The microbial diversity of mesophilic starter cultures used in cheese production

Den mikrobielle diversiteten i mesofile starterkulturer anvendt i ostproduksjon

Philosophiae Doctor (PhD) Thesis

Cyril Frantzen

Norwegian University of Life Sciences
Faculty of Chemistry, Biotechnology and Food Science

Ås (2017)
# Table of contents

Acknowledgements ........................................................................................................... 1
Summary ............................................................................................................................. 2
Sammendrag ......................................................................................................................... 4
List of papers ....................................................................................................................... 6
Introduction .......................................................................................................................... 7
Milk fermentation by lactic acid bacteria ............................................................................. 7
Starter cultures ..................................................................................................................... 8
Mesophilic starter cultures ................................................................................................. 10
   *Lactococcus lactis* ........................................................................................................... 10
   *Leuconostoc spp.* .......................................................................................................... 11
Bacteriophages ................................................................................................................... 13
Identification and differentiation of strains ......................................................................... 18
Microbial diversity and culture dynamics in milk ............................................................... 22
Aim of study ......................................................................................................................... 23
Main results and discussions ............................................................................................. 24
   Use of M17 and a milk-based medium enables isolation of two distinct and diverse 
   populations of *Lactococcus lactis* strains from undefined mesophilic starter cultures .... 24
   Genomic characterization of dairy associated *Leuconostoc* species and diversity of 
   leuconostocs in undefined mixed mesophilic starter cultures ........................................ 26
   Diversity of *Lactococcus lactis* in undefined mixed dairy starter cultures revealed by 
   comparative genome analyses and targeted amplicon sequencing of *epsD* ....................... 29
Conclusion and future perspectives .................................................................................... 32
References .......................................................................................................................... 34

PAPER I-III
Acknowledgements

The work presented in this thesis was carried out during the period 2013-2017 at the Laboratory for Microbial Gene Technology and Food Microbiology, Faculty of Chemistry, Biotechnology and Food Science, Norwegian University of Life Sciences. This work was supported by TINE SA and The Research Council of Norway.

First, I am sincerely thankful to my supervisor Helge Holo, for invaluable guidance and patience during the PhD project period, for all the help with my research, and the commentary on manuscripts and this thesis. I would also like to thank my co-supervisors Hans Petter Kleppen and Hilde M. Østlie for their most helpful advices and comments on the manuscripts and the thesis, and deliver a huge thank you to all past and present colleagues at LMG, and a special thank you to Linda Godager a.k.a. “The Boss”. I am also grateful for the research collaborators and co-authors that I have been fortunate enough to meet to during the project, with a special thanks to Witold Kot, Finn Vogensen and Horst Neve for tolerating my barrage of questions at the early stages of the project, and for the invaluable advices given during the later stages of the project.

To all of my family and friends, thank you for all the encouragement throughout this period. Thanks to Lars Adrian Johnsen for the lesson in Norwegian grammar. Finally yet most importantly, eternal thanks and love to my two awesome boys Theo and Odin. You always lift my spirits, even when my scientific work weighs heavily on my shoulders. You make me a stronger and better person <3.

*If you aren't scandalized by the work you did five or even three years ago, you're not learning anywhere near enough.* (Nick Black)
Summary

For ages, humankind has preserved various foods by fermentation by lactic acid bacteria (LAB), and fermentation of milk to obtain cheese can be traced back to the domestication of cattle, at least seven millennia ago. An essential ingredient in contemporary production of Dutch-type cheeses are the undefined mixed mesophilic (DL) starter cultures, which contains unknown mixtures of *Lactococcus lactis* strains and *Leuconostoc* spp..

Bacteriophages infecting *Lactococcus lactis*, the major contributors in the acidification of milk using mesophilic starter cultures, are recognized as the major cause of fermentation failures in dairy fermentations, disrupting the acidification process and negatively affecting the quality of the final product. The undefined mixed (DL) starter cultures are considered more robust against phage attack than the defined cultures, a characteristic gained from their large number of strains with diverse phage sensitivity. Starter cultures from different manufacturers are known to give cheeses qualitatively different characteristics, and performance differences are reported for different batches of the same starter culture, which indicates dissimilar culture compositions. Information on the microbial diversity of starter cultures is not publically available and tools to quantify the strain diversity or compare compositional differences between starter cultures does not exist. The information provided by the culture manufacturer with culture purchase does not include details beyond genus for leuconostocs, or beyond subspecies for the lactococci.

In this study, the diversity of bacteria and their bacteriophages in starter cultures and dairy samples collected from three major cheese plants in Norway was investigated using molecular and DNA-sequencing based approaches. Use of a milk based-medium (GMA) in addition to the traditional M17 was instrumental in capturing a larger diversity of bacteria from starter cultures, which consequently increased the capacity to isolate bacteriophages from the dairy samples. The bacteria and bacteriophages were discriminated from each other use phage typing, revealing a large number of different bacteria as well as different bacteriophages. Interestingly, many of the strains that were only able to grow in a milk-based media, demonstrated unique phage sensitivities. A large number of phenotypically different starter bacteria with dissimilar phage sensitivities were whole-genome sequenced and characterized in pan-genome analyses. Pan-genome analyses discriminated between 21 *Lactococcus lactis* subsp. *lactis*, 28 *Lactococcus lactis* subsp. *cremoris*, as well as 12
Leuconostoc spp. lineages. Interestingly, the analyses did not discriminate *Leuconostoc mesenteroides* subsp. mesenteroides from *Leuconostoc mesenteroides* subsp. *dextranicum*, and showed that genomic variation between the isolates was much greater than between the subspecies. The diversity of *Lactococcus lactis* of three DL starter cultures was analyzed by targeted-amplicon sequencing of 16S rDNA, the core gene *purR*, and the softcore gene *epsD*, present in over 95% of starter culture isolates, but absent in most of the reference strains. The results revealed significant differences between the three starter cultures as well as compositional shifts during cultivation in milk. Compositional analyses of the *Leuconostoc* population in the five DL starters by targeted-amplicon sequencing of *eno*, the gene encoding for enolase, also revealed significant differences between the cultures. Three of the cultures were dominated by *Leuconostoc mesenteroides* subsp. *cremoris* while *Leuconostoc pseudomesenteroides* dominated in the other two. *Leuconostoc mesenteroides* subsp. *mesenteroides* and subsp. *dextranicum* was found in all DL cultures, while *Leuconostoc lactis*, reported to be a major constituent in fermented dairy products, was only identified in one of the cultures.

This work shows that starter cultures are different both with regards to both lactococci and leuconostocs, and provides tools to describe the microbial diversity of mesophilic starter cultures. The dairy industry and starter culture manufacturers can vastly improve their ability to monitor all phases of starter culture and cheese production by implementing the methods described in this work. Routine analysis of the microbial composition of starter cultures will enable quality control of starter cultures, and enable the industry to make competent decisions regarding starter culture rotations in the event of phage attack.
I årtusener har mennesker utnyttet melkesyrebakterier (LAB) til å konservere mat via fermentering. Produksjon av ost via fermentering av melk kan spores minst 7.000 år tilbake til domestiseringen av storfe. En essensiell ingrediens i moderne produksjon av gulost er starterkulturene, som oftest såkalt udefinerte mesofile blandingskulturer (DL) som inneholder et ukjent antall forskjellige *Lactococcus lactis* stammer og *Leuconostoc* spp.. Kjent som den hyppigste årsaken til fermenteringsfeil, er bakteriofager som angriper *Lactococcus lactis*, den viktigste bidragsyteren i forsuringen av melk ved bruk av mesofile starterkulturer. Bakteriofagangrep kan forstyrre forsuringsprosessen og redusere kvaliteten på sluttproduktet. Fordi de inneholder et stort antall stammer med ulik følsomhet for bakteriofager, anses de udefinerte blandingskulturene som mer robuste mot bakteriofagangrep enn definerte kulturer. Det er kjent at starterkulturer fra ulike produsenter gir ostene forskjellige kvalitetsmessige karakteristikker, en indikasjon på ulikheter i kulturkomposisjonen. Informasjon om den mikrobielle diversiteten i starterkulturene er ikke offentlig tilgjengelig og verktøy for kvantifisering av stammediversiteten eller for å sammenligne kulturkomposisjonen mellom kulturene eksisterer ikke. Kulturprodusentene oppgir ikke detaljer utover genus for *Leuconostoc*, eller utover underart for *Lactococcus lactis*.

Leuconostoc mesenteroides subart dextranicum, og viste at den genomiske variasjonen mellom isolatene var mye større enn mellom subartene. Diversiteten av Lactococcus lactis ble undersøkt i tre DL starterkulturer ved «amplicon» sekvensering av 16S rDNA, «core»-genet purR, og «softcore»-genet epsD som var tilstede i over 95% av starterkultur isolatene, men var fraværende i flesteparten av referansestammene. Resultatene avslørte betydelige forskjeller mellom de tre starterkulturene og endringer i kulturkomposisjonen under kultivering i melk. Komposisjonsanalysen av Leuconostoc i fem DL starterkulturer ved «amplicon» sekvensering av eno, genet som koder for Enolase, et essensielt enzym i glykolysen avslørte også signifikante forskjeller mellom starterkulturene. Tre av kulturene var dominerende av Leuconostoc mesenteroides subsp. cremoris mens de to resterende kulturene var dominerede av Leuconostoc pseudomesenteroides. Et lavt antall av Leuconostoc mesenteroides subsp. mesenteroides and subsp. dextranicum ble identifisert i alle DL starterkulturene, mens Leuconostoc lactis, beskrevet i litteraturen som høyst relevant, ble kun identifisert i lave antall i en av kulturene.

Dette arbeidet viser at mesofile starterkulturer er forskjellige, både med hensyn til laktokokker og leukonostokkene og inkluderer verktøy for å beskrive den mikrobielle diversiteten i mesofile starterkulturer. Ved å implementere metodene beskrevet i dette arbeidet kan meierinæringen og starterkulturprodusentene oppnå en betraktelig bedre evne til å overvåke alle faser av starterkultur og osteproduksjonen. Kvalitetskontroll av meieriproduksjonen ved å regelmessig analysere den mikrobielle komposisjonen i starterkulturene kan bidra til å redusere svinn, effektivisere produksjonen, og styrke evnen til å avgjøre hvilke kulturer som benyttes i produksjonen, samt hvilke kulturer som inkluderes i et rotasjonssystem, skulle produksjonen være utsatt for bakteriofagangrep.
List of papers

Papers included in this thesis:

Paper I:

Paper II:

Paper III:
**Frantzen C, Kleppen HP, Holo H.** 2017. Diversity of *Lactococcus lactis* in undefined mixed dairy starter cultures revealed by comparative genome analyses and targeted amplicon sequencing of *epsD*. (Submitted manuscript)

Co-authored papers published during the project period:


Introduction

Milk fermentation by lactic acid bacteria

For ages, humankind has preserved various foods by fermentation by lactic acid bacteria (LAB). The fermentation of milk to obtain cheese has been linked all the way back to the domestication of cattle, millennia before the common era [1]. After the development of pasteurization, the essential role of microorganisms in fermentation of foods has been apparent, and with the industrial revolution, a large shift to large-scale food production was necessary to accommodate the dramatic increase in population densities. LAB traditionally used in contemporary food fermentations include certain species of the genera Lactobacillus, Lactococcus, Streptococcus, Leuconostoc, Pediococcus, Enterococcus and Weissella [2, 3], which all contribute to the preservation of the food by the production of organic acids. This greatly increases product shelf life compared to the raw product, and may significantly alter the taste and texture of the product.

In the processing of milk to obtain cheese, the fat and milk proteins are concentrated, while a variable proportion of the water-soluble fraction (whey) is removed. The four main ingredients in this process are; milk, microorganisms, rennet, and salt. Microorganisms are highly relevant in two of the main steps involved in producing cheese, the acidification of milk and the ripening of cheese. These microorganisms are commonly referred to by three main terms, starter bacteria, adjunct bacteria and non-starter lactic acid bacteria (NSLAB) based on their source, use and function in the production [4-6]. The starter bacteria are essential to the dairy fermentation process and the common species used are Lactococcus lactis, Leuconostoc spp., Streptococcus salivarius subspecies thermophilus (S. thermophilus), Lactobacillus delbrueckii subspecies lactis (Lb. lactis), subspecies bulgaricus (Lb. bulgaricus) and subspecies helveticus (Lb. helveticus) [3]. Adjunct bacteria can be defined as microorganisms that are added during cheesemaking with intent in order to achieve a particular quality, taste, or characteristic in the final product [7]. The NSLAB comprises the lactic acid bacteria that are not intentionally added to the production process, yet are present in substantial amounts, sometimes even dominating the cheese microflora in late stages of cheese ripening [8]. The presence of adjunct bacteria, adjunct fungi, or NSLAB during ripening are commonplace in a variety of cheeses, such as the Swiss-type cheeses [9] and the surface mould-ripened cheeses [10].
Starter cultures

The starter bacteria used in dairy production are merchandised as starter cultures, which are categorized by the general metabolic and growth characteristics of the included bacteria. Several different starter compositions, usually including different species or several strains of the same species, are available. An initial distinction is made between dairy starter cultures containing thermophilic bacteria, with an optimum growth temperature of ~42 °C (37-50 °C), and starter cultures containing mesophilic bacteria, which grow at a lower temperature range (optimal temperature between 20-37 °C) [11]. The thermophilic starter cultures contain the lactobacilli (Lb. lactis, Lb. bulgaricus, Lb. helveticus) and S. thermophilus, and are applied in the production of yoghurt and cheeses with high fermentation temperatures, such as the Italian- and Swiss-type cheeses [12, 13]. The mesophilic starter cultures contain the lactococci (Lactococcus lactis subsp. lactis, subsp. cremoris, and subsp. lactis biovar diacetylactis), and the leuconostocs. These are used in the production of Dutch- and Scandinavian-type, continental, and Cheddar cheeses [14]. Mixing mesophilic and thermophilic cultures is uncommon in traditional cheese production. However, S. thermophilus has been used as an adjunct to mesophilic starter cultures in the manufacture of Cheddar cheese [15]. This composition of bacteria is commercially available as so-called RST cultures [16].

Beyond the distinction between mesophilic and thermophilic cultures, starter cultures are also divided into categories by their microbial content of species and strains, e.g. as described by Pogaku Ravindra [17]:

- Single-strain starters: Cultures that contain one strain of a certain species.
- Multiple-strain starters: Cultures that contain multiple known strains of a certain species.
- Mixed starters: Cultures that contain multiple known species of bacteria. Mixed starters can be further divided into two sub-categories. Cultures containing a known number of strains (defined) and cultures containing an unknown number of strains (undefined).
Typically, the contemporary starter cultures originate traditional dairy-farm cheese production, where starter bacteria from a successful production was used to inoculate the next, a process called back-slopping [14]. Through back-slopping, artisan dairy farmers facilitated the evolution of diverse microbial communities with geographically distinct compositions. Naturally, dairy farmers favored starter cultures known to result in good quality cheese, and the exchange of artisanal starter cultures between cheesemakers was common practice [14], adding a layer of hands-on selection to the natural evolution of microbial starter communities. Commercial starter cultures are manufactured from these artisanal cultures, which were frozen down sometime during the middle of the twentieth century and has been stored frozen since then [8]. Starter cultures can be inoculated directly for cheese-making as a direct vat set (DVS) culture or inoculated and propagated prior to cheese-making to produce so-called bulk starter cultures [18]. Special care is taken by culture developers to preserve the microbial composition of the frozen seed stock culture, and strict control of growth parameters is enforced to minimize compositional changes during culture propagation. Although culture propagation by back-slopping regimes have been shown to ultimately sustain the microbial community [19], the composition of the culture may change significantly over shorter time periods, depending on growth conditions and bacteriophage attack [20]. The dairy industry, which are dependent on reliable and reproducible culture performance, avoid potential day-to-day fluctuations by using frozen or freeze-dried seed stock cultures, effectively resetting the microbial composition every day of production. The works included in this thesis are focused on mesophilic starter cultures used in the manufacture of European continental cheeses. An assortment of mesophilic starter cultures are available, and the choice of starter culture is important as it affects the taste, aroma, and quality of the final product.
Mesophilic starter cultures

Essential in the manufacture of continental cheese types, the mesophilic starter cultures are primarily composed of homo-fermentative *Lactococcus lactis* subsp. *lactis* (*L. lactis*) and *Lactococcus lactis* subsp. *cremoris* (*L. cremoris*) [6, 21]. Mesophilic starter cultures may also contain additional naturally occurring bacteria (secondary starter bacteria) and are divided into four sub-categories, O, D, L and DL, indicating which (if any) secondary starter bacteria are present. The O-starter only includes strains of *L. lactis* and *L. cremoris*, while the other also contains either *L. lactis* subsp. *lactis* biovar. *diacetylactis* (*L. diacetylactis*) (D-starter), *Leuconostoc* spp. (L-starter), or both *L. diacetylactis* and *Leuconostoc* spp (DL-starter) [21]. The mesophilic starter cultures are divided into cultures of unknown strains (undefined) or cultures of known strains (defined). Defined starters are used mostly in the manufacture of cheeses like Cheddar, while in the production of European continental cheeses the use of undefined mixed DL starter cultures is more common [22]. With DL starter cultures, the *L. diacetylactis* and *Leuconostoc* spp. provide aroma and texture to the cheese product, while *L. lactis* and *L. cremoris* are the major contributors in the acidification process through fermentation of lactose [23].

*Lactococcus lactis*

*L. lactis* is the main constituent of mesophilic dairy starter cultures and has GRAS (Generally Regarded As Safe) status, based on its long history of safe use in food fermentations [24]. Although predominantly associated with the dairy environment, *L. lactis* strains have been isolated from a number of sources but are believed to originate from the plant environment [24-27]. *L. lactis* is one of the most widely studied lactic acid bacteria, and dairy strains are distinguishable from their non-dairy counterparts by both phenotype and genotype [11, 28]. *L. lactis* found in dairies have adapted to the environment and genomic analyses corroborate its hypothesized origin by revealing the telltale signs of degenerative evolution [28]. A large genetic diversity within the dairy *L. lactis* has been identified using a wide range of DNA fingerprinting and culture-dependent sequencing methods [28-31]. The *L. lactis* species includes four subspecies, *lactis*, *cremoris*, *hordniae*, and *tructae*. The former two, subsp. *lactis*, and subsp. *cremoris* are routinely employed in the making of cheese, primarily contributing in the acidification of milk through fermentation of lactose [21], but also influencing the texture and taste of the cheese product.
The distinction between subspecies *lactis* and *cremoris* was initially based on phenotypic features. *L. lactis* subspecies *lactis* has the ability to grow at 40° C, in 4 % NaCl, at up to pH 9.2, is able to deaminate arginine, and to ferment maltose, while *L. lactis* subspecies *cremoris* does not [33, 34]. Moreover, a biovariant of *L. lactis*, biovar *diacetylactis* distinguishes itself by metabolizing citrate to produce diacetyl [35-37], a flavour and aroma compound important to the characteristics of products such as the continental cheeses. Detailed studies on the genetic relation of the subspecies have shown that phenotypic features alone are inadequate to identify subspecies [38]. Moreover, there is a discrepancy between the subspecies identification determined by phenotypic features and genotypic identification determined using 16S rDNA sequences [33]. Strains of *L. lactis* identified as subspecies *cremoris* by genotype have been reported to show a subspecies *lactis* phenotype, and *vice versa*, making accurate identification and differentiation of isolates a difficult task [33, 39]. Pan-genome analysis of *L. lactis* has revealed a number of genes that are unique to each subspecies, and sequence analysis of core genes show a clear phylogenetic division between the subspecies [28]. This highlights the importance of implementing genome sequencing in taxonomic characterization to achieve precise and robust identification and differentiation of *L. lactis* strains.

