomonas	
		Stenotrophomonas	
		Streptomyces	

		Thermus	
Oxford	*Acinetobacter*	*Escherichia*	*g__Pseudomonas*
Nanopore	*Anoxybacillus*	*Pseudomonas_E*	*d__Bacteria*
Technologies	*Clostridium_F*		*p__Proteobacteria*
	Escherichia		*f__Enterobacteriaceae*
	Lactococcus		*g__Escherichia-Shigella*
	Pseudomonas_E		*c__Alphaproteobacteria*
	Rahnella		*c__Gammaproteobacteria*
	Staphylococcus		*g__Bacillus*
	Streptococcus		
	Thermus		
	Turicibacter		
	Yersinia		
	Hafnia		
	Leuconostoc		
	Parageobacillus		
	Stenotrophomonas		

Chapter 6. General Discussion

6.1 Discussion

The global objective of this thesis was to employ community-based high-throughput sequencing methods to investigate microbial communities along the food chain, in particular the dairy microbiome. From the findings of this thesis, we provide further evidence that the use of shotgun metagenomic sequencing is a valuable culture-independent tool for investigating microbial communities in various environments, and of milk in particular. Culture-based methods target only specific subsets of microbes and often require further confirmatory tests to determine resolution. Whole genome sequencing, on the other hand, has been very useful in the monitoring of strain-level dynamics but is difficult to apply to samples containing multiple microorganisms at the same time (Shen, McFarland et al. 2021). Metagenomics expands the scope to the microbial community level, enabling greater levels of detail on the taxonomic composition, diversity, structure, functional potential and antimicrobial genes present.

While the value of shotgun metagenomic sequencing is readily apparent, it comes with translational challenges. The untargeted nature of shotgun sequencing allows for an unbiased analysis of the community present in samples, which means that DNA from the host and both live and dead microorganisms will be sequenced. In chapter 3, three methods of host depletion/microbial DNA enrichment were evaluated and the MolYsis Complete5 kit was found to yield significantly higher microbial reads than the other approaches. This significantly improved microbial sequencing depth and allowed for further characterisation of both bovine and human milk microbiomes through the recovery of metagenome assembled

genomes (MAGs). Since our study, other studies have highlighted the importance of microbial DNA enrichment prior to metagenomic sequencing, though mainly focusing on human samples (Shi, Wang et al. 2022). Even though the MolYsis kit was sufficient for our current application, new strategies to overcome this challenge are being developed that could potentially further improve host depletion for metagenomic sequencing (Kovaka, Fan et al. 2021). Regarding the issue of microbial viability, the pilot study undertaken in chapter 5 investigated several methods (two DNA-based and two RNA-based) to assess their effectiveness with respect to differentiating between live and dead microorganisms, with two different sequencing platforms (Illumina and Oxford Nanopore Technologies). Variations in outcomes across library types were attributed to different molecular targets, and differences between sequencing technologies were possibly due to a combination of different sequencing depths, error rates and bioinformatics pipelines. Other studies have also concluded that the use of chemical methods, such as propidium monoazide (PMA) used in chapter 5, require further optimisation as the effectiveness of such assays are influenced by the source and concentration of dye, microbial community composition, dye incubation time and temperature, sample turbidity and more (Heise, Nega et al. 2016, Wang, Yan et al. 2021). Although there is a lack of a standardised protocol and a need for prior testing before implementation across sample types and organisms, PMA tagging has already been used in a variety of environments successfully (Emerson, Adams et al. 2017) and protocols involving PMA are relatively easy to perform once optimised, which shows promise for its use in future sequencing studies to detect viable microbial communities. In contrast, as only metabolically active microbes should be

transcriptionally active, RNA-based methods can be used to characterise the viable component of microbial communities. However, RNA-based protocols are generally more laborious and complicated, not least because RNA is less stable than DNA, which could result in RNA losses (Li, Tun et al. 2017). It should also be noted that studies have also questioned the merits of using 16S rRNA as a viability marker as rRNA molecules have been found to persist in dead bacterial cells (Weigel, Jones et al. 2013). Further research is needed with more complex model communities and matrices to better determine the most appropriate sequencing-based approach to characterise the viable microbial community in a particular environment.

Concerning the different sequencing technologies and more generally short- and long-read sequencing methods, platform-specific biases have been reported (similar to our findings in chapter 5) with discrepancies in differentially abundant taxa, although similar profiles of dominant taxa, diversity and richness have also been found (Rubiola, Macori et al. 2022, Stevens, Creed et al. 2022). Each sequencing technology and platform has both advantages and disadvantages, and the choice of method and platform being dependent on the aim of the research (Tedersoo, Albertsen et al. 2021). A comparison of long read technologies found that Pacific Biosciences Sequel was more suited for amplicon sequencing of complex samples, while ONT's MinION was ideal for rapid and accurate identification of dominant microbes through both metagenetic and metagenomic sequencing (Loit, Adamson et al. 2019). Although ONT sequencing enables rapid lower cost sequencing compared to Illumina, its high error rates are currently still the limiting factor to its widespread application, despite evidence of benefits with

respect to microbial genome assembly, when combined with Illumina short reads or alone relative to Illumina based sequencing (Sevim, Lee et al. 2019, Commichaux, Javkar et al. 2021). With improvements in sequencing technology or chemistry and bioinformatics tools, these long-read sequencing technologies will hopefully soon be comparable to the current standard set by Illumina in the application of meta-omic approaches.

In this thesis, shotgun metagenomics was applied to the bulk tank milk microbiota where it successfully determined that the use of chlorine-based, relative to chlorine-free cleaning methods, had little influence on the bulk tank microbiota, while season and location had a significant influence on the taxonomic variation observed (chapter 2). This information is valuable for the dairy industry, giving farmers and milk processors greater confidence to change cleaning methods to remove the use of chlorine-based detergents from farms due to their potential detrimental effect on safety and quality of milk products. Thus, this study represents an excellent example of the successful use of metagenomics to examine microbial community composition at specific points in the production chain to ensure food safety and quality standards are met. Further study of the influencing factors in chapter 4 revealed correlations between the bulk tank microbiota, raw milk chemical composition and climate variables. This information could potentially be a starting point for an omics-based model to predict raw milk quality and safety. Current microbial risk assessment models predict the growth rate and behaviour of a microorganism in response to specific environmental conditions (Koutsoumanis, Tsaloumi et al. 2021). While these models are successful for formulating food safety

regulations and guidelines, inferences can only be drawn as they fail to take into account the community aspect of microorganisms in food. This is a gap that meta-omic sequencing data can fill, with the potential to enhance current models and more accurately mimic real world food safety and quality issues. Information on the interactions that occur between key components of the system through the correlation of microbiota data to intrinsic product data (physicochemical composition, pH, water activity) and extrinsic factors (climate, storage conditions, packaging) would be valuable to the industry and regulatory bodies. From the more prevalent use of metagenetics, food bacterial communities from metagenetic sequencing have been compiled into databases like FoodMicrobionet that could be used for modelling the presence/absence of specific bacteria in the context of other microbes in the community with other parameters (Parente, Zotta et al. 2022). However, the use of shotgun metagenomics or metatranscriptomics would provide more useful information on the behaviour of the microbial communities (Cocolin, Mataragas et al. 2018). The virulence factors expressed and metabolic or spoilage activities of the detected microbial communities and their relationships with the intrinsic and extrinsic factors could allow for modelling of the behaviour of pathogens or spoilage bacteria. In this thesis, findings from chapter 4 investigated intrinsic (raw milk microbiota and composition) and extrinsic (climate) factors which have identified correlations that could be indicative of potential markers for spoilage or safety in bulk tank milk. If applied together with optimised methods from chapters 3 and 5, the viable microbiome, with a greater microbial sequencing depth from host depletion, could be used in the model. This would offer more

information relating to likely changes in the viable microbiota in response to other parameters and more accurately identify safety and quality issues.