**Leuconostoc spp.**

The heterofermentative leuconostocs grow associatively with the acid-producing lactococci in fermented dairy products. The importance of the *Leuconostoc* in cheese production is widely recognized [23]. In addition to providing aroma and texture to the product by metabolizing citrate, producing diacetyl, acetoin and CO₂, they have been suggested to play a role in promoting the growth of citrate positive *Lactococcus* strains [23, 40, 41]. The starter culture manufacturers do not provide information on the content of *Leuconostoc* beyond genus, and the details on the strain diversity of *Leuconostoc* species in DL-cultures is not available to the scientific community. Due to their low initial number and comparatively slow growth, *Leuconostoc* spp. are not believed to have a significant effect in the acidification process in the early stages of cheese making [42]. However, leuconostocs have been shown to dominate the cheese LAB in the late stages of ripening in productions using adjunct propionic acid bacteria [4, 43]. The genus *Leuconostoc* includes 13 species, with the species *Leuconostoc mesenteroides* divided into subspecies *mesenteroides*,
dextranicum, cremoris, and suionicum [44, 45]. The Leuconostoc species and subspecies found in dairy production are Leuconostoc mesenteroides subsp. mesenteroides (Ln. mesenteroides), Leuconostoc mesenteroides subsp. dextranicum (Ln. dextranicum), Leuconostoc mesenteroides subsp. cremoris (Ln. cremoris), Leuconostoc pseudomesenteroides (Ln. pseudomesenteroides) and Leuconostoc lactis (Ln. lactis) [46, 47]. Scientific literature and product information on starter cultures pre-dating the genomic age list Ln. cremoris and Ln. lactis as the key Leuconostoc in undefined mixed mesophilic starter cultures [23, 48, 49]. However, in recent years, isolation of Ln. mesenteroides, Ln. dextranicum, and Ln. pseudomesenteroides is more common from starter cultures or from cheese derivatives [4, 50-53]. The taxonomy of dairy relevant leuconostocs is based on the accumulative work of Ellen Garvie and John Farrow, who characterized leuconostocs decades ago using biochemical phenotype traits or non-specific molecular methods [54-60]. Since then, several molecular methods have been employed to differentiate and identify Leuconostoc isolates [61-71]. However, concerns about the stability and reproducibility of culture-dependent methods have been raised [47, 72], and comparison of results between the methods and between different laboratories is challenging. To our knowledge, the Leuconostoc genus has not been subject to extensive genomic research, and the information on the diversity of Leuconostoc species, the culture dynamics, and compositional fluctuations through the cheese production is not available.
**Bacteriophages**

Bacteriophages, or “phages”, are viruses that depend on bacterial hosts to propagate. All known LAB-infecting phages belong to the *Caudovirales* order, possessing double stranded DNA (dsDNA) genomes and isometric or prolate capsids. This order is further divided into three families, *Siphoviridae* with long non-contractile tails, *Podoviridae* with short non-contractile tails, and *Myoviridae* with long contractile tails [73, 74].

Phages may follow one of two life cycles, the lytic cycle or the lysogenic cycle. Phages with the ability for both are called temperate phages, while phages that can only reproduce by the lytic cycle are called virulent phages. In both life cycles, phage infection initiates with the interaction between the tip of the phage tail and a receptor on the bacterial surface [75]. Following attachment to the host cell surface, phage DNA injected into the host cell is either; (i) replicated and transcribed to produce progeny phages which are subsequently released through lysis of the host cell, (ii) integrated into the chromosome of the host bacterium and replicated along with the host chromosome [76, 77]. Phages integrated into the host bacterium chromosome are termed prophages. Prophages can persist in a dormant state as the host replicates, or exit its host by switching to the lytic life cycle. A switch to the lytic cycle can occur spontaneously, because of stressful conditions, or be induced *in vitro* by DNA-damaging agents such as UV-light or mitomycin C [75, 78].

In addition to the two defined life cycles, phages sometimes interact with their hosts in less defined infections termed pseudolysogeny [79]. The state of pseudolysogeny exists as a grey area between the lysogenic and lytic life cycles, in which the phage neither establishes itself in, nor kills its host [80]. Pseudolysogeny could be an important aspect of phage-host interactions, and may be involved in facilitating the survival of phages in hosts that are nutrient-limited or otherwise debilitated [81]. Indeed, delayed cell lysis of nutrient-depleted hosts harboring pseudolysogenic phages has been reported when nutrients are added to the host environment, the phage reverting to the lytic life cycle when the host is no longer in a retarded state [82].
**Bacteriophages infecting Lactococcus lactis**

Disruption of the milk fermentation process by dairy phages that infect starter bacteria was first reported in 1934 [83, 84]. Since, phages infecting *L. lactis* have been the leading cause of fermentation failures, negatively affecting the production process and quality of the final product [85, 86]. Due to their prominent negative effects on milk fermentations and the economic impact of fermentation failures, phages infecting *L. lactis* are among the best studied groups of bacteriophages [73, 87]. Lactococcal phages are classified into ten groups (Table 1) [88]. Of these, the 936, c2 and P355 phage groups predominate in dairy environments, although members of the c2 group appear to have become less prominent in the recent years [89]. Phages belonging to the P335 group can be temperate or virulent whereas members of the 936 and c2 groups are exclusively virulent. Based on genomic and morphological analysis, the P355 phages have been divided into four sub-groups [89]. Neither the 936 nor c2 phage groups include such sub-division, both groups displaying a highly conserved genomic organization [90, 91]. Traditionally, members of the lytic 936 and c2 groups have been distinguished from each other by their host range. However, the rapid advancement and availability of high-throughput sequencing technologies in the recent decade has facilitated analysis of phages on the genomic level. Since the genomic characterization of phage sk1, the first 936-group phage, more than a hundred more phages have been added to the 936 group, which currently includes 123 publically available genomes. In contrast, only ten genomes are available in the c2 group. Although c2 phages are still of great scientific interest, a substantial proportion of the scientific efforts to characterize dairy phages have been focused on the P355 and 936 phage groups.
Table 1: Overview of the ten taxonomic groups of lactococcal phages (adapted from [73]).

<table>
<thead>
<tr>
<th>Group</th>
<th>Phage family</th>
<th>Genome size (Kb)</th>
<th>ORFs(^a)</th>
<th>TEM(^b)</th>
<th>Virion dimensions (nm)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(capsid diameter, tail width, tail length)</td>
</tr>
<tr>
<td>936</td>
<td>Siphoviridae</td>
<td>26-32 [91]</td>
<td>49-63 [91]</td>
<td></td>
<td>50, 11, 126</td>
</tr>
<tr>
<td>P335</td>
<td>Siphoviridae</td>
<td>31-41 [89]</td>
<td>47-60 [89]</td>
<td></td>
<td>49, 7, 104</td>
</tr>
<tr>
<td>c2</td>
<td>Siphoviridae</td>
<td>21-23 [90]</td>
<td>37-40 [90]</td>
<td></td>
<td>54x41, 10, 95</td>
</tr>
<tr>
<td>1358</td>
<td>Siphoviridae</td>
<td>37 [92]</td>
<td>43 [92]</td>
<td></td>
<td>45, 10, 93</td>
</tr>
<tr>
<td>Q54</td>
<td>Siphoviridae</td>
<td>26.5 [93]</td>
<td>47 [93]</td>
<td></td>
<td>56x43, 11, 109</td>
</tr>
<tr>
<td>P087</td>
<td>Siphoviridae</td>
<td>60 [94]</td>
<td>88 [94]</td>
<td></td>
<td>59, 14, 163</td>
</tr>
<tr>
<td>949</td>
<td>Siphoviridae</td>
<td>114.7 [95]</td>
<td>154 [95]</td>
<td></td>
<td>70, 12, 490</td>
</tr>
<tr>
<td>1706</td>
<td>Siphoviridae</td>
<td>55.6 [96]</td>
<td>76 [96]</td>
<td></td>
<td>58, 11, 276</td>
</tr>
<tr>
<td>P034</td>
<td>Podoviridae</td>
<td>18.7 [97]</td>
<td>28 [97]</td>
<td></td>
<td>57x40, 5, 19</td>
</tr>
<tr>
<td>KSY1</td>
<td>Podoviridae</td>
<td>79.2 [98]</td>
<td>134 [98]</td>
<td></td>
<td>223x45, 6, 32</td>
</tr>
</tbody>
</table>

\(^a\) Open reading frame (ORF).

\(^b\) Transmission electron microscopy (TEM). The black bar represents a length of 50nm.
The detailed interactions between phage and bacteria that initiate or bring the infection to completion are not fully understood, but a host-range determinant has been proposed for c2 phages [90, 99], and a correlation between host specificity and the phylogeny of the receptor-binding protein (RBP) has been shown for the P335 and 936 phages [89, 91]. All three phage groups initially recognize carbohydrate structures in attachment to the host cell wall [100, 101]. However, while a secondary protein receptor is involved in the infection process for c2 phages, no such evidence for a secondary receptor exists for the P355 or 936 phages. Bacterial sensitivity to some 936 phages has been related to which cell-wall polysaccharide (CWPS) type the bacteria belongs to [102, 103]. Attempts to distinguish the 936 phages from each other using the amino acid sequence of RBP, and relating that to the CWPS-type of their host bacteria has enabled clustering of certain 936 phages into five RBP-groups [91]. However, this does not provide a classification in full agreement with the observed host-range of phages, and conclusive proof of interactions between phage RBP and bacterial CWPS has yet to be obtained.

For c2 phages, reversible attachment to a carbohydrate receptor is followed by irreversible interactions with the phage infection protein (Pip) or the recently discovered YjaE-protein, both of which are membrane spanning [90, 104]. The evidence for a causal link between bacterial sensitivity to a number of c2 phages and the presence of Pip has existed for over a decade, and is quite compelling. However, reports of c2 phages unaffected by mutations in Pip divided the c2 group into two sub-types. Those which depended on Pip (c2 type), and those which did not depend on Pip (bIL67 type), but depend on YjaE instead [104, 105]. Host recognition by c2 phages have suggested to be encoded by the genes l14, l15, l16 in phage c2 corresponding to orf34, 35 and 35 in bIL67 [90, 99]. Comparative genomic analysis shows a good correlation between the genetic variation found within these three genes, and the division of Pip-dependent and YjaE-dependent c2 phages. Moreover, European isolates are more similar to each other than they are to American isolates, and *vice versa*, indicating a history of divergent evolution [90].
**Bacteriophages infecting Leuconostoc species**

First described in 1978, phages that infect leuconostocs are found regularly in dairy products [106-108]. Phages infecting *Leuconostoc* species can negatively influence the flavor and texture of the final product [109]. However, since they do not cause fermentation failure like the lactococcal phages do, their presence may sometimes be overlooked. The *Leuconostoc* phages are divided into subgroups based on the host species they infect, a subdivision corroborated by comparative genomic analysis [110, 111]. To date, thirteen complete genome sequences of virulent *Leuconostoc* phages and one temperate phage (phiMH1) have been sequenced [51, 110, 112-115]. The lytic phages range from 25.7 to 29.5 kb in size, with a GC content of 36.0 to 36.8%, seven of these infecting *Ln. mesenteroides*, the remaining six infecting *Ln. pseudomesenteroides*. Phages infecting *Ln. lactis* have been detected [49], but have yet to be sequenced.

Dissimilar host ranges has been demonstrated both for phages infecting *Ln. pseudomesenteroides* and for phages infecting *Ln. mesenteroides*. A host-determining RBP has been described in phages infecting *Ln. pseudomesenteroides* [114], and a RBP homolog has been deduced in phages that infect *Ln. mesenteroides* [111]. Sequence analysis of the RBP sequences indicate a good correspondence between host range and RBP sequence similarities. The construction of a chimeric *Ln. pseudomesenteroides* phage with a replacement RBP from a closely related phage has been shown to alter its host range. However, no such experiment exists for phages that infect *Ln. mesenteroides*. Recently, cross-species infectivity between some dairy phages infecting *Ln. mesenteroides* and *Ln. pseudomesenteroides* was reported [111]. However, this contradicts previous results, which clearly separates between the two species and their phages [106, 116], and is disputed by a number of experts in the field of *Leuconostoc* phages (Finn Vogensen, Witold Kot, Horst Neve, personal communications, 26.09.2017, not published). A more thorough investigation on the possibility of cross-species infectivity is necessary in order to provide conclusive evidence for this claim.
Identification and differentiation of strains

Robust identification and differentiation of strains is essential for the analysis of the microbial diversity of starter cultures. As in other fields in microbiology, strain identification and differentiation in dairy products can be performed using culture-dependent or culture-independent methods (Figure 1). Culture-dependent methods consist of isolating and culturing microorganisms prior to identification according to the morphological, biochemical or genetic characteristics of the isolate. The culture-independent methods omit the growth and enrichment steps by extracting DNA or RNA directly from the sample. Due to their cultivation requirement, culture-dependent methods are often time-consuming and require elaborate equipment, reagents or culture techniques to be performed.

![Cheese microbial ecosystem](image)

Figure 1. Overview of culture-dependent and culture-independent methodology, with examples of methods within each category. Adapted from [31].

**Culture-dependent methods** are divided into the traditional and the molecular methods. The traditional methods comprises the classical biochemical tests and phenotypic characterization, while the molecular methods include characterization, profiling, or
differentiation by genotype or protein content. The molecular methods encompass the finger printing techniques and DNA sequencing techniques, and often provide more reliable, reproducible results without any extra cost or time spent compared to the traditional methods [117]. Finger printing techniques refers to the genotyping methods that distinguish between samples based on non-sequence characteristics and/or pattern(s) of DNA. Commonly, finger printing techniques include treating DNA with restriction enzymes or using the DNA as template for a PCR, in order to generate fragments that are distinguishable using gel electrophoresis. The sequence-based methods traditionally involve either:

- Sequencing the DNA of one or more loci, using the nucleotide variation within the loci to differentiate between the strains.
- Sequencing the genomic DNA of the strain (whole-genome sequencing).

In the last decades, complete or partial DNA sequencing of the 16S rDNA gene has been the most widely used method for identification of bacterial species [118], sometimes enabling identification at the subspecies level. The 16S rDNA genes contain nine “hypervariable regions” named V1-V9 [119]. In the analysis of mesophilic lactic acid bacteria by partial 16S rDNA sequencing, the most consistent variable regions have been V1-V3 [120, 121], made even more relevant by the advent of next-generation sequencing (NGS) techniques. Since the introduction of NGS techniques, analysis of bacterial communities using targeted-amplicon sequencing schemes to acquire large number of sequences has become common practice. Several NGS platforms exist, each with their own advantages and limitations. The platform most commonly used to perform community analysis using targeted-amplicon sequencing is the Illumina platform, methodologically limited to a maximum of 600 nucleotides per sequence, but practically limited to somewhere between 450-500 nucleotides per sequence, given the requirements for successful downstream analysis. This limitation prohibits the complete sequencing of 16S rDNA, and demands selection of variable region(s) to sequence. For distinguishing between very closely related species or beyond subspecies, the sequence of protein-coding genes is more discriminative than 16S rDNA [122], and the sequence variation of single-copy target genes is shown to be sufficient for improving the phylogenetic resolution within species [31]. A large number of multi locus sequencing typing (MLST) schemes have been developed to differentiate and
identify bacteria. These involve the selection of several protein-coding loci present in all members of the genus or species, and which contain nucleotide variation to enable differentiation and identification. MLST schemes are usually genus or species specific [71, 123, 124], and was considered the “gold standard” for typing bacteria preceding whole-genome sequencing (WGS). The declining cost of DNA sequencing in combination with the significant improvement and inflation of WGS-associated software for downstream analysis has made WGS increasingly available to scientists, and solidified the position of WGS as the new gold standard for culture-dependent identification and differentiation. However, albeit extremely useful, the culture-dependent approaches present with several disadvantages, especially in providing comprehensive information on the composition of microbial communities [125]. Intrinsic to the in vitro cultivation of bacteria, is the growth medium, which is not suited to reproduce the complex natural environment and microbial dynamics that facilitated the established microbial diversity. The media selection bias promotes the growth of some species or strains [126], while suppressing or completely preventing the growth of others [127, 128]. Moreover, the culture-dependent methods significantly favor the fast-growing self-dependent species, which distorts the microbial composition of the cultivable fraction. Recent studies employing culture-independent approaches have shown that the cultivable fraction poorly represent the microbial community as a whole [129], and enumeration of bacteria by traditional culturing techniques have been shown to produce inaccurate results [130], also with LAB used in cheese production [129, 131, 132].

Many of the different culture-dependent methods that do not utilize DNA sequencing, produce arbitrary results and can rarely be combined or compared between the methods. Even when using the same method, results can be hard to reproduce or to compare between different laboratories. Nowadays, strain-level studies are incorporating WGS, the “gold standard” of culture-dependent methods, which overcomes some of the analytical limitations concerning the cultivable fraction. In constrast, community-level studies are relying more and more on culture-independent methods, based on the direct analysis of DNA or RNA without any culturing prerequisite. The increasing volume of bacterial genomic data available to the scientific community has facilitated the development of culture-independent methods, which in combination with WGS enable studies on the diversity of complex microbial communities at unprecedented resolution and accuracy.
Culture-independent methods, which do not require cultivation of the microorganisms, have emerged to face the limitation of culture-dependent analysis of microbial communities. Most of these methods are initiated by PCR amplification of total DNA or RNA extracted directly from the sample (Figure 1), but non-PCR approaches also exist, like fluorescent in situ hybridization (FISH), which has previously been successfully applied to study the microbial composition in cheese [133, 134]. A number of in vitro culture-independent methods such as denaturing gradient gel electrophoresis (DGGE), temperature gradient gel electrophoresis (TGGE), automated ribosomal intergenic spacer analysis (ARISA), terminal restriction fragment length polymorphisms (T-RFLP), amplified fragment length polymorphism (AFLP) exist, all of which use a fingerprinting approach to differentiate between samples. These methods have all previously been applied to study cheese microbiota [29, 117]. However, novel NGS-based approaches that are applicable to both culture-dependent and culture-independent analysis [31], render many of these fingerprinting approaches obsolete. Indeed, the potential for fast and cost-effective NGS-based genomic analyses of strains and high-resolution analysis of complex microbial communities is extensive. However, the large amounts of genomic data produced using NGS require a substantial amount of in silico processing to produce interpretable results.

Targeting protein-coding genes shared by all members of a species or subspecies instead of the variable regions of 16S rDNA enables identification and differentiation of strains beyond subspecies, but requires extensive work on genomic analyses of a sufficient number of strains in order to identify a conserved yet adequately variable amplicon target. Moreover, the maximum amplicon size of current NGS platforms limits the sensitivity and versatility of the analysis. The alternative to targeted-amplicon sequencing of DNA is shotgun metagenomics sequencing of total DNA, which has the advantage that it is more representative of the microbial community, requires a sequencing depth that is not possible to achieve with any of the current methods, and produces a gigantic amount of data extremely challenging to interpret. In addition, shotgun metagenomic sequencing data include a large proportion of genes encoding basic cell functions, not always related to any specific activity or suitable for differentiating between the members of the microbial community [135]. As such, shotgun metagenomic sequencing only reveals the functional content of the community as a whole, and is commonly complemented with culture-
dependent approaches or targeted-amplicon sequencing of ribosomal DNA to determine the taxonomical diversity of the sample [20, 136, 137].

**Microbial diversity and culture dynamics in milk**

The quality of the final cheese product is dependent on the combined contributions of the starter bacteria, adjunct bacteria, and NSLAB. Bacteriophages infecting *L. lactis* subsp. *lactis* and *L. lactis* subsp. *cremoris* are ubiquitous in dairies, often present in very high titers [86, 138, 139], and can negatively affect the production process and the quality of the final product [85, 86]. However, in causing fermentation failures with DL starter cultures, the diversity of phages, rather than their quantity appears to be more important [86]. Undefined are considered more robust against phage attack compared to defined cultures [19], a characteristic gained from their large number of strains with diverse bacteriophage sensitivities [140]. Throughout history, an evolutionary arms race has existed between phages and their host bacteria promoting genetic diversification of species [141]. Bacteria that are attacked by virulent phages are either killed or survive through; (i) acquiring defensive systems or (ii) evolving to evade the infection. *Vice versa*, phages must evolve to overcome the defensive systems of their host(s), adapt to the diversification of their host(s), or face extinction. The use of frozen or freeze-dried batch starter cultures effectively halts the lactococcal evolution, while phages have the advantage of evolving in the dairy environment [85]. Thus, the dairy industry experiences significant disruption of cheese production due to phage attack. One countermeasure to phage attack is to employ a starter rotation strategy, in which two or more starters with minimal overlap in bacteriophage sensitivity are used alternately [85]. However, choosing which cultures to implement into a rotation strategy is challenging [142], as the microbial diversity and composition of starter cultures is not known beyond sub-species [19, 20]. Characterizing the strain diversity of DL and other undefined starter cultures is of the utmost importance to advise functional culture rotation and predict production performance. Moreover, identifying key starter culture strains central to the character of the product, will improve the capability to assess and predict the impact of phage attack(s).
Aim of study

Undefined mixed (DL) starter cultures used in the production of continental cheeses contain unknown mixtures of *Lactococcus lactis* strains as well as *Leuconostoc* spp.. Starter cultures from different manufacturers give cheeses with qualitatively different characteristics, and each culture shows batch-to-batch performance differences. In addition, bacteriophages infecting starter bacteria can disrupt the production process and negatively affect the quality of the final product. In order to better predict culture performance and assess the impact of phage attach, characterizing the microbial diversity of starter cultures and dairy samples is important, but tools for these analyses are lacking.