6.2 Potential applications

This thesis provides useful information relating to the factors that shape the bulk tank milk microbiota and demonstrates the application of enhanced community-based sequencing approaches for characterising dairy microbial communities. There are many possible applications for the results of this thesis.

First, specifically for the dairy industry, understanding the factors that influence the bulk tank milk microbiota is valuable with respect to maintaining safety and quality in raw milk. Using these culture-independent approaches to understand the impact of implemented changes to the dairy product, like chlorine or chlorine-free cleaning in chapter 2, gives a valuable insight into the influence of these changes on the overall microbial community, and not just single species or groups of microbes as is usually the case when culture-dependent methods are used. Furthermore, if shotgun sequencing or other meta-omic approaches are used, additional information on the changes in gene expression or metabolites can be inferred, which is useful for quality control purposes. In addition, an increased knowledge of the interactions between the milk microbiota, climatic factors, and milk's chemical composition, such as that in chapter 5, can allow predictions relating to the presence of spoilage or pathogenic microorganisms and an understanding of the conditions that cause a deterioration in milk quality. The use of data from community-based sequencing approaches in a predictive model or risk

management system can provide a pre-emptive view to the changes in the microbial consortia with respect to the changes in other intrinsic or extrinsic factors (as mentioned in the Discussion above).

Furthermore, the methods in chapters 3 and 5 are not limited to the dairy microbiome and can be applied to the characterisation of other microbiomes. Reductions in the levels of contaminating host cells when performing untargeted sequencing can be applied to samples with a low biomass or samples with high host DNA present (soil and human-associated samples among others). The increased microbial sequencing depth arising as a result of using the MolYsis Complete5 kit allowed for a deeper characterisation of the milk microbiome through a greater number of observed species detected and high-quality MAGs recovered (Yap, Feehily et al. 2020). An evaluation would be needed before adopting this kit for other sample types as this was only applied to milk sample matrices, but a similar concept of depleting host reads to increase the microbial sequencing depth would be beneficial for any sample type when untargeted shotgun sequencing methods are employed. In addition, the differentiation of viable and non-viable members of a microbial community would add value to the use of sequencing-based methods in various settings. In the food industry, this has food safety and quality-related implications as it allows the detection of viable and active pathogenic or spoilage bacteria along the food processing chain (Zwirzitz, Wetzels et al. 2020). Besides the food industry, distinguishing live and dead cells is useful in public health, manufacturing, or environmental applications. During outbreak investigations of emerging pathogens, environmental sampling or surveillance, being able to focus

on the viable microbes alone would help determine causative agents or sources of contamination with greater confidence and/or establish if compliance cut-offs are met (Emerson, Adams et al. 2017). From a public health perspective, the COVID pandemic has shown the value of sequencing and data sharing, and continued adoption of metagenomics and other meta-omic approaches to enhance public health surveillance is needed as a One Health approach (Ko, Chng et al. 2022).

6.3 Future work

Based on the findings and discussion in this thesis, there are several areas that warrant further study, as outlined below:

1. Further work on sequencing-based methods to distinguish viable and non-viable microorganisms in microbiomes

From the pilot study in chapter 5, differences in data produced by different library types (DNA and RNA) and sequencing technologies (Illumina and ONT) were detected. For future studies, several changes in the study design might be useful. First, the use of a single kit for combined DNA and RNA extraction would help standardise and reduce potential kit biases. A more complex model community with varying levels of viability could be used to determine if the methods tested could quantify the levels of live and dead cells. Separate optimisation studies for both Illumina and ONT sequencing technologies might be better in terms of methodology and bioinformatic analysis, so that suitable tools can be used for each type of sequencing data. Ultimately, once a particular method successfully and

accurately differentiates between viable and non-viable microorganisms, further studies with actual samples (e.g., raw milk, dairy powders, cheese) will be required to further validate the method.

2. Additional and continued data collection to develop a predictive model for food safety and quality issues in bulk tank milk

The findings from this thesis evidently showed the influence of season and location on the bulk tank milk microbiota. Other studies have shown other factors that influence bulk tank milk such as environment, lactation stage, feed, storage conditions and farm system, among others (Kable, Srisengfa et al. 2016, Doyle, Gleeson et al. 2017, Porcellato, Aspholm et al. 2018). It would be beneficial to collect data at the farm level (animal health, feed, farm system and more) to add to the analysis to see if these factors influence the bulk tank milk microbiota from a farm-level perspective, which would allow extrapolations to be made to other countries/continents where farm practices may differ significantly from those studied in Ireland.

As a continuation of chapter 4, the collection of more data over a longer period of time would provide more certainty in the interactions between climate, raw milk chemical composition and microbiota. Positive and negative associations and trends can be identified that may be useful to the dairy industry and with all the different types of data, a predictive model could be developed that might be able to pre-empt potential spoilage or quality issues based on climate and chemical composition.

3. Further investigation of the core microbiota and antimicrobial resistance genes

The use of metagenomic sequencing in this thesis has allowed for the identification of a core microbiota and antimicrobial resistance genes (ARGs) that were reported in chapters 2 and 4. However, greater taxonomic resolution of the identified core microbiota and uncovering patterns of these core taxa and detected ARGs from various metagenomic studies on Irish milk samples, not just those in this thesis, would be of interest to the dairy industry. Changes in the core microbiota of bulk tank milk may provide indications of a shift in quality which may in turn affect downstream applications and the presence of ARGs can have potential public health implications that could be monitored.

4. Further taxonomic resolution for selected taxa - MAGs

This thesis focused on the application of community-based approaches to characterise milk microbiomes. While generally genus or species level taxonomy was used in analysis, it is possible to obtain further taxonomic resolution through additional bioinformatics approaches. Reconstruction of metagenome assembled genomes (MAGs) from metagenomes of food or environment samples can enable strain-level resolution. In a previous instance, assemblies from a shotgun metagenomics approach allowed the accurate identification of a pathogenic *Salmonella* strain that caused a foodborne outbreak. In that case, environmental and food strains that were sequenced were included in a SNP-level phylogenetic

tree containing human isolates to determine if they were the causative agents (Buytaers, Saltykova et al. 2021). Strain-level resolution could be useful in situations when there is a particular species/strain of interest, which might be the case when these methods are used to track specific strains across the food chain or for technologically relevant or epidemiological purposes.

6.4 Conclusion

This thesis aimed to utilise community-based sequencing approaches to characterise the dairy microbiome, while at the same time, enhancing these current sequencing approaches for improved characterisation of microbiomes. Shotgun metagenomic sequencing was employed to understand microbial communities present in raw milk and their functional profiles across locations and seasons in Ireland. This allowed for a better resolution of the impact of cleaning method, as well as extrinsic factors that influence the dairy microbiome such as season and location. It highlighted the interconnected nature of environmental/climatic factors, raw milk composition and its microbiota, which could be exploited for monitoring or predicting food safety and quality issues.

Two challenges associated with the use of sequencing-based techniques were addressed. First, high host reads resulting from the untargeted nature of shotgun metagenomic sequencing was overcome through the evaluation of methods revealing that the use of the MolYsis kit depleted greater proportions of host DNA in milk, enabling greater microbial sequencing depth. Furthermore, several methods were studied on their ability to distinguish between viable and non-viable

cells, with the findings providing a good starting point for further study and improvement in this area for potential application in the characterisation of various microbiomes.

Overall, this thesis has presented the value of community-based sequencing approaches in studying the dairy microbiome. Improvements can still be made to current methods that will allow for better understanding of food microbiomes, which would be beneficial for food safety, quality, and public health.