The primary aim of this study was to describe the diversity of bacteria and bacteriophages in starter cultures and dairy samples using molecular or sequencing-based methods. In order to achieve this, the work was divided into the following parts:

- High diversity sampling of bacteria and bacteriophages from starter cultures and dairy samples. (PAPER 1).
- Characterize the diversity of *Lactococcus lactis* and *Leuconostoc* in starter cultures using next-generation sequencing technologies and comparative genomics. (PAPER 2 and 3)
- To develop novel methods for culture-independent quantification and differentiation of starter bacteria. (PAPER 2 and 3)
- To apply these novel methods in order to describe and compare the microbial diversity of different DL starter cultures. (PAPER 2 and 3)
Main results and discussions

PAPER I

Use of M17 and a milk-based medium enables isolation of two distinct and diverse populations of Lactococcus lactis strains from undefined mesophilic starter cultures

Undefined mesophilic starter cultures (DL cultures) are used in the production of continental cheeses. These starter cultures contain undefined mixtures of Lactococcus lactis and Leuconostoc strains, where the lactococci are the major contributors in the acidification process. Bacteriophages that infect Lactococcus lactis are ubiquitous in dairies and can disrupt the production process and negatively affect the quality of the final product. Development of dairy starter cultures and bacteriophage research depends on adequate growth media for isolation of relevant starter culture strains and their bacteriophages.

Using M17 and a milk-based medium (GMA), bacterial counts were compared in two commonly used commercially available DL starter cultures (A and B). The bacterial counts were three times higher on GMA than on M17 for culture A, while counts for culture B were the same on both media. This suggested that culture A, but not culture B, had a bacterial subpopulation that would only grow on GMA. The difference in counts for culture A was similar to results described by Erkus et al. (2013) [20], and for this reason, the focus of the remaining analyses in this study was on culture A. The isolated L. lactis strains were used in plaque assays to isolate bacteriophages from bulk starter samples collected at three major Norwegian cheese plants. Plaque assays depend on visual measurements and are difficult to perform using opaque media. Therefore, the GMA isolated had to be transferred to M17 prior to phage studies. Interestingly, as well as inconveniently, most of the bacteria isolated using GMA would not grow in M17 when inoculated directly from the GMA-plate, indicating that bacteria isolated using GMA have properties that are different from the bacteria isolated using M17. However, pure isolates were successfully grown in M17 after two to four passages in GMA without agar, and would readily grow in M17 following transfer.

Using 96 M17- and 96 GMA-bacterial isolates as indicators, 123 bacteriophages, 68 M17-derived, and 55 GMA-derived, were isolated. The phage collections from each media were pooled (Φ-M17 and Φ-GMA) and used in bacteriophage inhibition arrays to assess the
overlap in phage sensitivity between the M17 and GMA subpopulations. A surprisingly low overlap in phage sensitivity was observed between the two bacterial subpopulation. Of the 96 M17 isolates, 68 were inhibited by Φ-M17 but only 7 were inhibited by Φ-GMA. Furthermore, only 8 of the GMA-isolates were inhibited by Φ-M17, while 55 were inhibited by Φ-GMA. This unexpected result prompted us to analyze culture B, which had the same bacterial counts on both media, by the same approach as described for culture A. Although the same bacterial counts were observed on both media for culture B, the analysis showed that both the M17 and the GMA subpopulations contained a number of unique strains. Indeed, the M17- and GMA-subpopulations could be discriminated by their bacteriophage sensitivity. These findings show that both media are biased, and that they complement each other in strain isolation.

The diversity of the M17 and GMA subpopulations of culture A were assessed by using bacteriophage sensitivity as a phenotypic metric (phage typing). The sensitivity spectrum of the 147 sensitive bacterial isolates were analyzed using our 68 M17- and 55 GMA-derived bacteriophages, each bacteriophage applied separately in phage typing plaque assays. The phage typing results show large diversity within each subpopulation, with regards to both bacteria and bacteriophages. Fifty-seven bacterial (47 M17 and 20 GMA) and 85 (50 M17 and 35 GMA) bacteriophage profiles were found. Bacterial isolates that were not sensitive to any of our bacteriophage isolates could not be differentiated from each other. Acidification tests were performed using culture A in the presence of Φ-M17- and Φ-GMA, both separately and in combination. A standard for the acidification process was determined at ΔpH of 0.99 by incubating reconstituted skim milk inoculated with culture A for 4 hours at 30 °C without the presence of bacteriophages. Separately, Φ-M17 and Φ-GMA reduced the ΔpH over 4 hours to 0.79 and 0.80, respectively. In combination, the ΔpH over 4 hours was reduced to 0.67, demonstrating the importance of both bacterial subpopulations in the successful acidification of milk.
Genomic characterization of dairy associated *Leuconostoc* species and diversity of leuconostocs in undefined mixed mesophilic starter cultures

Undefined mixed (DL) starter cultures are composed of predominantly *Lactococcus lactis* and 1–10% *Leuconostoc* spp. The composition of the *Leuconostoc* population in the starter culture ultimately affects the characteristics and the quality of the final product. The genus *Leuconostoc* includes 13 species. Three of these species, *Leuconostoc pseudomesenteroides* (*Ln. pseudomesenteroides*), *Leuconostoc lactis* (*Ln. lactis*), and *Leuconostoc mesenteroides* which is further divided into the four subspecies *mesenteroides* (*Ln. mesenteroides*), *dextranicum* (*Ln. dextranicum*), *cremoris* (*Ln. cremoris*), and *suionicum* (*Ln. suionicum*) are relevant for dairy production using DL cultures, with the exception of *Ln. suionicum* that is associated with Asian wine and sake production. The leuconostocs have not been subject to extensive genomic investigation, and the bases for *Leuconostoc* taxonomy results from cultivation-dependent methods, phenotypic characterization or non-specific molecular methods. Traditionally, *Ln. cremoris* and *Ln. lactis* are reported as the key *Leuconostoc* in DL cultures. However, reports of finding *Ln. mesenteroides*, *Ln. dextranicum*, and *Ln. pseudomesenteroides* in starter cultures and cheese derivates has increased in the recent years. Present-day product information on starter cultures commonly does not include details beyond genus, and the genomic diversity or population dynamics of leuconostocs through the production processes is not known.

In this study, the *Leuconostoc* populations in five DL starter cultures were analyzed using traditional cultivation methods augmented by high-throughput sequencing techniques, and by amplicon sequencing of *eno*, the gene encoding for Enolase, essential for the degradation of carbohydrates via glycolysis. Enumeration on MRS-agar has been reported to underestimate the number of leuconostocs, especially *Ln. cremoris*. To compare bacterial counts in starter cultures, two different media, MRS and milk-containing MPCA were used, both supplemented with vancomycin to select for leuconostocs. In two of the starter cultures (A and D), substantially higher bacterial counts were found on MPCA compared to MRS, while cultures B, C and E had similar counts on both media. Most of the leuconostocs in two of the starter cultures (A and D) were unable to grow on MRS, emphasizing the importance of careful media selection and highlighting the limitations of the culture-based methods.
Forty-six isolates were whole-genome sequenced and analyzed together with thirteen publically available *Leuconostoc* sp. genomes acquired from the National Center of Biotechnology Information (NCBI). Pan-genomic analysis clearly discriminated *Leuconostoc* species and sub-species from each other, and enabled differentiation into twelve robust lineages. These included three lineages of *Ln. cremoris* (C1-C3), four lineages of *Ln. pseudomesenteroides* (P1-P4), four lineages of *Ln. mesenteroides* (M1-M4), and one lineage of *Ln. lactis* (L1). Moreover, the pan-genome analysis revealed that several of the strains previously identified as *Ln. mesenteroides* subspecies were actually *Ln. pseudomesenteroides*, and the NCBI strain LbT16, previously identified as *Ln. cremoris*, was shown to be *Ln. mesenteroides*. The analyses show that the dairy-associated leuconostocs are highly adapted to their environment, clearly differentiated from the non-dairy leuconostocs, and characterized by the acquisition of niche-genotype traits, such as the ability to metabolize citrate. *Ln. cremoris* isolates, when compared to *Ln. mesenteroides* or *Ln. dextranicum* isolates, were missing several genetic elements and contained several truncated genes and deletions, likely the result of a degenerative evolution after a long period of exclusively growing in milk. Interestingly, the analyses did not discriminate *Ln. mesenteroides* from *Ln. dextranicum*, and showed that genomic variation between the isolates was much greater than between the subspecies.

Genomic analyses revealed a multitude of dissimilarities between intra-species lineages. (i) *Ln. cremoris*. All the genomes in *Ln. cremoris* lineages C1-C3 were highly similar to each other. However, phenotypically, C2 and C3 readily grow on MRS, in contrast to C1, which did not. Genetic analysis revealed the absence of four orthologous groups (OGs) *rmlA*, *rmlB*, *rmlC*, and *rmlD* in all C1 isolates. However, these genes are associated with polysaccharide biosynthesis, and their absence does not explain the inability to grow on MRS. (ii) *Ln. mesenteroides*. A large variation in the pan-genomic content of the *Ln. mesenteroides* lineages (M1-M4) was observed. Interestingly, several glucosyl transferases were found within all lineages, several of them encoding for dextranucrases. The ability to produce dextran from sucrose is the phenotypic marker separating *Ln. mesenteroides* from *Ln. dextranicum*. Genotypically, the potential for dextran production was found within all *Ln. mesenteroides* isolates, and could not be used to differentiate between them. Functional comparative analyses showed that the presence of the *cit* operon necessary for metabolism of citrate, and the *lacLM* genes, is a characteristic of dairy-
associated *Ln. mesenteroides*, *Ln. cremoris*, and *Ln. pseudomesenteroides*. All the isolates in lineages M3, and M4 contained both the *cit* operon and the *lacLM* genes, while strains in lineages M1 and M2 were all lacking the *cit* operon, while half of them also lacked the *lacLM* genes. The M4 isolates also contained the deletion in *lacZ*, commonly associated with *Ln. cremoris* type strains. **(iii) *Ln. pseudomesenteroides***. Despite significant pan-genomic differences and core-genomic sequence variation between the *Ln. pseudomesenteroides* lineages (P1-P4), the functional differences were surprisingly few. Lineages P1, P2 and P3 were highly similar to each other, while P4 isolates were missing the genes for reduction of diacetyl, and contained genes for a different capsular or extracellular polysaccharide, when compared to the other three lineages.

Compositional analyses of the *Leuconostoc* population in the five DL starters by targeted-amplicon sequencing of *eno* revealed significant differences between the cultures. Three of the cultures were dominated by *Ln. cremoris* while *Ln. pseudomesenteroides* dominated in the other two. *Ln. mesenteroides* and *Ln. dextranicum* was found in all DL cultures, while *Ln. lactis*, reported to be a major constituent in fermented dairy products, was only identified in one of the cultures.

The compositional differences of the *Leuconostoc* populations between the starter cultures could affect the characteristics of the cheese product. *Ln. cremoris* lacks a wide range of genes involved in carbohydrate metabolism and proteolytic activity, and *Ln. cremoris* and *Ln. pseudomesenteroides* differ significantly in their rate of growth and production of volatile compounds. A better understanding of the microbial composition of starter cultures and the functional dynamics of different dairy leuconostocs may be of great value to the dairy industry and to the starter culture manufacturers.
Diversity of *Lactococcus lactis* in undefined mixed dairy starter cultures revealed by comparative genome analyses and targeted amplicon sequencing of *epsD*

Undefined mixed (DL) starter cultures used in the production of continental cheeses, contain unknown strain mixtures of *Lactococcus lactis* as well as *Leuconostoc* spp., where the lactococci are the major contributors in the acidification process through the fermentation of lactose. Bacteriophages that infect *L. lactis* subsp. *lactis* and *L. lactis* subsp. *cremoris* are ubiquitous in high numbers in the dairy environment, and can disrupt the acidification process and negatively affect the quality of the final product. In industrialized cheese production, frozen or freeze-dried batches of starter cultures are used to ensure predictable and reproducible production. This effectively halts lactococcal evolution, while the phages evolve in the dairy environment. Thus, the dairy industry experiences substantial disruption of the cheese production due to phage attack. One countermeasure to phage attack is to employ a starter rotation strategy, where two or more starters with minimal overlap in phage sensitivity are alternated between. However, because the lactococcal strain diversity and thus the compositional differences between the strain diversity of DL cultures is not known beyond sub-species, deciding which starters to alternate between is challenging.

In this study, pan-genomic analysis of *Lactococcus lactis* isolates obtained from three DL starter cultures in combination with publically available genomes acquired from the National Center of Biotechnology Information (NCBI) enabled differentiation of 21 subsp. *lactis* and 28 subsp. *cremoris* lineages. Most of these lineages were culture specific and phylogenetic analysis of 551 core-genes clearly discriminated dairy and non-dairy lactococci from each other, and also distinguished the DL culture isolates from non-DL culture isolates. Previously, a undefined mesophilic starter culture was divided into seven groups based on AFLP [29], which were later quantified in a metagenome dataset using group-specific gene markers [20]. None of our isolates contained the gene markers specific for TIFN1-6, and only 19 of our isolates contained the gene marker specific for the TIFN7 group. These isolates were scattered amongst several different pan-genomic lineages and consisted of strains isolated on both the M17 and GMA growth media. Interestingly, none of the isolates in lineages C1, C3, C5, C9, C27 and C28 contained any of the gene markers, highlighting the limitation of unique loci as genetic markers when analyzing complex microbial communities.
The *Lactococcus lactis* composition in three DL starter cultures was analyzed by targeted-amplicon sequencing of 16S rDNA, the core gene *purR*, and the softcore gene *epsD*, found to be present in 95 of the 97 starter culture isolates, but absent in most of the reference strains. Using targeted-amplicon sequencing, the downstream data analysis cluster the sequences together into Operational Taxonomic Units (OTUs). The OTU assignments are dependent on the DNA sequence similarity threshold, which can be set by the user, and has traditionally been set at 97% in studies involving 16S rDNA. Several authors have previously pointed out that this threshold is excessively low, and have suggested the use of a higher threshold. To assess the genetic diversity in our three starter cultures A, B and C, amplicon-targeted sequencing of *purR* and *epsD* was performed and clustered using a 99.5% similarity threshold, grouping single-SNP sequences together, but allocating new groups to sequences with SNP distances of 2 and higher. This increased the resolution of the analysis to differentiate between lineages, and revealed substantial differences in the lactococcal composition between the starter cultures.

The 16S rDNA amplicon analysis showed that all cultures to a varying degree were dominated by *L. lactis* subsp. *cremoris*, most prominently culture B with more than 70% *L. lactis* subsp. *cremoris*, and also showed that the content of leuconostocs varied from 1% in culture B to 24.6% in culture A, and 29.4% in culture C. By comparing the *purR* and 16S rDNA amplicon data, a significant underestimation of *L. lactis* subsp. *cremoris* by 16S rDNA was identified in all the samples. The discrepancy varied from 4.5% in the bulk starter of culture C to 15.5% in the frozen culture of culture B. Previous studies have also reported discrepancies in subspecies identification of lactococci using 16S rDNA [25, 124]. Further analysis of 16S rDNA revealed that a number of isolates, which were all identified as subsp. *cremoris* in the pan- and core-genome analysis, contained a novel and unique 16S rDNA sequence more similar to subsp. *lactis* type than subsp. *cremoris*, which leading to the misidentification of isolates in the 16S rDNA analysis, but not in the *purR* analysis. These findings highlight the advantages of using conserved genes instead of, or in combination with 16S rDNA, when analyzing complex lactococcal communities.

The *purR* amplicon analysis enabled relative quantification of 17 OTUs, corresponding to the core-genomic differentiation of strains, and showed considerable differences in the *purR* diversity in the three starter cultures and their corresponding bulk starters. Of the 17 *purR* OTUs, 10 were found in Culture A, 8 in culture B, and 13 in
culture C. Some of the OTUs were culture specific, and accounted for a substantial proportion of the total population. However, the sequence variation within the purR amplicon did not allow for discrimination between many of the lineages. Thus, the variance within the amplicons found among core genes is not high enough to expose the complexity of the DL starter cultures. Therefore, we expanded the analyses to also include softcore-genes, which revealed the highly conserved yet highly variable epsD locus.

The EPS genotype was highly conserved amongst our starter culture isolates and the high sequence variability of the epsD amplicon enabled differentiation between most of the genetic lineages. This allowed unprecedented discrimination of starter culture *Lactococcus lactis*, and revealed significant differences between the three starter cultures as well as compositional shifts during cultivation of cultures in milk. Interestingly, the phylogenetic analysis of epsD did not separate subspecies *lactis* from subspecies *cremoris* at the root of the tree like purR and 16S rDNA. Rather, subspecies separation was made on branches further out on the tree, a strong indication of horizontal gene transfer. The epsD sequences clustered into 52 OTUs, enabling high-resolution quantification of genetic lineage diversity among eps positive strains present in the starter cultures. Of these 52 OTUs, 31 were found in culture A, 28 in culture B, and 18 in culture C. Most of these epsD OTUs, 13 in culture A, 9 in culture B, and 11 in culture C, were culture specific and accounted for a large proportion of the total population in each culture.

In order to better predict production performance and advise functional culture rotation strategies it is important to characterize the strain diversity of DL and other undefined starter cultures. By using comparative genome analyses of whole-genome sequenced *Lactococcus lactis* isolates a robust foundation is made for discovering intra-species gene markers for targeted-amplicon sequencing. The use of purR and epsD as gene markers for *Lactococcus lactis*, enables intra-species differentiation of genetic lineages in undefined mixed mesophilic starter cultures. Combining pan-genome analyses with targeted-amplicon sequencing is an approach that could also be applied to other microbial niches. Using this method, a better understanding of the lactococcal diversity in DL starter cultures can be achieved, which in turn will enable the development of more robust starter cultures and assist in the efforts to maintain the stability and performance of dairy starter cultures.
Conclusion and future perspectives

This thesis provides details on the microbial diversity of DL starter cultures, and advances the capability to analyze microbial communities using next-generation sequencing technologies. Using two complementary growth media, bacteria and bacteriophage strains were isolated and used to estimate the microbial diversity of starter cultures, as well as the diversity of bacteriophages in Norwegian cheese production facilities. Whole-genome sequencing of selected isolates and the subsequent pan-genomic analysis enabled characterization and differentiation of both *Leuconostoc* and *Lactococcus lactis*, facilitating the development of culture-independent targeted-amplicon quantification of different starter cultures and their bulk starters. Our analysis revealed substantial dissimilarities between starter cultures produced by different manufacturers, and demonstrated compositional fluctuations of starter cultures during cultivation of bulk starters. The significant differences in composition between DL starter cultures explains why different starter cultures display a low overlap of bacteriophage sensitivity, and provides a quantitative explanation for the qualitative differences reported for starter cultures by the dairy producers, who often tend to prefer one starter culture to another. Furthermore, the details on the microbial diversity of each starter culture indicate the important distinction between phage quantity and phage diversity in the disrupting acidification of milk. The dairy industry and starter cultures can vastly improve their ability to monitor all phases of starter culture and cheese production by implementing the methods described in this thesis. Routine analysis of the microbial composition of starter cultures will enable quality control of starter cultures, and enable the industry to make competent decisions regarding starter culture rotations in the event of phage attack.