6.5 References

Buytaers, F. E., A. Saltykova, W. Mattheus, B. Verhaegen, N. H. Roosens, K. Vanneste, V. Laisnez, N. Hammami, B. Pochet and V. Cantaert (2021). "Application of a strain-level shotgun metagenomics approach on food samples: resolution of the source of a Salmonella food-borne outbreak." Microbial genomics 7(4).

Cocolin, L., M. Mataragas, F. Bourdichon, A. Doulgeraki, M.-F. Pilet, B. Jagadeesan, K. Rantsiou and T. Phister (2018). "Next generation microbiological risk assessment meta-omics: the next need for integration." International Journal of Food Microbiology 287: 10-17.

Commichaux, S., K. Javkar, P. Ramachandran, N. Nagarajan, D. Bertrand, Y. Chen, E. Reed, N. Gonzalez-Escalona, E. Strain and H. Rand (2021). "Evaluating the accuracy of Listeria monocytogenes assemblies from quasimetagenomic samples using long and short reads." BMC genomics 22(1): 1-18.

Doyle, C. J., D. Gleeson, P. W. O'Toole and P. D. Cotter (2017). "Impacts of seasonal housing and teat preparation on raw milk microbiota: a high-throughput sequencing study." Applied and environmental microbiology 83(2): e02694-02616.

Emerson, J. B., R. I. Adams, C. M. B. Román, B. Brooks, D. A. Coil, K. Dahlhausen, H. H. Ganz, E. M. Hartmann, T. Hsu and N. B. Justice (2017). "Schrödinger's microbes: tools for distinguishing the living from the dead in microbial ecosystems." Microbiome 5(1): 1-23.

Heise, J., M. Nega, M. Alawi and D. Wagner (2016). "Propidium monoazide treatment to distinguish between live and dead methanogens in pure cultures and environmental samples." Journal of microbiological methods 121: 11-23.

Kable, M. E., Y. Srisengfa, M. Laird, J. Zaragoza, J. McLeod, J. Heidenreich and M. L. Marco (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(4): e00836-00816.

Ko, K. K., K. R. Chng and N. Nagarajan (2022). "Metagenomics-enabled microbial surveillance." Nature Microbiology 7(4): 486-496.

Koutsoumanis, K., S. Tsaloumi, Z. Aspridou, C. Tassou and M. Gougouli (2021). "Application of Quantitative Microbiological Risk Assessment (QMRA) to food spoilage: Principles and methodology." Trends in Food Science & Technology 114: 189-197.

Kovaka, S., Y. Fan, B. Ni, W. Timp and M. C. Schatz (2021). "Targeted nanopore sequencing by real-time mapping of raw electrical signal with UNCALLED." Nature biotechnology 39(4): 431-441.

Li, R., H. M. Tun, M. Jahan, Z. Zhang, A. Kumar, W. Dilantha Fernando, A. Farenhorst and E. Khafipour (2017). "Comparison of DNA-, PMA-, and RNA-based 16S rRNA Illumina sequencing for detection of live bacteria in water." Scientific reports 7(1): 1-11.

Loit, K., K. Adamson, M. Bahram, R. Puusepp, S. Anslan, R. Kiiker, R. Drenkhan and L. Tedersoo (2019). "Relative performance of MinION (Oxford Nanopore Technologies) versus Sequel (Pacific Biosciences) third-generation sequencing

instruments in identification of agricultural and forest fungal pathogens." Applied and Environmental Microbiology 85(21): e01368-01319.

Parente, E., T. Zotta and A. Ricciardi (2022). "FoodMicrobionet v4: a large, integrated, open and transparent database for food bacterial communities." International Journal of Food Microbiology 372: 109696.

Porcellato, D., M. Aspholm, S. B. Skeie, M. Monshaugen, J. Brendehaug and H. Mellegård (2018). "Microbial diversity of consumption milk during processing and storage." International journal of food microbiology 266: 21-30.