The combined application of culture-dependent analysis and collection of isolates, whole-genome sequencing to perform pan-genomic analysis and the development and use of culture-independent targeted-amplicon sequencing provides robust and exhaustive analysis of microbial communities. Although dairy starter cultures are simple compared to the complexity of other environmental samples such as soil or mammalian gut, starter cultures could serve as a good model for the development of methods for differentiating bacteria in other environments. Considering the past 100 years research on LAB, the application of the genomics approach from the early 1980s and onwards has redefined the science of
molecular biology. The great rate of advancement in next-generation sequencing technologies over the past decade has been accompanied by a rapid development of bioinformatics applications. The reduced cost of sequencing has promoted whole-genome sequencing of bacterial isolates, and the vast improvements to the downstream analysis of genomic data has taken comparative analysis to another level. Next-generation sequencing techniques have been used to overcome the disadvantages of culture-based molecular methods; however, they are not without limitations. While the culture-independent methods provide high-resolution details on the diversity of microbial communities, one limitation in particular, is that they cannot be used to isolate strains, which is an important part of the work with starter cultures and with bacteriophages. Implementation of emerging “omics” approaches, such as transcriptomics, proteomics, and metabolomics, and improvements to genomics that increase the sequencing depth or fragment length of DNA sequencing methods, will greatly improve the scientific capacity to characterize strains and differentiate between the members of microbial communities. Characterization of strains and their nutritional and environmental requirements, in combination with the development of improved selective and suitable media, will help to expand the fraction of cultivable bacteria, and provide details on the selective criteria necessary to achieve better accuracy in the isolation of strains. The capacity to sample increased genetic diversity that can be incorporated into genomic analyses will in turn improve the precision and applicability of the culture-independent methods, and as an important part of the future work with LAB and their bacteriophages.
References


34. Fernández E, Alegría Á, Delgado S, Martín MC, Mayo B. 2011. Comparative phenotypic and molecular genetic profiling of wild Lactococcus lactis subs. lacticis strains of the L. lactis subsp. lactis and L. lactis subsp. cremoris genotypes, isolated
from starter-free cheeses made of raw milk. Applied and Environmental Microbiology 77:5324-5335.


Paper I
Use of M17 and a milk-based medium enables isolation of two distinct and diverse populations of \textit{Lactococcus lactis} strains from undefined mesophilic starter cultures

Cyril Frantzen\textsuperscript{a,}* , Hans Petter Kleppen\textsuperscript{b}, Helge Holo\textsuperscript{a, c}

\textsuperscript{a} Department of Chemistry, Biology and Food Science, Norwegian University of Life Sciences, P.O. Box 5003, N-1432, Ås, Norway
\textsuperscript{b} ACD Pharmaceuticals AS, Leknes, Norway
\textsuperscript{c} TINE SA, Oslo, Norway

\begin{abstract}
Dairy starter culture development and bacteriophage research depends on adequate growth media for isolation of relevant starter culture strains and their bacteriophages. We show that the use of two growth media, M17 and a milk-based medium (GMA), enables isolation of two distinct subpopulations of \textit{Lactococcus lactis} ssp. from undefined mesophilic starter cultures. Phage typing revealed large diversity within each subpopulation, and interestingly, that there was very little overlap in phage sensitivity between the two sets of bacteria. Acidification activity tests performed in the presence of M17- and GMA-derived bacteriophages, both separately and in combination, demonstrated the relevance of both bacterial subpopulations in the acidification process. Use of both media enables higher diversity sampling of \textit{Lactococcus} ssp., and their bacteriophages.
\end{abstract}

© 2015 Elsevier Ltd. All rights reserved.

1. Introduction

Mesophilic mixed (LD-type) starter cultures are used in the production of Dutch-type cheeses. These starter cultures contain undefined mixtures of \textit{Lactococcus lactis} and \textit{Leuconostoc} strains, where the lactococci are the major contributors in the acidification process. Bacteriophages infecting \textit{L. lactis} are ubiquitous in dairies and can negatively affect the production process and quality of the final product (Kleppen, Bang, Nes, & Holo, 2011; Rousseau & Moineau, 2009).

Lactococcal phages are classified into ten groups, where the most frequent groups found in dairy environments are the 936-, c2-, and P335-like bacteriophages (Deveau, Labrie, Chopin, & Moineau, 2006). In an earlier study by our group, no c2-like bacteriophages were found in Norwegian dairies and P335-like bacteriophages were found in titres too low to be of significant consequence in fermentations. However, the strictly lytic 936-like bacteriophages were found in high titres in all bulk starters regardless of fermentation activity, emphasising the relevance of bacteriophage diversity (host range) in disrupting fermentation (Kleppen et al., 2011).

Undefined starter cultures gain their robustness against phage attack from their large number of strains with diverse phage sensitivity (Boucher & Moineau, 2001). To keep up production and maintain reproducibility, the cheese industry prefer frozen batches to back slopping. However, the use of frozen starter culture batches effectively halts lactococcal evolution, whilst giving phages the advantage to evolve freely (Rousseau & Moineau, 2009). The composition of starter cultures is often unknown, and knowledge about the individual strains and their phages is of great importance to the dairy industry. Starter culture lactococci are commonly cultivated on M17, a medium also well suited for isolation of lactococcal phages (Johansen, Øregaard, Serensen, & Derksen, 2015; Terzaghi & Sandine, 1975). However, working with a Dutch mixed starter culture, Erkus et al. (2013) reported that the majority of lactococci could only be isolated on Reddy’s medium (Reddy, Vedamuthu, Washam, & Reinbold, 1972), a medium with milk as one of its ingredients, and not on M17. This prompted us to compare strains and bacteriophages isolated using M17 and the milk-based medium GMA (Hugenholtz, Splint, Konings, & Veldkamp, 1987) from two commonly used undefined mesophilic starter cultures. We show that different phagovars can be isolated on the two
media. Thus, combined use of M17 and GMA enables isolation of a wider diversity of lactococci and their bacteriophages.

2. Materials and methods

2.1. Growth of bacteria and bacteriophages

The media used for cultivation of Lactococcus sp. were M17 (Oxoid, Hampshire, UK) (Terzaghi & Sandine, 1975) supplemented with 0.5% (w/v) lactose (Merck, Oslo, Norway), or 10% (w/v) skimmed milk (TINE SA, Oslo, Norway) supplemented with 50 mM β-glycerophosphate (Sigma–Aldrich, Munich, Germany) (GMA; Hugenholtz et al., 1987). Commercial starter cultures were suspended in Luria–Bertani broth (Bertani, 1951) to an optical density (OD) at 600 nm of 1.0, serially diluted and spread plated on M17 and GMA agar plates in triplicate. Plates were incubated at 22 °C. Frantzen et al. (2016) performed using complete linkage UPGMA using Manhattan distances. The resulting distance-matrix was used to construct a heatmap with dendrograms using the heatmap.2 function from the gplots package (Version 2.16; Warnes et al., 2015) supplemented by the dendextend package (Version 0.18.3; Galili, 2015). A cut-off for the number of clusters was determined using the knee of the curve according to Salvador and Chan (2004) to establish robust clusters.

2.3. Bacteriophage inhibition array

Bacteriophage isolates were distributed into M17-and GMA-collections in microtiter plates and used to phage type isolates that were inhibited in the bacteriophage inhibition arrays. Indicator bacteria were grown in M17-c broth to an OD at 620 nm of about 0.2, and cast in 0.8% (w/v) M17-c soft agar over 1.2% (w/v) M17-c bottom agar. Phage suspensions from the microtiter plates were spotted on top using a stainless steel 48-pin replicator delivering approximately 5 µL per pin. Assays were performed in triplicate. Results were logged as sensitive when clear plaque was observed, or insensitive when no plaque was observed. No turbid plaques were observed. Hierarchal clustering of the results was performed in R (www.r-project.org) with RStudio (RStudio, 2012; Version 0.98.1103; RStudio, Boston, Massachusetts, US) using complete linkage UPGMA using Manhattan distances. The resulting distance-matrix was used to construct a heatmap with dendrograms using the heatmap.2 function from the gplots package (Version 2.16; Warnes et al., 2015) supplemented by the dendextend package (Version 0.18.3; Galili, 2015). A cut-off for the number of clusters was determined using the knee of the curve according to Salvador and Chan (2004) to establish robust clusters.

2.5. Genotyping

Bacterial isolates were typed by Polymerase Chain Reaction (PCR) as described by Mahony et al. (2013) for CWPS-genotyping. Random samples from both bacterial collections were analysed by PCR as described by Erkus et al. (2013) to identify TIFN7-genotype bacterial isolates.

2.6. Starter culture acidification activity tests

Starter culture acidification activity tests were performed in triplicate largely as previously described (IDF, 2009) using reconstituted skim milk 10% (w/v) (RSM), heated to 95 °C for 45 min and cooled to room temperature prior to use. A pre-culture in RSM inoculated with 0.1% (v/v) frozen starter culture and grown overnight at room temperature was diluted 33 times in 10 mL RSM aliquots and incubated at 30 °C for 4 h before pH was measured. The effects of phages were studied by including at least 107 plaque-forming-units (pfu) mL⁻¹ in each assay.

3. Results and discussion

3.1. Bacterial counts on M17 and GMA

Two starter cultures (A and B) commonly used in the production of Dutch-type cheese were used to compare bacterial counts on M17 and a milk-based medium. For the latter we chose GMA, which yielded the same counts for culture A as the more complex milk-based medium of Nickels and Leesment (1964) in earlier experiments (data not shown). For culture A, the counts on GMA were three times higher than on M17 (2.1 × 10⁹ versus 7 × 10⁸ cfu mL⁻¹), while counts on the two media both gave 1.3 × 10⁹ cfu mL⁻¹ for culture B. The difference in counts for culture A was similar to the results described by Erkus et al. (2013). For this reason, the focus of diversity analyses in this study was on culture A.

3.2. Transfer of GMA-isolates to M17 for phage studies

The bacterial counts suggested that culture A, but not culture B, had a bacterial subpopulation that would only grow on milk-based media. However, most of the GMA isolates from both culture A and B would not grow when inoculated directly from the GMA-plate into M17. This demonstrates that for both cultures, many of the strains isolated using GMA (GMA-population), have properties different from those isolated using M17 (M17-population). Phage assays often rely on visual measurements and are difficult to perform using opaque media. Therefore, transfer of the GMA-isolates to M17 prior to phage studies was important. The GMA-isolates were successfully grown in M17 after cultivation in GMA without agar for two to four passages. After successful transfer, isolates would readily grow in M17.

3.3. Bacteriophage isolation

From three major cheese plants producing Dutch-type cheese in Norway, bulk starter samples were collected over a period of two months. The bulk starter samples were stored at 4 °C for up to 12 months or at −20 °C with 15% (w/v) glycerol. Filtered phage lysates and bacteriophages isolated from plaques were stored at 4 °C for 5 days, after which the number of colonies on each plate was counted. Isolates were transferred into the same media without agar, and cultivated in 22 °C with daily transfers until they would grow in M17 (a minimum of two days). Finally, all cultures were transferred to M17, grown for two passages before aliquots were stored frozen at −70 °C in M17 supplemented with 15% (w/v) glycerol (Sigma–Aldrich). Bacteriophages were isolated by plaque assays (Lillehaug, 1997). When propagating phages, M17 was supplemented with 5 mM calcium chloride (Merck, Oslo, Norway) (M17-c). Filtered phage lysates and bacteriophages isolated from plaques were stored at 4 °C for up to 12 months or at −20 °C with 15% (w/v) glycerol.

2.2. Dairy sample collection

Bulk starter samples were collected from several cheese plants and treated as described by Kleppen et al. (2011).
were screened for bacteriophages using bacteriophage inhibition arrays with 96 M17-and 96 GMA-isolates. Inhibited indicators were used as hosts to isolate 68 M17-derived bacteriophage isolates (ɸ-M17) and 55 GMA-derived bacteriophage isolates (ɸ-GMA).

3.4. Low overlap in phage sensitivity between subpopulations

The ɸ-M17 and ɸ-GMA bacteriophage collections were pooled (ɸ-M17pa and ɸ-GMapa) and used in bacteriophage inhibition arrays containing 96 M17-and 96 GMA-bacterial isolates from culture A. The results revealed a surprisingly low overlap in phage sensitivity between the two collections of bacteria. Of the M17 isolates, 68 were inhibited by ɸ-M17pa, but only seven were inhibited by ɸ-GMapa. On the other hand, only eight of the GMA isolates were inhibited by ɸ-M17pa while 55 were inhibited by ɸ-GMapa. This difference in response to the phage mixtures (P < 0.001 by Fisher’s exact test) shows that the two bacterial collections are different. Moreover, if strains isolated using M17 could be isolated on both media, one third of the GMA isolates should respond to phages sensitive to similar to the M17 collection and 23 (68/3) GMA strains should be inhibited on the two media. Using bacteriophage inhibition arrays with 96 M17-and 96 GMA-isolates. Inhibited indicators were screened for bacteriophages using bacteriophage inhibition arrays. Only two of these isolates did not represent a fraction of the GMA collection but are strains not recovered on GMA. This shows that the bacteria isolated on the two media represent distinct subpopulations of the starter culture, each characterised by specific phagovars.

The unexpected bias of GMA prompted us to perform similar analysis on culture B, which diverged significantly from culture A in enumeration on the two media. Using bacteriophage inhibition arrays constructed with culture B bacteria, a mixture of 55 culture B-M17-derived bacteriophages (ɸ-M17pa), and a mixture of 53 culture B-GMA-derived bacteriophages (ɸ-GMapa), culture B was analysed by the same approach as described for culture A. As with culture A, the M17-and GMA-bacteria responded differently to the collections (P < 0.01 by Fisher’s exact test). Of the M17 isolates 55 were sensitive to ɸ-M17 and 32 were sensitive to ɸ-GMA. Of the GMA isolates 36 were sensitive to ɸ-M17 while 53 were sensitive to ɸ-GMA. Thus, although the same counts were recorded on GMA and M17 for culture B, isolation from the two media yielded two subpopulations, where a significant number of strains are unique to its subpopulation. Our data shows that GMA and M17 are biased, and that they complement each other in strain isolation. Use of both media thus enables higher diversity sampling of bacteria and bacteriophages. The coherence of the results from culture A and B indicates that this is a general phenomenon.

3.5. Large diversity of bacteria and bacteriophages in both subpopulations

Metagenomics has become the gold standard in describing microbial diversity, but has major limitations compared with culture based methods and cannot be used for isolating strains. Using bacteriophage sensitivity as a phenotypical metric provides deep and applicable insight into the population diversity of dairy starter cultures. Our M17 and GMA bacteriophage collections were used to phage type the 147 bacterial isolates inhibited in the bacteriophage inhibition arrays. Only two of these isolates did not show sensitivity during phage typing. The phage typing results show large diversity with regards to both bacteria and bacteriophages (Figs. 1 and 2). The number of clusters was established using a 90% similarity stringency cut-off determined according to Salvador and Chan (2004), and the dendrograms drawn for both provided distances within each cluster. With the 90% cut-off, bacterial isolates sensitive to M17 bacteriophages produced 47 phage sensitivity profiles, and the corresponding M17 bacteriophages produced 50 host range profiles (Fig. 1). The GMA bacterial and bacteriophage isolates produced 20 and 35 profiles, respectively (Fig. 2). The profile separation is indicated in Figs. 1 and 2 by the grey-scale gradients between the dendrograms and the heatmaps. In total, 70 bacterial and 82 bacteriophage profiles could be discerned. Only three of the bacterial profiles include isolates sensitive to both bacteriophage collections (Supplementary Fig. S1).

The large diversity of isolates allowed differentiation of phages with a resolution comparable with the MLST-type differentiation as described by Moisan and Moineau (2012), and at the same time robust differentiation of lactococci. From the results, the M17-population appears more diverse than the GMA-population. However, 41 GMA and 4 M17 bacterial isolates were not sensitive to any bacteriophage isolates and could not be differentiated from each other. The phage-typing revealed a wide host-spectrum for some of the bacteriophages, but also demonstrated a large variation in spectra for both subpopulations. Interestingly, some bacteria were sensitive to a wide range of phages. Such isolates can be very useful when isolating, detecting or quantifying bacteriophages. However, the presence of these isolates is ominous as they can provide means for bacteriophages to propagate and evolve within dairy environments.

3.6. M17-and GMA-populations are both important in acidification

The observation that GMA-isolates initially did not grow on M17, combined with the low overlap of phage sensitivity questioned the importance of M17 strains in the acidification process. To investigate this, acidification activity tests in were performed using culture A in the presence of ɸ-M17 and ɸ-GMA bacteriophage collections. Omitting phages and inoculating RSM with culture A was used as standard, providing a ΔpH of 0.99. Two filtered dairy samples derived from fermentation batches with reported complications, and showing high inhibition efficiency in phage inhibition arrays, reduced ΔpH to 0.71 and 0.73, respectively. The results (Table 1) showed that the presence of ɸ-M17 phages inhibited the acidification process and reduced ΔpH to 0.79 whilst the presence of ɸ-GMA reduced ΔpH to 0.80. In the presence of both phage collections in combination, the acidification inhibition was increased to provide a ΔpH of 0.67. This demonstrates the complementary activity of the two subpopulations in acidification, and shows that cultivation using both M17 and GMA is necessary to isolate relevant bacteriophages.

3.7. Confirmation of TIFN7-isolates and high abundance of CWPS Type-1 and Type-2 in both subpopulations

Erkus et al. (2013) reported that the majority of bacterial counts were represented by one lineage of strains (L. lactis subsp. cremoris TIFN7-lineage) only isolated through Reddy’s medium, a complex medium with milk as one of its ingredients. Using the TIFN7-specific primers described by Erkus et al. (2013), isolation of the L. lactis subsp. cremoris TIFN7 lineage was confirmed within both our M17-and GMA-subpopulations (Supplementary Table S1). This contrasts to the culture used in Erkus et al. (2013), which only gave isolates of the TIFN7-genotype on the milk based medium and not M17. Recently, Mahony et al. (2013) implicated a genetic determinant for host specificity of some lactococcal phage isolates within the operon termed the CWPS-operon. Furthermore, Mahony et al. (2013) defined three discriminatory criteria for CWPS-types, where type B (IL/KF) is a common L. lactis subsp. lactis type operon, whilst type C (MG/SK) and type A (UC/CV) are common L. lactis subsp. cremoris type operons. CWPS-typing revealed a high
Fig. 1. Phage-host spectra for the M17 bacteriophages of culture A. All bacterial isolates inhibited by $\phi$-M17 in the bacteriophage inhibition array were analysed by plaque-assays. Bacterial isolates not sensitive to any $\phi$-M17 bacteriophages were omitted from the figure. The black squares indicate sensitivity and the white squares insensitivity. The number of clusters was established by complete-linkage UPGMA with Manhattan distances using a cut-off at 90% similarity. With this cut-off, the bacterial isolates sensitive to M17 bacteriophages produced 47 phage sensitivity profiles, and the corresponding M17 bacteriophages produced 50 host range profiles. This profile separation is indicated by the grey-scale gradient between dendrograms and heatmap. Dendrograms provide clustering of isolates using a stringency of 100% similarity to depict variance within each of the profiles.
abundance of both type B and type C for both M17-and GMA-populations (Supplementary Table S2). CWPS type B was detected once, in the M17-population. Interestingly, no correlation between CWPS-types and phage-host spectra existed within our data. We found different CWPS-types with identical sensitivity spectra, and vice versa.

4. Conclusions

This study shows that choice of culture medium significantly influences the diversity of bacteria isolated from starter cultures and subsequently, the diversity of bacteriophages isolated from dairy samples. Both the M17 and GMA media are biased, but differently so, enabling isolation of two diverse but distinct subpopulations of Lactococcus lactis ssp. strains. The two bacterial subpopulations complement each other and facilitate isolation of a wider diversity of phages. Acidification activity tests performed in the presence of M17- and GMA-derived phages underlines the relevance of both M17 and GMA bacteria in milk acidification. The dairy industry has a demand for novel, preferably defined blends of strains for cheese making, and characterization of starter strains is pivotal to ensure culture performance. Our work demonstrates the
importance of using complementary media in isolation of starter culture strains.