Rubiola, S., G. Macori, T. Civera, S. Fanning, M. Mitchell and F. Chiesa (2022). "Comparison Between Full-Length 16S rRNA Metabarcoding and Whole Metagenome Sequencing Suggests the Use of Either Is Suitable for Large-Scale Microbiome Studies." Foodborne Pathogens and Disease 19(7): 495-504.

Sevim, V., J. Lee, R. Egan, A. Clum, H. Hundley, J. Lee, R. C. Everroad, A. M. Detweiler, B. M. Bebout and J. Pett-Ridge (2019). "Shotgun metagenome data of a defined mock community using Oxford Nanopore, PacBio and Illumina technologies." Scientific data 6(1): 1-9.

Shen, J., A. G. McFarland, V. B. Young, M. K. Hayden and E. M. Hartmann (2021). "Toward accurate and robust environmental surveillance using metagenomics." Frontiers in Genetics 12: 600111.

Shi, Y., G. Wang, H. C.-H. Lau and J. Yu (2022). "Metagenomic Sequencing for Microbial DNA in Human Samples: Emerging Technological Advances." International Journal of Molecular Sciences 23(4): 2181.

Stevens, B., T. Creed, C. Reardon and D. Manter (2022). "Comparison of Oxford Nanopore Technologies and Illumina MiSeq sequencing with mock communities and agricultural soil."

Tedersoo, L., M. Albertsen, S. Anslan and B. Callahan (2021). "Perspectives and benefits of high-throughput long-read sequencing in microbial ecology." Applied and environmental microbiology 87(17): e00626-00621.

Wang, Y., Y. Yan, K. N. Thompson, S. Bae, E. K. Accorsi, Y. Zhang, J. Shen, H. Vlamakis, E. M. Hartmann and C. Huttenhower (2021). "Whole microbial community viability is not quantitatively reflected by propidium monoazide sequencing approach." Microbiome 9(1): 1-13.

Weigel, K. M., K. L. Jones, J. S. Do, J. Melton Witt, J.-H. Chung, C. Valcke and G. A. Cangelosi (2013). "Molecular viability testing of bacterial pathogens from a complex human sample matrix." PLoS One 8(1): e54886.

Yap, M., C. Feehily, C. J. Walsh, M. Fenelon, E. F. Murphy, F. M. McAuliffe, D. van Sinderen, P. W. O'Toole, O. O'Sullivan and P. D. Cotter (2020). "Evaluation of methods for the reduction of contaminating host reads when performing shotgun metagenomic sequencing of the milk microbiome." Scientific reports 10(1): 1-11.

Zwirzitz, B., S. U. Wetzels, E. D. Dixon, B. Stessl, A. Zaiser, I. Rabanser, S. Thalguter, B. Pinior, F.-F. Roch and C. Strachan (2020). "The sources and transmission routes of microbial populations throughout a meat processing facility." NPJ biofilms and microbiomes 6(1): 1-12.

Acknowledgements

Never in my wildest dreams would I have ever thought I would move halfway across the world to Ireland to do a PhD, but God opened this door, and it has been 4 challenging, yet wonderful and fulfilling years. All praise and glory to Him, He has carried me every step of the way, through the late nights/early mornings, uncertainties and so much more. And without Him this journey would not have been possible.

First, I would like to thank my primary supervisor, Prof. Paul Cotter, for taking a chance on me, for all the guidance and advice over the past 4 years. His patience, optimism, clarity and love for science is so inspiring. I feel so fortunate and grateful to have had him as my primary supervisor. I would like to thank Dr Orla O'Sullivan for the encouragement and support. And sincere gratitude to my university supervisor, Prof. Paul O'Toole, for advice and support through the years. His honest and constructive feedback and comments have helped me develop as a scientist. I would like to thank Dr Ranil Coorey, my honours supervisor at Curtin University. He encouraged me over the years to keep pursuing research and science and to not give up as I looked for a PhD. Thank you for the advice and support over the years and for the part you played in helping me get this opportunity.

I would like to thank the Irish Dairy Levy for the funding for this project. I would like to thank Dr David Gleeson and Dr Lizandra Paludetti, for their help with raw milk collection for chapter 2, and Dr Diarmuid Sheehan, Sarah Cooney and Arlene McGrath, for their help with raw milk collection for chapter 4. Dr Fiona Crispie, Laura Finnegan and Amanda Bréchon, for all the help with sequencing, without which this project could not have been completed.

The biggest thanks to the people in Vision 1, the best lab family one could ask for. To those from V1 past and present, I'm so grateful to have had so many memories, learnt so much and grown so much during my time with you all. Special thanks to Elaine Lawton, who was always there to listen and give good lab advice. Dr Conor

Feehily, for willingly guiding me through lab work, writing and publication of chapter 3. Dr Calum Walsh, for always answering my SOS emails when it came to bioinformatics. Dr Aoife McHugh, for allowing me to shadow you in the lab when I first came and for always being so helpful and offering great advice for the lab and for this PhD. Those in the office and lab, Chloe, Ciara, John, Wiley, Samuel, Raul, Ramya, Enriqueta and more, for the chats and troubleshooting discussions whenever any of us got stuck. Cathy, for being my neighbour in the office, for the catch ups over coffee and so willingly covering me in the lab when I was away. Very special thanks to Laura, Katie and Amy - I am so grateful for your friendship over the 4 years, for the smoke breaks, meals, check ins, laughs and walks. I am so glad to have gone through this journey with you and I definitely would not have been able to get here without the support from the 3 of you.

Many thanks to my “Irish” family - all those at Fermoy Presbyterian Church, it has been a joy to do life with you over the past few years. Jelvon, for being an honest friend and partner for worship, for always pointing me back to what is important. Iris, for Sunday mornings and random weekend hikes. Natan and Larissa, for roadtrip memories and constant encouragement. Jialing, for pushing me to pick up the guitar and helping me settle in to church. William and Sharon, for all the love and care I’ve received over the past years, for opening your home and welcoming me in. Granny Margaret, for your generosity and love every time I visit you up north. Yihong, Minee, and those from Jie’s lab, thank you for the meals, chats and outings. Special thanks to the ones I’ve come to share a home with over the past few years. For Tais, who reached out and welcomed me into the house. Maria, for all the Italian food and good laughs. Fabiana, for your bravery to try new things and optimism and support. And especially Jie, for putting up with me for 4 years, including several lockdowns, for driving me around, and supporting and loving me the way I am.

My gratitude to my friends near and far – Michelle, Weisian, Jasmine & family, Rosey & family, Ethel, Michelle, Nicholas, Emilia & family, Jess, Grace, and Kalada. Thank you for checking in time and again, and for the words for encouragement and support. Corli, for the lovely handwritten letters and texts out of the blue. Shermaine, for being the only person in the family who understands what I do/talk about and for the support from afar. Mian, for the texts and videocalls to keep in touch, for still keeping me part of your life and for listening whenever I needed someone. Thank you for the advice and love.

Special thanks to my parents, who have been a constant support throughout. I am grateful for all the prayers and the love they have shown me over the years as I once again packed up and left Singapore to study. My brother, Xiong, for the practical advice about life and future and for taking time to update me every week about life and happenings back home. My sister, Yen, for the love, overwhelming support and thoughtfulness throughout my PhD, in the form of letters, packages, visits and texts. I am not sure why I said yes to us being apart once again, as if the first time wasn't bad enough, but I am so grateful for all the hours we spent talking through things, figuring things out and chatting about sports. Lem, Leah and McCabe, for being good calming distractions when things are not working as expected.

And to anyone else who has helped in any way big or small, thank you!