Acknowledgements

This work was supported by the Norwegian Research Council and by TINE SA. We acknowledge Michelle Høgger for her work with starter culture B, and thank Hailay Gebreselassie for his assistance in the lab. We are grateful to TINE SA for providing dairy samples.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.idairyj.2015.09.005.

References

### Table S1

**TIFN7-genotyping results.**

<table>
<thead>
<tr>
<th>Name</th>
<th>PCR result</th>
<th>Name</th>
<th>PCR result</th>
</tr>
</thead>
<tbody>
<tr>
<td>M17-01</td>
<td>-</td>
<td>GMA-01</td>
<td>-</td>
</tr>
<tr>
<td>M17-11</td>
<td>-</td>
<td>GMA-02</td>
<td>+</td>
</tr>
<tr>
<td>M17-12</td>
<td>-</td>
<td>GMA-03</td>
<td>-</td>
</tr>
<tr>
<td>M17-13</td>
<td>-</td>
<td>GMA-04</td>
<td>-</td>
</tr>
<tr>
<td>M17-14</td>
<td>-</td>
<td>GMA-05</td>
<td>-</td>
</tr>
<tr>
<td>M17-16</td>
<td>-</td>
<td>GMA-06</td>
<td>-</td>
</tr>
<tr>
<td>M17-18</td>
<td>-</td>
<td>GMA-08</td>
<td>+</td>
</tr>
<tr>
<td>M17-20</td>
<td>-</td>
<td>GMA-09</td>
<td>-</td>
</tr>
<tr>
<td>M17-21</td>
<td>-</td>
<td>GMA-10</td>
<td>-</td>
</tr>
<tr>
<td>M17-22</td>
<td>-</td>
<td>GMA-13</td>
<td>-</td>
</tr>
<tr>
<td>M17-24</td>
<td>-</td>
<td>GMA-14</td>
<td>-</td>
</tr>
<tr>
<td>M17-25</td>
<td>-</td>
<td>GMA-15</td>
<td>-</td>
</tr>
<tr>
<td>M17-27</td>
<td>-</td>
<td>GMA-16</td>
<td>-</td>
</tr>
<tr>
<td>M17-30</td>
<td>-</td>
<td>GMA-17</td>
<td>+</td>
</tr>
<tr>
<td>M17-31</td>
<td>+</td>
<td>GMA-18</td>
<td>-</td>
</tr>
<tr>
<td>M17-32</td>
<td>-</td>
<td>GMA-19</td>
<td>+</td>
</tr>
<tr>
<td>M17-34</td>
<td>-</td>
<td>GMA-20</td>
<td>+</td>
</tr>
<tr>
<td>M17-35</td>
<td>-</td>
<td>GMA-21</td>
<td>-</td>
</tr>
<tr>
<td>M17-36</td>
<td>-</td>
<td>GMA-26</td>
<td>-</td>
</tr>
<tr>
<td>M17-38</td>
<td>-</td>
<td>GMA-28</td>
<td>-</td>
</tr>
<tr>
<td>M17-40</td>
<td>-</td>
<td>GMA-32</td>
<td>+</td>
</tr>
<tr>
<td>M17-41</td>
<td>-</td>
<td>GMA-33</td>
<td>+</td>
</tr>
<tr>
<td>M17-43</td>
<td>-</td>
<td>GMA-34</td>
<td>-</td>
</tr>
<tr>
<td>M17-44</td>
<td>-</td>
<td>GMA-37</td>
<td>-</td>
</tr>
<tr>
<td>M17-45</td>
<td>+</td>
<td>GMA-39</td>
<td>+</td>
</tr>
<tr>
<td>M17-53</td>
<td>-</td>
<td>GMA-41</td>
<td>+</td>
</tr>
<tr>
<td>M17-54</td>
<td>-</td>
<td>GMA-44</td>
<td>-</td>
</tr>
<tr>
<td>M17-56</td>
<td>-</td>
<td>GMA-47</td>
<td>-</td>
</tr>
<tr>
<td>M17-57</td>
<td>+</td>
<td>GMA-48</td>
<td>+</td>
</tr>
<tr>
<td>M17-06</td>
<td>-</td>
<td>GMA-51</td>
<td>+</td>
</tr>
<tr>
<td>M17-61</td>
<td>-</td>
<td>GMA-52</td>
<td>-</td>
</tr>
<tr>
<td>M17-67</td>
<td>-</td>
<td>GMA-60</td>
<td>-</td>
</tr>
<tr>
<td>M17-70</td>
<td>-</td>
<td>GMA-61</td>
<td>-</td>
</tr>
<tr>
<td>M17-71</td>
<td>-</td>
<td>GMA-64</td>
<td>-</td>
</tr>
<tr>
<td>M17-77</td>
<td>-</td>
<td>GMA-81</td>
<td>-</td>
</tr>
<tr>
<td>M17-80</td>
<td>+</td>
<td>GMA-82</td>
<td>-</td>
</tr>
<tr>
<td>M17-82</td>
<td>-</td>
<td>GMA-95</td>
<td>-</td>
</tr>
<tr>
<td>M17-83</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M17-87</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M17-88</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M17-90</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M17-92</td>
<td>-</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table S2

CWPS-genotyping results.

<table>
<thead>
<tr>
<th>M17 isolates</th>
<th>GMA isolates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Name</td>
<td>Name</td>
</tr>
<tr>
<td>M17-1</td>
<td>GMA-1</td>
</tr>
<tr>
<td>M17-2</td>
<td>GMA-2</td>
</tr>
<tr>
<td>M17-3</td>
<td>GMA-3</td>
</tr>
<tr>
<td>M17-4</td>
<td>GMA-4</td>
</tr>
<tr>
<td>M17-5</td>
<td>GMA-5</td>
</tr>
<tr>
<td>M17-6</td>
<td>GMA-6</td>
</tr>
<tr>
<td>M17-7</td>
<td>GMA-7</td>
</tr>
<tr>
<td>M17-8</td>
<td>GMA-8</td>
</tr>
<tr>
<td>M17-9</td>
<td>GMA-9</td>
</tr>
<tr>
<td>M17-10</td>
<td>GMA-10</td>
</tr>
<tr>
<td>M17-11</td>
<td>GMA-11</td>
</tr>
<tr>
<td>M17-12</td>
<td>GMA-12</td>
</tr>
<tr>
<td>M17-13</td>
<td>GMA-13</td>
</tr>
<tr>
<td>M17-14</td>
<td>GMA-14</td>
</tr>
<tr>
<td>M17-15</td>
<td>GMA-15</td>
</tr>
<tr>
<td>M17-16</td>
<td>GMA-16</td>
</tr>
<tr>
<td>M17-17</td>
<td>GMA-17</td>
</tr>
<tr>
<td>M17-18</td>
<td>GMA-18</td>
</tr>
<tr>
<td>M17-19</td>
<td>GMA-19</td>
</tr>
<tr>
<td>M17-20</td>
<td>GMA-20</td>
</tr>
<tr>
<td>M17-21</td>
<td>GMA-21</td>
</tr>
<tr>
<td>M17-22</td>
<td>GMA-22</td>
</tr>
<tr>
<td>M17-23</td>
<td>GMA-23</td>
</tr>
<tr>
<td>M17-24</td>
<td>GMA-24</td>
</tr>
<tr>
<td>M17-25</td>
<td>GMA-25</td>
</tr>
<tr>
<td>M17-26</td>
<td>GMA-26</td>
</tr>
<tr>
<td>M17-27</td>
<td>GMA-27</td>
</tr>
<tr>
<td>M17-28</td>
<td>GMA-28</td>
</tr>
<tr>
<td>M17-29</td>
<td>GMA-29</td>
</tr>
<tr>
<td>M17-30</td>
<td>GMA-30</td>
</tr>
<tr>
<td>M17-31</td>
<td>GMA-31</td>
</tr>
<tr>
<td>M17-32</td>
<td>GMA-32</td>
</tr>
<tr>
<td>M17-33</td>
<td>GMA-33</td>
</tr>
<tr>
<td>M17-34</td>
<td>GMA-34</td>
</tr>
<tr>
<td>M17-35</td>
<td>GMA-35</td>
</tr>
<tr>
<td>M17-36</td>
<td>GMA-36</td>
</tr>
<tr>
<td>M17-37</td>
<td>GMA-37</td>
</tr>
<tr>
<td>M17-38</td>
<td>GMA-38</td>
</tr>
<tr>
<td>M17-39</td>
<td>GMA-39</td>
</tr>
<tr>
<td>M17-40</td>
<td>GMA-40</td>
</tr>
<tr>
<td>M17-41</td>
<td>GMA-41</td>
</tr>
<tr>
<td>M17-42</td>
<td>GMA-42</td>
</tr>
<tr>
<td>M17-43</td>
<td>GMA-43</td>
</tr>
<tr>
<td>M17-44</td>
<td>GMA-44</td>
</tr>
<tr>
<td>M17-45</td>
<td>GMA-45</td>
</tr>
<tr>
<td>M17-46</td>
<td>GMA-46</td>
</tr>
<tr>
<td>M17-47</td>
<td>GMA-47</td>
</tr>
<tr>
<td>M17-48</td>
<td>GMA-48</td>
</tr>
<tr>
<td>M17-49</td>
<td>GMA-49</td>
</tr>
<tr>
<td>M17-50</td>
<td>GMA-50</td>
</tr>
<tr>
<td>M17-51</td>
<td>GMA-51</td>
</tr>
<tr>
<td>M17-52</td>
<td>GMA-52</td>
</tr>
<tr>
<td>M17-53</td>
<td>GMA-53</td>
</tr>
<tr>
<td>M17-54</td>
<td>GMA-54</td>
</tr>
</tbody>
</table>
Supplementary Figure 1: Phage-host spectra for the M17 and GMA bacteriophages

**Fig. S1.** Combined phage-host spectra for the M17 and GMA bacteriophages of culture A. All bacterial isolates inhibited in the bacteriophage inhibition array were analysed by plaque-assays. Bacterial isolates not sensitive to any bacteriophages were omitted from the figure. The black squares indicate sensitivity and the white squares insensitivity. The number of clusters was established by complete-linkage UPGMA with Manhattan distances using a cut-off at 90% similarity. With this cut-off, the bacterial isolates produced 70 phage sensitivity profiles, and the bacteriophages produced 82 host range profiles. This profile separation is indicated by the grey-scale gradient between dendrograms and heatmap. Dendrograms provide clustering of isolates using a stringency of 100% similarity to depict variance within each of the profiles.
Paper II
Genomic Characterization of Dairy Associated Leuconostoc Species and Diversity of Leuconostocs in Undefined Mixed Mesophilic Starter Cultures

Cyril A. Frantzen, Witold Kot, Thomas B. Pedersen, Ylva M. Ardö, Jeff R. Broadbent, Horst Neve, Lars H. Hansen, Fabio Dal Bello, Hilde M. Østlie, Hans P. Kleppen, Finn K. Vogensen and Helge Holo

Undefined mesophilic mixed (DL-type) starter cultures are composed of predominantly Lactococcus lactis subspecies and 1–10% Leuconostoc spp. The composition of the Leuconostoc population in the starter culture ultimately affects the characteristics and the quality of the final product. The scientific basis for the taxonomy of dairy relevant leuconostocs can be traced back 50 years, and no documentation on the genomic diversity of leuconostocs in starter cultures exists. We present data on the Leuconostoc population in five DL-type starter cultures commonly used by the dairy industry. The analyses were performed using traditional cultivation methods, and further augmented by next-generation DNA sequencing methods. Bacterial counts for starter cultures cultivated on two different media, MRS and MPCA, revealed large differences in the relative abundance of leuconostocs. Most of the leuconostocs in two of the starter cultures were unable to grow on MRS, emphasizing the limitations of culture-based methods and the importance of careful media selection or use of culture independent methods. Pan-genomic analysis of 59 Leuconostoc genomes enabled differentiation into twelve robust lineages. The genomic analyses show that the dairy-associated leuconostocs are highly adapted to their environment, characterized by the acquisition of genotype traits, such as the ability to metabolize citrate. In particular, Leuconostoc mesenteroides subspp. cremoris display telltale signs of a degenerative evolution, likely resulting from a long period of growth in milk in association with lactococci. Great differences in the metabolic potential between Leuconostoc species and subspecies were revealed. Using targeted amplicon sequencing, the composition of the Leuconostoc population in the five commercial starter cultures was shown to be significantly different. Three of the cultures were dominated by Ln. mesenteroides subspp. cremoris. Leuconostoc pseudomesenteroides dominated in two of the
INTRODUCTION

Mesophilic mixed (DL-type) starter cultures used in the production of Dutch-type cheeses are composed of undefined mixtures of homofermentative Lactococcus lactis subsp. lactis (Lc. lactis), Lactococcus lactis subsp. cremoris (Lc. cremoris), Lactococcus lactis subsp. lactis biovar. diacetylactis (Lc. diacetylactis) and heterofermentative Leuconostoc spp. The latter two provide aroma and texture by metabolizing citrate, producing diacetyl, acetoin and CO₂, while Lc. cremoris and Lc. lactis are the major acid producers through fermentation of lactose. In many cheeses, diacetyl is an important aroma compound, and CO₂ is important for the eye formation (Hugenholtz, 1993). In fermented dairy products, Leuconostoc grows in association with the acid-producing lactococci and have been suggested to play a role in promoting the growth of citrate positive Lactococcus strains (Vedamuthu, 1994; Bandell et al., 1998; Hache et al., 1999). The importance of Leuconostoc in cheese production is widely recognized. DL-type starter cultures are predominantly Lactococcus spp., Leuconostoc spp. commonly accounting for 1–10% of the starter culture population (Cogan and Jordan, 1994). However, knowledge on the species diversity of Leuconostoc included in these starter cultures, or the composition of Leuconostoc through the culture production is sparse. Due to the low initial number and relatively weak ability to ferment lactose, Leuconostoc spp. are not believed to have a significant effect in the acidification process in the early stages of cheese making (Ardö and Varming, 2010). However, leuconostocs have been shown to dominate the cheese microbiota in the later stages of acidification with added propionic acid bacteria (Porcellato et al., 2013; Østlie et al., 2016). The genus Leuconostoc is comprised of 13 species, with the species Leuconostoc mesenteroides divided into subspecies mesenteroides, dextranicum, cremoris, and suionicum (Hemme and Foucaud-Scheunemann, 2004; Gu et al., 2012). The Leuconostoc species (or subspecies) relevant for dairy production are Leuconostoc mesenteroides subsp. mesenteroides (Ln. mesenteroides), Leuconostoc mesenteroides subsp. dextranicum (Ln. dextranicum), Leuconostoc pseudomesenteroides subsp. cremoris (Ln. cremoris), Leuconostoc pseudomesenteroides (Ln. pseudomesenteroides) and Leuconostoc lactis (Ln. lactis) (Cogan and Jordan, 1994; Thunell, 1995).

The bases for Leuconostoc taxonomy are results from cultivation-dependent methods, followed by phenotypic/biochemical characterization or non-specific molecular methods. In addition to being tedious and time-consuming, classical cultivation-dependent methods are known to underestimate the number of Leuconostoc spp., especially Ln. cremoris (Vogensen et al., 1987; Ward et al., 1990; Auty et al., 2001). In addition, concerns on the lack of stability and reproducibility of phenotypical methods have been raised (Thunell, 1995; Barrangou et al., 2002). Several molecular typing methods, such as RAPD, PFGE, RFLP, Rep-PCR, MLST, MALDI-TOF MS, plasmid profiling and 16S rRNA targeted differentiation have been employed to characterize or identify Leuconostoc isolates (Villani et al., 1997; Björkroth et al., 2000; Cibik et al., 2000; Pérez et al., 2002; Sánchez et al., 2005; Vihavainen and Björkroth, 2009; Nieto-Arribas et al., 2010; Alegria et al., 2013; Zeller-Péronnet et al., 2013; Dan et al., 2014; Zhang et al., 2015). However, most of these techniques requiring a preliminary stage of cultivation and comparison of results between the methods and between different laboratories remains challenging. Often, these methods were developed to work with only one or two species of Leuconostoc, so they do not provide subspecies differentiation, yield inconclusive results, yield results that are hard to reproduce, or provide arbitrary differentiation of isolates not sufficiently tethered to phenotypic traits. So far, the work by Dr. Ellen Garvie on the growth and metabolism of Leuconostoc spp. (Garvie, 1960, 1967, 1969, 1979, 1983; Garvie et al., 1974), and DNA-DNA hybridization studies (Farrow et al., 1989) remains the basis for the taxonomical division of dairy relevant leuconostocs.

The Leuconostoc genus has also not been subject to extensive genomic research, and information on the genomic diversity or species population dynamics through the cheese production processes is scarce if available at all. Scientific literature and product information on starter cultures pre-dating the genomic age list Ln. cremoris and Ln. lactis as the key Leuconostoc in undefined mixed mesophilic starter cultures (Lodics and Steenson, 1990; Johansen and Kibeniich, 1992; Vedamuthu, 1994). However, in recent years, isolation of Ln. mesenteroides, Ln. dextranicum, and Ln. pseudomesenteroides is more common from starter cultures or from cheese derivatives (Olsen et al., 2007; Kleppen et al., 2012; Pedersen et al., 2014a,b; Østlie et al., 2016).

Here we present genomic comparative analysis of Leuconostoc spp. and present data on the diversity and composition of Leuconostoc populations in five commercially available DL-type starter cultures. Using traditional cultivation methods in combination with high-throughput sequencing techniques, we provide robust species and subspecies differentiation, and direct population composition analysis using targeted amplicon-sequencing techniques. To our knowledge, this is the first in-depth genomic work performed on the Leuconostoc genus, and the first data published on Leuconostoc diversity in DL-type starter cultures.
MATERIALS AND METHODS

Cultivation of Bacterial Strains and Starter Cultures
All bacterial strains used in this study are listed in Supplementary Table S1. The two different media used for cultivation were de Man Rogosa Sharpe (MRS) (Difco, Detroit, Michigan, USA), and modified PCA (MPCA). PCA (Sigma-Aldrich, Oslo, Norway) was supplemented with 0.5 g/L Tween 80, 5.0 g/L ammonium-citrate, 1 g/L skim milk powder (TINE SA, Oslo, Norway), 0.04 g/L FeSO₄, 0.2 g/L MgSO₄, 0.05 g/L MnSO₄, and 10.0 g/L glucose. Glucose was sterile filtered separately and added after autoclaving. Both media were supplemented with 40 μg/mL vancomycin to select for *Leuconostoc*. Three separate extractions from one batch of each starter cultures (A, B, C, D, and E) were suspended in MPCA to an optical density at 600 nm (OD₆₀₀) of 1.0, serially diluted in 10% (w/v) skim milk and spread plated on MRS and MPCA agar plates in triplicate. The plates were incubated at 22 °C for 5 days before colony enumeration. Isolates were transferred to MRS and MPCA broth media, respectively, and cultivated at 22 °C for two passages before aliquots were supplemented with 15% (w/v) glycerol (Sigma-Aldrich) and stored at −70°C.

Genome Sequencing, Assembly, and Annotation
Genomic DNA from *Leuconostoc* isolates was extracted from 1 mL of overnight culture using Qiagen DNeasy Blood & Tissue Kit (Qiagen, Hilden, Germany). The cells were lysed with 40 mg/mL lysozyme (Qiagen, Hilden, Germany) and bead-beating in a FastPrep®-24 (MP Biomedicals, Santa Ana, California) using 0.5 g acid-washed beads (<106 μm) (Sigma-Aldrich) prior to column purification. DNA libraries were made using the Nextera XT DNA Sample Prep kit (Illumina, San Diego, California, USA) according to manufacturer instructions and sequenced with Illumina MiSeq (Illumina, San Diego, California, USA) using V3 chemistry for 33 isolates sequenced at the Norwegian University of Life Sciences, and V2 chemistry for 13 isolates sequenced at the Aarhus University. Raw sequences were adapter trimmed, quality filtered (Q>20), de novo assembled using SPAdes V3.7.1 (Nurk et al., 2013) and annotated using the Prokka pipeline (Seemann, 2014). Contigs shorter than 1000 bp or with < 5 times coverage were removed from each assembly prior to gene annotation. Thirteen publicly available genomes of *Leuconostoc* obtained from the National Center for Biotechnology Information (NCBI) database were also included in the dataset (Jung et al., 2012; Meslier et al., 2012; Erkus et al., 2013; Pedersen et al., 2014a,b; Campedelli et al., 2015; Østlie et al., 2016). This whole genome project has been deposited at DDBJ/ENA/GenBank under the BioProject PRJNA352459.

Genomic Analysis
The protein coding sequences of all *Leuconostoc* isolates were compared by an all-against-all approach using BLASTP (Camacho et al., 2009) and grouped into orthologous clusters using GET_HOMOLOGUES (Version 2.0.10) (Contreras-Moreira and Vinuesa, 2013). Pan and core genomes were estimated using the pan-genomic analysis tool PanGP v.1.0.1 (Zhao et al., 2014). Orthologous groups (OGs) were identified via the Markov Cluster Algorithm (MCL) with an inflation value of 1.5 (Enright et al., 2002) and intersected using the compare_clusters.pl script provided with GET_HOMOLOGUES. The orthologous clusters were curated to exclude significantly divergent singletons, which is likely the result of erroneous assembly or annotation. A presence/absence matrix for each gene cluster and each genome was constructed for the pan-genome before statistical and clustering analysis of the matrix was performed in R (http://www.r-project.org/). Hierarchical clustering of the pan-genome matrix was performed using complete-linkage UPGMA with Manhattan distances, and a distance cut-off for the number of clusters was determined using the knee of the curve approach (Salvador and Chan, 2004), binning the isolates into genomic lineages. The resulting distance-matrix was used to construct a heatmap with dendrograms using the heatmap.2 function included in the Gplots package (Version 2.16; Warnes et al., 2015) supplemented by the Dendextend package (Version 0.18.3; Galili, 2015).

Comparative Genomics Analysis
The genetic potential of individual *Leuconostoc* lineages that were identified by the pan-/core-genome analysis was investigated by producing intra-lineage pan-genomes using GET_HOMOLOGUES (Version 2.0.10). The pan-genome for each lineage was analyzed using Blast2GO v4 (Conesa et al., 2005) to identify functionality, and Geneious 8.1.8 (Kearse et al., 2012) to identify sequence variation within orthologous clusters. The lineage pan-genomes were then compared using KEGG databases (Kanehisa and Goto, 2000) and the functional comparative comparison tool found in The SEED Viewer (Overbeek et al., 2014). CRISPR sequences and spacers were identified using the CRISPRFinder tool (Grissa et al., 2007).

Relative Quantification of *Leuconostoc* Species in Starter Cultures
Compositional analysis of *Leuconostoc* in five commercially available starter cultures was performed in triplicates on total DNA isolated from the starter cultures using 1 mL of starter culture diluted to an OD₆₀₀ of 1. The cultures were treated with 20 mg/mL lysozyme (Sigma-Aldrich) and 3U/L mutanolysin (Sigma-Aldrich), mechanically lysed using FastPrep (MP Biomedicals) with 0.5 g of acid-washed beads (<106 μm) (Sigma-Aldrich) and purified using the Qiagen DNeasy Blood & Tissue Kit (Qiagen). A suitable amplicon target was identified by screening the core-genome for nucleotide sequence variation using the sequence alignment metrics functions available in the DECIPHER package v1.16.1 (Wright, 2015). Genes without flanking consensus regions within a 500 bp variable region adequate for differentiation, or did not provide sufficient discrimination from similar sequences in species likely to be present in dairy, were excluded. The locus eno encoding for enolase was amplified by PCR using the KAPA HiFi PCR Kit (KAPA Biosystems, Wilmington, Massachusetts, USA) with primers Eno-F (5′-AACACGAAGCTGTGGAATTTGCGTG-3′) and Eno-R (5′-GCCAATCCACCTTCATCAACAACTGA-3′). Forward (5′-TCGTGGAGGCGTACAGATGTTAAGACAG-3′) and
reverse (5’ GTCTCGTGCTGAGATGTTAAGAGACAG-) Illumina adapter overhangs were added to the 5' end of the primers to allow for Nextera XT DNA indexing of the PCR-products. The resulting libraries were sequenced on an Illumina MiSeq with V3 (2 × 300 bp) reagents. The resulting data were paired-end-joined and quality filtered using PEAR (Zhang et al., 2014) and clustered with a 100% identity level threshold using usearch v7 (Edgar, 2010) with error-minimization from uparse (Edgar, 2013). The resulting sequences were matched against a local BLAST-database produced from the Leuconostoc genomes for identification.

RESULTS

**Leuconostoc in Dairy Starters**

Enumeration on MRS-agar has been reported to underestimate the number of leuconostocs, especially *Ln. cremoris* (Vogensen et al., 1987; Ward et al., 1990; Auty et al., 2001). Bacterial counts were compared in five starter cultures (A, B, C, D, and E) commonly used in the production of Dutch-type cheeses using MRS and MPCA agar with 40 µg/mL vancomycin. The results (Figure 1) showed large differences in the counts between starter cultures for the two media. Cultures A and D gave substantially higher counts on MPCA compared to MRS, while cultures B, C, and E had similar counts on both media. Thus, cultures A and D seemed to contain a large number of *Leuconostoc* strains unable to grow on MRS, while cultures B, C, and E did not.

**Genome Sequencing and Pan-Genomic Analysis**

*Leuconostoc* diversity was investigated by whole-genome sequencing of 20 isolates picked from MPCA- and MRS-plates of cultures A and D, and 26 isolates from cheese, including Dutch-type cheese produced using cultures B, C, and E. Lastly, 13 publicly available *Leuconostoc* spp. genomes were included in the dataset. All 59 *Leuconostoc* genomes were annotated and the coding sequences (CDS) were compared by a blast-all-against-all approach to identify OGs. Pan- and core-genomes were estimated (Figure 2) using the pan-genomic analysis tool PanGP. After curation, the pan-genome was determined to consist of 4415 OGs, and a core-genome was found to comprise 638 OGs. Differentiation of isolates using hierarchical clustering on the pan-matrix clearly separated *Leuconostoc* species and sub-species (Figure 3). Several of the strains previously identified as *Ln. mesenteroides* subspecies were shown to be *Ln. pseudomesenteroides* by the genomic analysis. Moreover, the NCBI strain LbT16 previously identified as *Ln. cremoris*, was an outlier to the *Ln. cremoris* species branch and was identified in the pan-genomic analysis as *Ln. mesenteroides*. This was further confirmed by alignment of the full-length 16S rRNA, revealing a 100% identity between *Ln. cremoris* LbT16 and *Ln. mesenteroides* type 16S rRNA. Based on sequence similarity and gene content, the pan-genomic clustering divided the 59 leuconostocs into 12 robust *Leuconostoc* lineages across the genus. These included three lineages of *Ln. cremoris* (C1-C3), four lineages of *Ln. pseudomesenteroides* (P1-P4), four lineages of *Ln. mesenteroides* (M1-M4), and one lineage of *Ln. lactis* (L1). The *Ln. cremoris* TIFN8 genome was excluded from further analysis because the genome data contained a high number of fragmented genes and redundant sequences, making it an outlier.

The differences between lineages (Table 1), species and subspecies level (in the case for *Ln. mesenteroides* subs.) include significantly smaller genomes for *Ln. cremoris* and *Ln. lactis* (1.6–1.8 Mb) compared to *Ln. mesenteroides*, *Ln. dextranicum,*

---

**FIGURE 1** | Bacterial counts for five starter cultures A–E on MRS and MPCA supplemented with vancomycin to select for *Leuconostoc*. The counts are the mean of three separate extractions made from the same culture batch and the error bar indicates the standard deviation. The blue bars represent the bacterial counts on MPCA, while the orange bars represent the bacterial counts on MRS. The Y-axis is cut at 1,0E+06 for better readability.
and \textit{Ln. pseudomesenteroides} (1.8–2.2 Mb). Moreover, the larger genome found in the latter three species contained up to 400 more coding sequences (CDS) than \textit{Ln. cremoris} and \textit{Ln. lactis}. Analysis of functional genomics indicated a closer relationship between \textit{Ln. lactis} and \textit{Ln. pseudomesenteroides}, than that of \textit{Ln. mesenteroides}. Comparison of genetic potential within and between the \textit{Ln. mesenteroides} subspecies showed only minor differences between \textit{Ln. mesenteroides} and \textit{Ln. dextranicum}. Rather, as shown in Figure 3, the variation between the isolates was much greater than the difference between \textit{Ln. mesenteroides} and \textit{Ln. dextranicum}. On the other hand, substantial difference was found between isolates of dairy origin and non-dairy origin. This environment adaptation was also observed for \textit{Ln. lactis}, where \textit{Ln. lactis} 91922, isolated from kimchi was clearly distinguishable from LN19 and LN24 isolated from dairy. Comparison of \textit{Ln. cremoris} and other \textit{Ln. mesenteroides} subspecies isolates revealed that a range of genetic elements found in these species that were missing in \textit{Ln. cremoris}. Apart from some enzymes encoding for rhamnose-containing glucans, \textit{Ln. cremoris} isolates did not have any genetic functionality absent in \textit{Ln. mesenteroides} or \textit{Ln. dextranicum}. Moreover, several truncated genes and deletions were found in \textit{Ln. cremoris} isolates, likely the result of a degenerative evolutionary process through a long period of growth in the milk environment.

**Comparative Genomics of Intra-Species \textit{Leuconostoc} Lineages**

To explore differences in functional genetic potential between the lineages within the species and subspecies, comparative analysis of intra-lineage pan-genomes was performed. The results are included in Supplementary Table S2.

**(I) \textit{Ln. cremoris} Lineages**

Comparison of the genetic content for \textit{Ln. cremoris} lineages showed that \textit{Ln. cremoris} C1, C2, and C3 were highly similar and differentiated from each other mostly because of sequence variation in shared OGs. \textit{Ln. cremoris} C1 (MPCA-type), which did not grow on MRS was missing four OGs found in both lineage C2 and C3 (MRS-type). These OGs were annotated \textit{rmlA, rmlB, rmlC,} and \textit{rmlD}, encoding for four enzymes identified in the subsystem “rhamnose containing glycans.” These enzymes are associated with polysaccharide biosynthesis and their presence likely does not explain the inability of C1-type strains to grow on MRS.

**(II) \textit{Ln. mesenteroides} and \textit{Ln. dextranicum} Lineages**

Comparison of the genetic content showed a large variance between and within the \textit{Ln. mesenteroides} lineages. Interestingly, no major difference between subspecies \textit{Ln. mesenteroides} and \textit{Ln. dextranicum} was found. \textit{Ln. dextranicum} 20484 is grouped together with \textit{Ln. mesenteroides} isolates LN32 and LN34, while \textit{Ln. dextranicum} LBe16 is grouped together with \textit{Ln. mesenteroides} LBe15 and LN08. This subspecies segregation of \textit{Ln. dextranicum} and \textit{Ln. mesenteroides} was based on the phenotypical ability to produce dextran from sucrose. Dextranase, the enzyme involved in this process, is a glucosyltransferase that catalyzes the transfer of glucosyl residues from sucrose to a dextran polymer and releases fructose. Several glucosyltransferases were found within all \textit{Ln. mesenteroides} isolates included in this study, among them several genes encoding for dextranases with 40–67% amino acid identity to each other. Genotypically, the potential for dextran production exists within many if not all \textit{Ln. mesenteroides} isolates, and does not differentiate \textit{Ln. mesenteroides} from \textit{Ln. dextranicum}. This finding was manifest by the separation of \textit{Ln. mesenteroides} and \textit{Ln. dextranicum} isolates into four lineages. Functional comparative analyses showed that the presence of the \textit{cit} operon necessary for metabolism of citrate, and the \textit{lacLM} genes is a characteristic of dairy-associated \textit{Ln. mesenteroides}, \textit{Ln. cremoris} and \textit{Ln. pseudomesenteroides}. In all of the strains in lineages M3 and M4, both the \textit{cit} operon and the \textit{lacLM} genes were present, while strains in lineages M1 and M2 were lacking the \textit{cit} operon, and half of them also lacked the \textit{lacLM} genes. Furthermore, the strains in lineages M1 and M2 contained the genetic potential for metabolism of arabinose, and the two isolates J18 and ATCC8293 also contained genetic potential for xylose and β-glucoside metabolism. The lineage M4 strains LhT16 and LN05 also contained the deletion in the \textit{lacZ} gene which is commonly identified in \textit{Ln. cremoris} type strains. A genetic potential for proteolysis of casein (\textit{prlP}) was identified in \textit{Ln. mesenteroides} lineages M1 and M4, but not in M2 or M3.

**(III) \textit{Ln. lactis} Lineages**

The pan-genomic differentiation grouped all the \textit{Ln. lactis} isolates into one lineage. However, differences in genetic potential were found between the kimchi isolate \textit{Ln. lactis} 91922 and...
dairy isolates LN19 and LN24. *L. lactis* 91922 lacked citrate metabolism genes *citCDEFG*, but carried genetic potential for a maltose and glucose specific PTS system, metabolism of arabinose and a CRISPR-Cas operon, that were not found in the other two *L. lactis* isolates.

### (IV) *L. pseudomesenteroides* Lineages

Despite the significant pan-genomic differences and the sequence variation in shared OGs, the functional differences between lineages of *L. pseudomesenteroides* were surprisingly few. *L. pseudomesenteroides* P4 was different from the other three lineages with regards to genome synteny and genetic potential. Functional functionality in the category of methionine biosynthesis, β-glucoside metabolism, sucrose metabolism, as well as an additional lactate dehydrogenase was identified in *L. pseudomesenteroides* P4 but not P1, P2, and P3. Moreover, P4 isolates were missing the genes for reduction of diacetyl to acetoin and 2,3-butandiol, and contained genes for a different capsular and extracellular polysaccharide biosynthesis pathway, compared to P1, P2, and P3 isolates.

### Genetic Potential of *Leuconostoc*

#### (I) Amino Acid Biosynthesis

The amino acid requirements of leuconostocs have been described as highly variable between strains. Glutamic acid and valine are required by most leuconostocs, methionine usually stimulates growth, while no *Leuconostoc* are reported to require alanine (Garvie, 1967). Comparative analysis of genes involved in amino acid biosynthesis showed that *L. cremoris* and *L. mesenteroides* subspecies carried the genetic potential to produce a wide range of amino acids while *L. lactis* and *L. pseudomesenteroides* did not (*Table 2*). This included genes encoding biosynthesis of histidine, tryptophan, methionine and lysine. Studies on the amino acid requirement of leuconostocs show that most of the *L. mesenteroides* subspecies do require isoleucine and leucine to grow. The *ilv* and *leu* operons involved in biosynthesis of the branched-chain amino acids isoleucine, leucine and valine were present in all *L. mesenteroides* isolates, however both operons were truncated when compared to functional *ilv* and *leu* operons from lactococci. The *leuA* gene in the *leuABCD* operon is truncated in *leuconostocs* (391 aa) compared to lactococci (513 aa) likely resulting in an inactive product and a nonfunctional pathway. This has been documented in the dairy strain *Lactococcus lactis* IL1403 where a similar truncation of the *leuA* gene led to an inactivation of the leucine/valine pathway (Godon et al., 1993). Likewise, the *ilv* operon of sequenced leuconostocs is missing the *ilvD* gene, and has truncated *ilvA* and *ilvH* genes when compared to the lactococcal *ilv* operon. The truncation of *ilvA* has been shown to result in inactivation of the product, and would by itself be sufficient to abort the biosynthesis pathway (Cavin et al., 1999). None of the leuconostocs had genes for biosynthesis of glutamic acid. *L. lactis* isolates also lacked the genetic potential for cysteine biosynthesis.

#### (II) Carbohydrate Metabolism

Differences in the genetic potential within and between the *Leuconostoc* species were analyzed by comparing intra-species
pan-genomes using Blast2GO and the Seed Viewer. The *Leuconostoc* genus is composed of heterofermentative bacteria that use the phosphoketolase pathway to ferment hexoses. Therefore, it was not surprising to find that none of the isolates contained the gene for phosphofructokinase, a key enzyme in the Embden-Meyerhof pathway. However, a gene encoding fructose-bisphosphate aldolase class II was present in all isolates binned into pan-genome lineages. Information on each individual isolate is included in Supplementary Table S1.

### TABLE 1 | Average genome size and coding sequences of *Leuconostoc* isolates binned into pan-genome lineages.

<table>
<thead>
<tr>
<th>Profile name</th>
<th>Average genome size (Mb)</th>
<th>Average CDS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ln. cremoris C1 (MPCA-type)</td>
<td>1.680 (±5)</td>
<td>1760 (±20)</td>
</tr>
<tr>
<td>Ln. cremoris C2 (MRS-type)</td>
<td>1.741 (±40)</td>
<td>1822 (±30)</td>
</tr>
<tr>
<td>Ln. cremoris C3</td>
<td>1.765 (±124)</td>
<td>1956 (±198)</td>
</tr>
<tr>
<td>Ln. mesenteroides M1</td>
<td>1.869 (±19)</td>
<td>1851 (±7)</td>
</tr>
<tr>
<td>Ln. mesenteroides M2</td>
<td>2.150 (±123)</td>
<td>2212 (±162)</td>
</tr>
<tr>
<td>Ln. mesenteroides M3</td>
<td>2.014 (±19)</td>
<td>2074 (±18)</td>
</tr>
<tr>
<td>Ln. mesenteroides M4</td>
<td>2.061 (±219)</td>
<td>2101 (±179)</td>
</tr>
<tr>
<td>Ln. pseudomesenteroides P1</td>
<td>2.028 (±47)</td>
<td>2081 (±61)</td>
</tr>
<tr>
<td>Ln. pseudomesenteroides P2</td>
<td>1.921 (±25)</td>
<td>1925 (±46)</td>
</tr>
<tr>
<td>Ln. pseudomesenteroides P3</td>
<td>2.063 (±44)</td>
<td>2133 (±60)</td>
</tr>
<tr>
<td>Ln. pseudomesenteroides P4</td>
<td>2.032 (±61)</td>
<td>2046 (±60)</td>
</tr>
<tr>
<td>Ln. lactis L1</td>
<td>1.718 (±26)</td>
<td>1700 (±43)</td>
</tr>
</tbody>
</table>

Information on each individual isolate is included in Supplementary Table S1.

pan-genomes using Blast2GO and the Seed Viewer. The *Leuconostoc* genus is composed of heterofermentative bacteria that use the phosphoketolase pathway to ferment hexoses. Therefore, it was not surprising to find that none of the isolates contained the gene for phosphofructokinase, a key enzyme in the Embden-Meyerhof pathway. However, a gene encoding fructose-bisphosphate aldolase class II was present in all isolates binned into pan-genome lineages. Information on each individual isolate is included in Supplementary Table S1.

### TABLE 2 | Presence of genes encoding enzymes for amino acid biosynthesis.

<table>
<thead>
<tr>
<th>Amino acid pathway</th>
<th>Ln. cremoris</th>
<th>Ln. mesenteroides</th>
<th>Ln. lactis</th>
<th>Ln. pseudomesenteroides</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alanine</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Arginine</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Aspartate</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Cysteine</td>
<td></td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Glutamine</td>
<td></td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Glutamic acid</td>
<td></td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Glycine</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Histidine</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Isoleucine</td>
<td></td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Leucine</td>
<td></td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Lysine</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Methionine</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Phenylalanine</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Proline</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Serine</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Threonine</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Tryptophan</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Tyrosine</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Valine</td>
<td></td>
<td></td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

+, presence of predicted pathway functionality; −, absence of predicted pathway functionality.

*Ln. cremoris* (<36% identity). Genes coding for maltose-phosphorylase (malP) and sucrose-6-phosphate hydrolase (scrB) were found in *Ln. lactis*, *Ln. pseudomesenteroides* P4, and *Ln. mesenteroides*, but not *Ln. cremoris*. These enzymes are central to the metabolism of maltose and sucrose. Isolates containing *malP* also contained genes *malR* and *malL*, as well as a maltose epimerase. *Ln. lactis* and *Ln. pseudomesenteroides* also contained the malEFG gene cluster encoding for an ABC transporter, however the malEFG genes were truncated in *Ln. pseudomesenteroides*. Genes encoding for β-glucosidase (*bglA*) enabling utilization of salicin and arbutin was found in all *Ln. pseudomesenteroides* and *Ln. lactis* isolates, as well as in *Ln. mesenteroides* M2 isolates. The *bglA* gene, was found to be present in all *Ln. cremoris* isolates, as well as *Ln. mesenteroides* M1, M3, and M4 isolates, however the gene was truncated and was identified as inactive by the Seed Viewer. A genetic potential for metabolism of trehalose was found, annotated as treA in *Ln. mesenteroides* and the *Ln. lactis* of dairy origin, and as TrePP in *Ln. mesenteroides* and *Ln. lactis* 91922. Genes encoding for trehalose transport were not found in *Ln. mesenteroides* M3 and M4, indicating that these lineages are not able to metabolize trehalose from the environment. Xylose isomerase (*xylA*) and xylose kinase (*xylB*) genes were found in all *Leuconostoc* isolates, but the genes were heavily truncated in *Ln. cremoris* isolates and *Ln. mesenteroides* M3 and M4 isolates. Isolates with full length *xylA* and *xylB* genes also contained the gene *xylG*, encoding for a xylose transport protein.
TABLE 3 | Genetic potential for metabolism of carbohydrates indicated by the presence or absence of enzymes crucial to metabolism of substrates.

<table>
<thead>
<tr>
<th>Gene(s)</th>
<th>Ln. cremoris</th>
<th>Ln. mesenteroides</th>
<th>Ln. pseudomesenteroides</th>
<th>Ln. lactis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>C1 (n = 13)</td>
<td>C2 (n = 5)</td>
<td>C3 (n = 2)</td>
<td>M1 (n = 3)</td>
</tr>
<tr>
<td>araBAD</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>maIP</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>maEFG</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>maX</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>maL</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>maR</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>lacL</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>lacM</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>lacZ</td>
<td>#</td>
<td>#</td>
<td>#</td>
<td>#</td>
</tr>
<tr>
<td>lacS</td>
<td>#</td>
<td>#</td>
<td>#</td>
<td>+</td>
</tr>
<tr>
<td>galEKT</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+ (75%)</td>
</tr>
<tr>
<td>manXYZ</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>manA</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>scrB</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>+</td>
</tr>
<tr>
<td>xyABG</td>
<td>#</td>
<td>#</td>
<td>#</td>
<td>#</td>
</tr>
<tr>
<td>treA</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>+</td>
</tr>
<tr>
<td>trePP</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>bgA</td>
<td>#</td>
<td>#</td>
<td>#</td>
<td>#</td>
</tr>
<tr>
<td>fruA</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>levE</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>frk</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>citCDEFGOS</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>fba</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>

+, gene presence. –, gene absence; #, gene present but truncated. Number in parenthesis signifies percentage of isolates where gene was present. All the isolates were able to metabolize glucose and lactose. The number given in parenthesis is given for the percentage of isolates within the lineage with the gene. Genes are abbreviated as follows: araBAD, arabinose metabolism pathway; maIP, maltose phosphorylase; maEFG, maltose transport genes; maX, maltose/maltodextrin binding precursor; manL, sucrose-isomaltose; maR, maltose operon regulatory gene; lacL, beta-galactosidase, big subunit; lacM, beta-galactosidase, small subunit; lacS, lactose permease; galEKT, galactose metabolism; manXYZ, mannose transport genes; manA, mannose-6-phosphate isomerase; scrB, sucrose-6-phosphate hydrolase; xyABG, xylose isomerase, xylose kinase, xylose transport protein; treA, trehalose-6-phosphate hydrolase; trePP, trehalase-6-phosphate phosphorylase; bgA, beta-D-glucosidase; fruA and levE, fructose PTS; frk, fructokinase; citCDEFGOS, citrate metabolism operon; fba, fructose bisphosphate aldolase.

(III) Citrate Metabolism

All the dairy strains in this study contained the genes necessary for uptake and metabolism of citrate. These genes are found in an operon comprised of citC (citrate lyase ligase), citDEF (citrate lyase), citG (holo-ACP synthase), citO (transcriptional regulator) and citS (Na+ dependent citrate transporter). A citrate/malate transporter annotated cinH was present in Ln. mesenteroides subspecies isolates, but was not present in any of the Ln. lactis or Ln. pseudomesenteroides isolates. In the Ln. cremoris and Ln. pseudomesenteroides genomes, the cit operon is flanked by two IS116/IS110/IS902 family transposases, suggesting it may have been acquired by horizontal gene transfer. In these bacteria, the operon appears to be located on the chromosome, a finding supported by the genome assembly, which organizes the cit operon on a contig containing a number of essential genes, and by read coverage analysis that shows a continuous gapless coverage through the contig, with no elevation in read coverage across the cit operon. The citCDEFGOS operons of Ln. mesenteroides and Ln. lactis, however, appear to be located on a plasmid, since in all cases they assembled on a contig, which includes a site of replication and not essential genes. The cit operon is highly conserved in the Ln. cremoris and Ln. pseudomesenteroides genomes with >97% DNA sequence identity between all the isolates. The likely to be plasmid-encoded cit operon found in Ln. mesenteroides and Ln. lactis genomes is also highly conserved between the isolates (>99% identity), however it is significantly different from the chromosomally encoded cit operon present in Ln. cremoris and Ln. pseudomesenteroides (50–65% DNA sequence identity for each gene). None of the strains of nondairy origin included in this study contained the citrate genes, indicating that the ability to metabolize citrate plays an important role in the successful adaption to the milk environment.

(IV) Proteolytic Activity

Leuconostocs grow in association with the lactococci in dairy fermentations, and commonly grow poorly in milk without the presence of lactococci. The general explanation for this poor growth is their lack of proteinase activity, making them dependent on small peptides from lactococcal proteinase activity. Screening all the isolates for genes involved in peptide and proteolytic activity revealed a number of differences between the lineages (Table 4). The genes encoding for the OppABCDF
system were found in all Leuconostoc genomes. However, in *Ln. cremoris* genomes, the oppA gene was missing, and the oppB gene was severely truncated. A gene encoding for a PII-type serine proteinase (PrtP) best known for its action on caseins was found in all *Ln. pseudomesenteroides* genomes, dairy *Ln. lactis* genomes, *Ln. mesenteroides* M4 and 33% of *Ln. mesenteroides* M1 genomes. All the sequenced Leuconostoc strains coded for a range of peptidases and aminotransferases. The *Ln. cremoris* isolates did not contain the pepN gene, but had the other general aminopeptidase gene, pepC, which was found to be missing from *Ln. lactis* genomes. The pepX gene, encoding for the enzyme x-prolyl dipetidyl aminopeptidase, was truncated in *Ln. cremoris* (334 amino acids) compared to the pepX of other Leuconostoc strains (778–779 amino acids). The pepA, pepF, pepO, pepQ, pepS, and pepT genes were present in all Leuconostoc isolates. Finally, all *Ln. pseudomesenteroides* have the pepV gene, encoding β-alal-dipeptidase. This dipetidase has been shown to cleave dipetides with an N-terminal β-Ala or D-alal residue, such as carnosine and to a lesser extent, was shown to catalyze removal of N-terminal amino acids from a few distinct tripeptides in *Lactobacillus delbrueckii* subsp. *lactis* (Vongerichten et al., 1994).

### CRISPR-Cas in *Ln. lactis* and *Ln. pseudomesenteroides*

*Ln. lactis* N91922 and all the *Ln. pseudomesenteroides* isolates included in this study contained CRISPR-Cas genes with repeat regions.

### Composition of Leuconostocs in Starter Cultures

The Leuconostoc core gene library was used to devise a scheme for species and subspecies quantification in starter cultures by amplicon sequencing. Core genes were screened for sequence variation and for targeted-amplicon suitability. After curation, the top three candidates were 16S rRNA, rpoB, and eno. While the full-length 16S rRNA sequence enables differentiation of species and subspecies, any region shorter than 500 bp is only able to differentiate between species, and then only when using the nucleotides between position 150–550, encompassing the V2 and V3 regions of 16S rRNA. However, the sequences of 16S rRNA and the rpoB loci were too similar to the same genes in lactococci to allow for primer design specific for leuconostocs, and thus were unsuitable for quantification of leuconostocs. The gene encoding enolase (eno) did allow for *Leuconostoc* specific primer design, and was used in targeted-amplicon sequencing to analyze the diversity of leuconostocs in the five starter cultures. The analysis revealed great differences between the starter cultures (Figure 4). *Ln. cremoris* dominated the Leuconostoc populations in cultures A, C, and E, *Ln. pseudomesenteroides* was most abundant in cultures B and C. Most of the *Ln. cremoris* in cultures A and D were of the MPCA type (*Ln. cremoris* Cl) unable to grow on MRS, while MRS type *Ln. cremoris* dominated in culture E (data not shown). Relatively low levels of *Ln. mesenteroides* and *Ln. dextranicum* were found in all cultures, the highest being 14% in culture B. *Ln. lactis* was only found in one of the starter cultures, culture E, where it constituted 17% of the leuconostocs.

### DISCUSSION

Decades have passed since Dr. Ellen Garvie laid the foundation for the taxonomy of dairy relevant leuconostocs, and Dr. John Farrow expanded this list to include *Ln. pseudomesenteroides*. Their work has been the basis for classification of leuconostocs since then.

The *Ln. pseudomesenteroides* species was described for the first time in 1898 (Farrow et al., 1989), however its presence in a dairy starter culture was not described before 2014 (Pedersen et al., 2014b). Identification of leuconostocs by phenotypical traits or by partial 16S rRNA sequencing does not reliably distinguish between all species and misidentification has been

### TABLE 4 | Genetic potential for proteolytic activity.

<table>
<thead>
<tr>
<th>Gene(s)</th>
<th><em>Ln. cremoris</em></th>
<th><em>Ln. mesenteroides</em></th>
<th><em>Ln. pseudomesenteroides</em></th>
<th><em>Ln. lactis</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>C1 (n=13)</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>C2 (n=5)</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>C3 (n=2)</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>M1 (n=3)</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>M2 (n=4)</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>M3 (n=3)</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>M4 (n=2)</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>P1 (n=6)</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>P2 (n=4)</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>P3 (n=5)</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>P4 (n=8)</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>L1 (n=3)</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

+, gene presence; −, gene absence; #, gene(s) present but truncated. Number in parenthesis indicates percentage of isolates where gene was present. Genes are abbreviated as follows: prtP, type-II serine proteinase; pepA, glutamyl aminopeptidase; pepC, aminopeptidase C; pepF, oligoendopeptidase; pepN, neutral endopeptidase; pepO, neutral endopeptidase; pepQ, β-alal-xaa dipetidase; pepV, xaa-pro dipetidyl-peptidase; oppABCDF, peptide ABC transporter operon.
common. After genomic analysis, several isolates previously identified as *Ln. mesenteroides* subspecies proved to be *Ln. pseudomesenteroides* and isolates may have been misidentified in other studies as well. Surprisingly, the strain LbT16 (Accession. No: LAYV00000000) reported to be *Ln. cremoris* by Campedelli et al. (2015) was identified as *Ln. mesenteroides* when characterized by its genomic content and its full length 16S rRNA sequence. Misidentification of *Ln. cremoris* is also uncommon. Compared to other dairy leuconostocs, *Ln. cremoris* grow slower, to a lower density and not at temperatures of 30°C or higher. In addition, a large proportion of *Ln. cremoris* type strains are not able to grow on MRS. These characteristics provide the means for reliable phenotypical identification of *Ln. cremoris*. However, phenotypical differentiation between other *Ln. mesenteroides* subspecies, *Ln. lactis* and *Ln. pseudomesenteroides* remains unreliable. In this study, dairy relevant leuconostocs are characterized using a genomics approach and the diversity of leuconostocs in five commercial DL-type starter cultures is analyzed.

The genomic analysis clearly separated leuconostocs by species, subspecies, and enabled intra-species differentiation. Interestingly, the genomic analysis did not distinguish *Ln. dextranicum* from *Ln. mesenteroides*. The strain-to-strain variation was higher than the differences between subspecies. The *dextranicum* subspecies has been previously defined by phenotypical traits only and separate subspecies distinction is not justified by the genomic data of this study. On the other hand, the pan-genomic analysis separated *Ln. mesenteroides* isolates by habitat. The dairy strains clearly differ from those isolated from plant material, the former have smaller genomes and utilize a more restricted range of carbohydrates. The two subspecies *Ln. mesenteroides* and *Ln. cremoris* share a large amount of genetic content with high identity scores, reflecting a close phylogenetic relationship. However, many genes present in *Ln. mesenteroides* are found to be truncated, contain deletions or are completely missing in *Ln. cremoris*. Adaptation of dairy strains to the milk environment involved acquisition of the plasmid-encoded *lacLM* by horizontal gene transfer (Obst et al., 1995), which in turn permitted loss of a functional *lacZ*. Some of the dairy *Ln. mesenteroides*, and all of the *Ln. cremoris* isolates carry a deletion in the *lacZ* gene. The dairy *Ln. mesenteroides* and in particular *Ln. cremoris* display telltale signs of a prolonged degenerative evolution, likely the result of a long period of growth in milk. In this environment, the leuconostocs have evolved alongside lactococci. All the dairy strains included in this study contain the *cit* operon comprised of *citC* (citrate lyase ligase), *citDEF* (citrate lyase), *citG* (holo-ACP synthase), *citO* (transcriptional regulator) and *citS* (Na⁺ dependent citrate transporter). The *citCDEFGOS* operon organization is different from the operon in *Lactococcus lactis*, which lacks *citO* and the *citS* transporter (Drider et al., 2004). In citrate positive *Lactococcus lactis*, homologs of *citO* (*citR*) and the *citS* (*citP*) are located on a plasmid (Magni et al., 1994). The presence of the *citCDEFGOS* genes enable so-called citrolactic fermentation, co-metabolism of sugar and citrate providing the cells with higher energy yield and proton motive force (Marty-Teyssset et al., 1996). In *Ln. lactis* and *Ln. mesenteroides*, this operon has been linked to a ~22-kb plasmid, inferred by phenotypical studies in combination with monitoring the presence of mobile genetic elements (Lin et al., 1991; Vaughan et al., 1995). In the study by Vaughan et al. (1995), *Ln. mesenteroides* was shown to retain its ability to metabolize citrate after losing three of its four plasmids. Moreover, after curing, a derivative isolate without the ability to degrade citrate still contained the fourth plasmid. Our data indicates that for *Ln. cremoris* and *Ln. pseudomesenteroides*, this is not the case. In all the *Ln. cremoris* and *Ln. pseudomesenteroides* genomes included in this study, the *cit* operon is located on the chromosome in a region with mobile element characteristics. A low level of genetic drift is indicated by the high sequence similarity between the *cit* operons of *Ln. cremoris* and *Ln. pseudomesenteroides* suggesting that the acquisition of these genes is quite recent.
possibly from a common donor. The chromosomally encoded cit operon of \textit{Ln. cremoris} and \textit{Ln. pseudomesenteroides} was significantly different from the highly conserved and likely to be plasmid-encoded cit operon found in \textit{Ln. lactis} and \textit{Ln. mesenteroides}. These results indicate that the plasmid encoded cit operon originates from a different source and time. None of the strains of non-dairy origin included in this study contained the citrate metabolism genes, indicating that the ability to metabolize citrate also plays an important role in the successful adaption to the milk environment. The manufacture of Dutch-type cheeses has been going on for centuries and the starter cultures have been maintained by so-called “back slopping” for the last one and a half century, where new milk is inoculated with whey from the previous batch. This technique for propagating starter cultures is still being used and recent studies have shown that the complex starter cultures maintain a highly stable composition with regards to lactococci (Erkus et al., 2013). Culture composition may change over a short period of time depending on growth conditions and bacteriophage predation, but the microbial community is sustained in the long run. In this study, we show a large variation in the amount and composition of the \textit{Leuconostoc} populations in cheeses starter cultures. Three of the starter cultures (A, D, and E) were dominated by \textit{Ln. cremoris}, and for culture A and D, the majority of these were unable to grow on MRS. The other two starter cultures (B and C) were dominated by \textit{Ln. pseudomesenteroides}. Interestingly, the cultures dominated by \textit{Ln. cremoris} also contain \textit{Ln. pseudomesenteroides} strains. \textit{Ln. pseudomesenteroides} growth rates in pure culture are significantly higher than that of \textit{Ln. cremoris} at temperatures above 20°C, so the microbial community is preserved, either by the starter culture developers, or by the microbial community itself. Little knowledge exists on how the diversity of leuconostocs is affected by manufacturing procedures. According to Thunell (1995) and Vedamuthu (1994) the only leuconostocs relevant in dairy are \textit{Ln. cremoris} and \textit{Ln. lactis}, but in this study, \textit{Ln. lactis} was detected only in culture E, which was dominated by \textit{Ln. cremoris}. In two of the starter cultures studies in this work, \textit{Ln. pseudomesenteroides} was the dominating \textit{Leuconostoc}, which shows that they are highly relevant in the production of cheese. This is also reflected by recent studies, where the presence of \textit{Ln. pseudomesenteroides} is more frequently reported (Callon et al., 2004; Porcellato and Skeie, 2016; Østlie et al., 2016). It is tempting to speculate that starter culture manufacturers have altered the conditions for culture propagation or manipulated the strain collections, thereby altering the culture dynamics between strains in favor of \textit{Ln. pseudomesenteroides}.

The differences between the starter cultures could have an impact on the characteristics of the cheese product. \textit{Ln. cremoris} lacks a wide range of genes involved in carbohydrate metabolism and proteolytic activity, and studies have shown that \textit{Ln. cremoris} and \textit{Ln. pseudomesenteroides} differ significantly in their ability to produce a wide range of volatile compounds (Pedersen et al., 2016). Most notably, the amount of acetoin and diacetyl in model-cheeses produced with only \textit{Ln. pseudomesenteroides} was negligible. This was supported by our data, which showed that the \textit{Ln. pseudomesenteroides} P4 isolates lack the genes necessary for reduction of diacetyl to acetoin and 2,3-butadiol. In addition, these isolates lacked the genes \textit{ilvB} and \textit{ilvH} encoding acetolactate synthetase large and small subunits, which is found in all \textit{Ln. mesenteroides} subspecies. However, a different gene \textit{alsS}, encoding the same function, was found in all leuconostocs, including \textit{Ln. pseudomesenteroides}. Studies on \textit{α}-acetolactate synthase (ALS) and \textit{α}-acetolactate decarboxylase (ALDC) activity in \textit{Ln. mesenteroides} subspecies and \textit{Ln. lactis} showed that the activity of both ALS and ALDC was higher for \textit{Ln. lactis} (which does not have the \textit{ilv} or \textit{leu} operon) than that of \textit{Ln. cremoris} (which does have part of these two operons) (Monnet et al., 1994). For comparison, the ALS activity of \textit{Lc. lactis} biovar diacetylactis was comparable or in some cases even higher than that of \textit{Ln. lactis}. \textit{Ln. pseudomesenteroides} was not included in the study, but data from semi-hard cheeses comparing the acetoin and diacetyl concentrations revealed lower concentrations in mock starters containing \textit{Ln. pseudomesenteroides} compared to mock starters containing \textit{Ln. cremoris} (Pedersen et al., 2016). This observation could be attributed to the rapid growth rate of \textit{Ln. pseudomesenteroides} when compared to that of \textit{Ln. cremoris}. The presence of the degenerated \textit{ilv} and \textit{leu} operons could somehow be negative to \textit{Ln. cremoris} growth rate. Indeed, when cloning of the \textit{ilv} operon into \textit{Escherichia coli}, the presence of \textit{Leuconostoc ilvB} was strongly detrimental to growth, while recombiant strains with an insertion in the \textit{Leuconostoc ilvB} genes displayed normal growth. Their hypothesis was that expression of \textit{ilvB} without a functional branched chain amino acid biosynthesis mechanism could interfere with energy metabolism via pyruvate (Cavin et al., 1999).

In dairy fermentations, the leuconostocs grow in association with the lactococci. Whether the associative growth is of mutual benefit to the leuconostocs and lactococci has not been determined. Literature often attributes the poor growth of leuconostocs to the lack of protease activity (Vedamuthu, 1994; Thunell, 1995). However, the ability to acidify milk in pure culture has been described for \textit{Ln. pseudomesenteroides} (Cardamone et al., 2011), and we identified genetic potential for caseinolytic activity in \textit{Ln. pseudomesenteroides} in our data. This would enable \textit{Ln. pseudomesenteroides} to grow better in milk than \textit{Ln. cremoris}, which lacks the capacity for protease, as well as a functional peptide uptake system due to the lack of OppA, which is responsible for the uptake of extracellular peptides. An argument for mutually beneficial growth has been made by superimposing metabolic pathways from lactococci and leuconostocs, indicating a potential for metabolic complementation between the two genera (Erkus et al., 2013). One can be forgiven for thinking \textit{Ln. pseudomesenteroides} the better bacteria of the two based on these tidbits of information alone. However, both \textit{Ln. cremoris} and \textit{Ln. pseudomesenteroides} have shown to be significant to the production of cheeses. It is difficult to conclude which \textit{Leuconostoc} species produces the highly subjective matter of the better cheese product. The concentration of volatile compounds, fatty acid derivatives, acetoin, diacetyl, and amino acid derivatives in products have been shown to diverge significantly, depending on which \textit{Leuconostoc} species is added to the mixture of lactococci (Pedersen et al., 2016).
In conclusion, the dairy-associated leuconostocs are highly adapted to grow in milk. Comparative genomic analysis reveals great differences between the *Leuconostoc* species and subspecies accustomed to the dairy environment, where they grow in association with the lactococci. The composition of the *Leuconostoc* population is significantly different between commercial starter cultures, which ultimately affects the characteristics and quality of the product. A better understanding of *Leuconostoc* microbial dynamics and the functional role of different dairy leuconostocs could be of great importance and be an applicable tool in ensuring consistent manufacture of high quality product. Currently, no detailed information on the relative amount or diversity of the *Leuconostoc* population in starter cultures is available to the industry. We provide a culture independent method for robust identification and quantification of *Leuconostoc* species in mixed microbial communities, enabling quantification of leuconostocs in starter cultures, as well as monitoring the diversity of leuconostocs through the cheese production process.

**AUTHOR CONTRIBUTIONS**

CF isolated and sequenced bacterial strains, performed the sequencing work in Norway (of all CF and H-isolates in CF isolated and sequenced bacterial strains, performed the production process.

**REFERENCES**


**FUNDING**

This work was funded by the Norwegian Research Council, TINE SA, and the Danish Council for Independent Research.

**ACKNOWLEDGMENTS**

We are grateful to TINE SA for providing culture samples and Dorota Dynda for providing isolates from Twarog. This work was funded by the Norwegian Research Council, TINE SA, and the Danish Council for Independent Research.

**SUPPLEMENTARY MATERIAL**

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fmicb.2017.00132/full#supplementary-material
bacteria in low-fat and full-fat Dutch-type cheese. *Int. Dairy J.* 33, 104–111. doi: 10.1016/j.dairyj.2013.01.007


Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.
<table>
<thead>
<tr>
<th>Isolate name</th>
<th>Accession No.</th>
<th>Isolation source</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>19254</td>
<td>NZ_ACKV000000000000</td>
<td>Homo sapiens</td>
<td>NCBI Genomes</td>
</tr>
<tr>
<td>T26</td>
<td>NZ_JAU000000000000</td>
<td>dairy starter culture</td>
<td>Pedersen, T. B. (2016)</td>
</tr>
<tr>
<td>LbT16</td>
<td>NZ_LAYV000000000000</td>
<td>Taleggio cheese</td>
<td>Campedelli, I. (2015)</td>
</tr>
<tr>
<td>TIFN8</td>
<td>NZ_ATAZ000000000000</td>
<td>dairy starter culture</td>
<td>Erkus, O. (2013)</td>
</tr>
<tr>
<td>20484</td>
<td>GCA_001047695</td>
<td>Cheese</td>
<td>Park, G. (2015)</td>
</tr>
<tr>
<td>J18</td>
<td>GCA_000234825</td>
<td>Kimchi</td>
<td>Jung (2012)</td>
</tr>
<tr>
<td>KACC 91922</td>
<td>NZ_JMEA000000000000</td>
<td>Kimchi</td>
<td>NCBI Genomes</td>
</tr>
<tr>
<td>1159</td>
<td>NZ_JAU000000000000</td>
<td>Dairy starter culture</td>
<td>Pedersen, T. B. (2014)</td>
</tr>
<tr>
<td>PS12</td>
<td>NZ_JDVA000000000000</td>
<td>Dairy starter culture</td>
<td>Pedersen, T. B. (2014)</td>
</tr>
<tr>
<td>4882</td>
<td>NZ_CAKV000000000000</td>
<td>Dairy starter culture</td>
<td>Meslier, V. (2012)</td>
</tr>
<tr>
<td>LN02</td>
<td>MPLG000000000000</td>
<td>Dairy starter culture</td>
<td>This study(c)</td>
</tr>
<tr>
<td>LN05</td>
<td>MPLH000000000000</td>
<td>Dairy starter culture</td>
<td>This study(c)</td>
</tr>
<tr>
<td>LN07</td>
<td>MPLI000000000000</td>
<td>Dairy starter culture</td>
<td>This study(c)</td>
</tr>
<tr>
<td>LN08</td>
<td>MPLJ000000000000</td>
<td>Culture collection</td>
<td>This study(c)</td>
</tr>
<tr>
<td>LN12</td>
<td>MPLK000000000000</td>
<td>Culture collection</td>
<td>This study(c)</td>
</tr>
<tr>
<td>LN19</td>
<td>MPLL000000000000</td>
<td>Whey isolate</td>
<td>This study(c)</td>
</tr>
<tr>
<td>LN23</td>
<td>MPLM000000000000</td>
<td>Dairy starter culture</td>
<td>This study(c)</td>
</tr>
<tr>
<td>LN24</td>
<td>MPLN000000000000</td>
<td>Culture collection</td>
<td>This study(c)</td>
</tr>
<tr>
<td>LN25</td>
<td>MPLO000000000000</td>
<td>Cheese isolate</td>
<td>This study(c)</td>
</tr>
<tr>
<td>LN27</td>
<td>MPLP000000000000</td>
<td>Culture collection</td>
<td>This study(c)</td>
</tr>
<tr>
<td>LN32</td>
<td>MPLQ000000000000</td>
<td>Dairy starter culture</td>
<td>This study(c)</td>
</tr>
<tr>
<td>LN34</td>
<td>MPLR000000000000</td>
<td>Dairy starter culture</td>
<td>This study(c)</td>
</tr>
<tr>
<td>TW1</td>
<td>MPBC000000000000</td>
<td>Twarog</td>
<td>This study</td>
</tr>
<tr>
<td>TW3</td>
<td>MPBD000000000000</td>
<td>Twarog</td>
<td>This study</td>
</tr>
<tr>
<td>TW6</td>
<td>MPBA000000000000</td>
<td>Twarog</td>
<td>This study</td>
</tr>
<tr>
<td>TW8</td>
<td>MPBB000000000000</td>
<td>Twarog</td>
<td>This study</td>
</tr>
<tr>
<td>CF01</td>
<td>MPA000000000000</td>
<td>Dairy starter culture A</td>
<td>This study</td>
</tr>
<tr>
<td>CF02</td>
<td>MPAN000000000000</td>
<td>Dairy starter culture A</td>
<td>This study</td>
</tr>
<tr>
<td>CF03</td>
<td>MAPA000000000000</td>
<td>Dairy starter culture A</td>
<td>This study</td>
</tr>
<tr>
<td>CF04</td>
<td>MAPA000000000000</td>
<td>Dairy starter culture A</td>
<td>This study</td>
</tr>
<tr>
<td>CF05</td>
<td>MAPA000000000000</td>
<td>Dairy starter culture A</td>
<td>This study</td>
</tr>
<tr>
<td>CF06</td>
<td>MPA000000000000</td>
<td>Dairy starter culture A</td>
<td>This study</td>
</tr>
<tr>
<td>CF07</td>
<td>MPA000000000000</td>
<td>Dairy starter culture A</td>
<td>This study</td>
</tr>
<tr>
<td>CF08</td>
<td>MPLS000000000000</td>
<td>Dairy starter culture A</td>
<td>This study</td>
</tr>
<tr>
<td>CF09</td>
<td>MAPA000000000000</td>
<td>Dairy starter culture A</td>
<td>This study</td>
</tr>
<tr>
<td>CF10</td>
<td>MA000000000000</td>
<td>Dairy starter culture A</td>
<td>This study</td>
</tr>
<tr>
<td>CF11</td>
<td>MPEA000000000000</td>
<td>Dairy starter culture A</td>
<td>This study</td>
</tr>
<tr>
<td>CF12</td>
<td>MPEB000000000000</td>
<td>Dairy starter culture A</td>
<td>This study</td>
</tr>
<tr>
<td>CF13</td>
<td>MPEC000000000000</td>
<td>Dairy starter culture A</td>
<td>This study</td>
</tr>
<tr>
<td>CF14</td>
<td>MPEE000000000000</td>
<td>Dairy starter culture A</td>
<td>This study</td>
</tr>
<tr>
<td>CF15</td>
<td>MPEF000000000000</td>
<td>Dairy starter culture A</td>
<td>This study</td>
</tr>
<tr>
<td>CF16</td>
<td>MPEG000000000000</td>
<td>Dairy starter culture A</td>
<td>This study</td>
</tr>
<tr>
<td>Code</td>
<td>MPEH00000000</td>
<td>Dairy starter culture A</td>
<td>This study</td>
</tr>
<tr>
<td>-------</td>
<td>--------------</td>
<td>-------------------------</td>
<td>------------</td>
</tr>
<tr>
<td>CF19</td>
<td>MPEI00000000</td>
<td>Dairy starter culture A</td>
<td>This study</td>
</tr>
<tr>
<td>CF20</td>
<td>MPEJ00000000</td>
<td>Dairy starter culture A</td>
<td>This study</td>
</tr>
<tr>
<td>HPKA1</td>
<td>MPAL00000000</td>
<td>Dairy starter culture C</td>
<td>This study</td>
</tr>
<tr>
<td>H61</td>
<td>MPHN00000000</td>
<td>Cheese</td>
<td>Østlie, M, H. (2016)</td>
</tr>
<tr>
<td>H83</td>
<td>MPH00000000</td>
<td>Cheese</td>
<td>Østlie, M, H. (2016)</td>
</tr>
<tr>
<td>H95</td>
<td>MHP00000000</td>
<td>Cheese</td>
<td>Østlie, M, H. (2016)</td>
</tr>
<tr>
<td>H97</td>
<td>MPHQ00000000</td>
<td>Cheese</td>
<td>Østlie, M, H. (2016)</td>
</tr>
<tr>
<td>H100</td>
<td>MPHR00000000</td>
<td>Cheese</td>
<td>Østlie, M, H. (2016)</td>
</tr>
<tr>
<td>H278</td>
<td>MPHS00000000</td>
<td>Cheese</td>
<td>Østlie, M, H. (2016)</td>
</tr>
<tr>
<td>H280</td>
<td>MPHT00000000</td>
<td>Cheese</td>
<td>Østlie, M, H. (2016)</td>
</tr>
<tr>
<td>H284</td>
<td>MPHU00000000</td>
<td>Cheese</td>
<td>Østlie, M, H. (2016)</td>
</tr>
<tr>
<td>BM2</td>
<td>MPLF00000000</td>
<td>Butter Milk</td>
<td>This study(d)</td>
</tr>
</tbody>
</table>

(a) Received as Leuconostoc mesenteroides subsp. cremoris
(b) Received as Leuconostoc mesenteroides subsp. dextranicum
(c) Provided by Sacco Srl, Cordorago, Italy
(d) Provided by Max-Rubner Institute, Kiel, Germany.
<table>
<thead>
<tr>
<th>Organism</th>
<th>Genome size (Mb)</th>
<th>Contigs</th>
<th>CDS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leuconostoc mesenteroides subsp. cremoris</td>
<td>1.638</td>
<td>126</td>
<td>1.720</td>
</tr>
<tr>
<td>Leuconostoc mesenteroides subsp. cremoris</td>
<td>1.832</td>
<td>123</td>
<td>1.931</td>
</tr>
<tr>
<td>Leuconostoc mesenteroides subsp. mesenteroides(a)</td>
<td>1.906</td>
<td>65</td>
<td>1.979</td>
</tr>
<tr>
<td>Leuconostoc mesenteroides subsp. cremoris</td>
<td>1.686</td>
<td>143</td>
<td>1.929</td>
</tr>
<tr>
<td>Leuconostoc mesenteroides subsp. dextranicum</td>
<td>2.007</td>
<td>63</td>
<td>2.059</td>
</tr>
<tr>
<td>Leuconostoc mesenteroides subsp. mesenteroides</td>
<td>2.036</td>
<td>85</td>
<td>2.094</td>
</tr>
<tr>
<td>Leuconostoc mesenteroides subsp. dextranicum</td>
<td>1.818</td>
<td>1</td>
<td>1.858</td>
</tr>
<tr>
<td>Leuconostoc mesenteroides subsp. mesenteroides</td>
<td>1.900</td>
<td>1</td>
<td>1.946</td>
</tr>
<tr>
<td>Leuconostoc lactis</td>
<td>1.683</td>
<td>30</td>
<td>1.626</td>
</tr>
<tr>
<td>Leuconostoc pseudomesenteroides</td>
<td>2.038</td>
<td>100</td>
<td>2.075</td>
</tr>
<tr>
<td>Leuconostoc pseudomesenteroides</td>
<td>1.934</td>
<td>91</td>
<td>1.963</td>
</tr>
<tr>
<td>Leuconostoc pseudomesenteroides</td>
<td>2.008</td>
<td>1</td>
<td>2.135</td>
</tr>
<tr>
<td>Leuconostoc pseudomesenteroides</td>
<td>1.936</td>
<td>123</td>
<td>1.929</td>
</tr>
<tr>
<td>Leuconostoc mesenteroides subsp. mesenteroides</td>
<td>2.212</td>
<td>252</td>
<td>2.094</td>
</tr>
<tr>
<td>Leuconostoc mesenteroides subsp. cremoris</td>
<td>1.636</td>
<td>215</td>
<td>1.683</td>
</tr>
<tr>
<td>Leuconostoc mesenteroides subsp. mesenteroides</td>
<td>1.998</td>
<td>166</td>
<td>2.068</td>
</tr>
<tr>
<td>Leuconostoc pseudomesenteroides</td>
<td>1.884</td>
<td>127</td>
<td>1.860</td>
</tr>
<tr>
<td>Leuconostoc lactis</td>
<td>1.737</td>
<td>70</td>
<td>1.728</td>
</tr>
<tr>
<td>Leuconostoc pseudomesenteroides</td>
<td>2.011</td>
<td>157</td>
<td>2.028</td>
</tr>
<tr>
<td>Leuconostoc lactis</td>
<td>1.730</td>
<td>69</td>
<td>1.721</td>
</tr>
<tr>
<td>Leuconostoc mesenteroides subsp. mesenteroides</td>
<td>2.277</td>
<td>68</td>
<td>1.705</td>
</tr>
<tr>
<td>Leuconostoc mesenteroides subsp. mesenteroides</td>
<td>2.231</td>
<td>69</td>
<td>2.229</td>
</tr>
<tr>
<td>Leuconostoc mesenteroides subsp. mesenteroides</td>
<td>1.860</td>
<td>94</td>
<td>1.843</td>
</tr>
<tr>
<td>Leuconostoc mesenteroides subsp. mesenteroides</td>
<td>1.891</td>
<td>151</td>
<td>1.852</td>
</tr>
<tr>
<td>Leuconostoc pseudomesenteroides</td>
<td>2.028</td>
<td>81</td>
<td>2.097</td>
</tr>
<tr>
<td>Leuconostoc pseudomesenteroides</td>
<td>1.938</td>
<td>56</td>
<td>1.969</td>
</tr>
<tr>
<td>Leuconostoc pseudomesenteroides</td>
<td>1.928</td>
<td>67</td>
<td>1.944</td>
</tr>
<tr>
<td>Leuconostoc pseudomesenteroides</td>
<td>2.054</td>
<td>56</td>
<td>2.125</td>
</tr>
<tr>
<td>Leuconostoc mesenteroides subsp. cremoris</td>
<td>1.776</td>
<td>146</td>
<td>1.867</td>
</tr>
<tr>
<td>Leuconostoc mesenteroides subsp. cremoris</td>
<td>1.785</td>
<td>80</td>
<td>1.928</td>
</tr>
<tr>
<td>Leuconostoc mesenteroides subsp. cremoris</td>
<td>1.743</td>
<td>92</td>
<td>1.856</td>
</tr>
<tr>
<td>Leuconostoc mesenteroides subsp. cremoris</td>
<td>1.702</td>
<td>88</td>
<td>1.790</td>
</tr>
<tr>
<td>Leuconostoc mesenteroides subsp. cremoris</td>
<td>1.699</td>
<td>88</td>
<td>1.809</td>
</tr>
<tr>
<td>Leuconostoc pseudomesenteroides</td>
<td>2.005</td>
<td>68</td>
<td>2.009</td>
</tr>
<tr>
<td>Leuconostoc mesenteroides subsp. cremoris</td>
<td>1.683</td>
<td>186</td>
<td>1.751</td>
</tr>
<tr>
<td>Leuconostoc pseudomesenteroides</td>
<td>2.146</td>
<td>163</td>
<td>2.148</td>
</tr>
<tr>
<td>Leuconostoc mesenteroides subsp. cremoris</td>
<td>1.675</td>
<td>81</td>
<td>1.753</td>
</tr>
<tr>
<td>Leuconostoc mesenteroides subsp. cremoris</td>
<td>1.679</td>
<td>83</td>
<td>1.762</td>
</tr>
<tr>
<td>Leuconostoc mesenteroides subsp. cremoris</td>
<td>1.678</td>
<td>89</td>
<td>1.757</td>
</tr>
<tr>
<td>Leuconostoc mesenteroides subsp. cremoris</td>
<td>1.682</td>
<td>115</td>
<td>1.76</td>
</tr>
<tr>
<td>Leuconostoc mesenteroides subsp. cremoris</td>
<td>1.669</td>
<td>223</td>
<td>1.736</td>
</tr>
<tr>
<td>Leuconostoc mesenteroides subsp. cremoris</td>
<td>1.685</td>
<td>74</td>
<td>1.773</td>
</tr>
<tr>
<td>Leuconostoc pseudomesenteroides</td>
<td>1.974</td>
<td>59</td>
<td>2.00</td>
</tr>
<tr>
<td>Leuconostoc mesenteroides subsp. cremoris</td>
<td>1.680</td>
<td>97</td>
<td>1.767</td>
</tr>
<tr>
<td>Leuconostoc mesenteroides subsp. cremoris</td>
<td>1.678</td>
<td>81</td>
<td>1.751</td>
</tr>
<tr>
<td>Species</td>
<td>Value 1</td>
<td>Value 2</td>
<td>Value 3</td>
</tr>
<tr>
<td>----------------------------------------------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
</tr>
<tr>
<td><strong>Leuconostoc mesenteroides subsp. cremoris</strong></td>
<td>1.686</td>
<td>71</td>
<td>1.775</td>
</tr>
<tr>
<td><strong>Leuconostoc mesenteroides subsp. cremoris</strong></td>
<td>1.694</td>
<td>94</td>
<td>1.787</td>
</tr>
<tr>
<td><strong>Leuconostoc mesenteroides subsp. cremoris</strong></td>
<td>1.685</td>
<td>74</td>
<td>1.773</td>
</tr>
<tr>
<td><strong>Leuconostoc pseudomesenteroides</strong></td>
<td>2.009</td>
<td>74</td>
<td>2.021</td>
</tr>
<tr>
<td><strong>Leuconostoc pseudomesenteroides.(b)</strong></td>
<td>2.009</td>
<td>62</td>
<td>2.02</td>
</tr>
<tr>
<td><strong>Leuconostoc pseudomesenteroides.(b)</strong></td>
<td>2.100</td>
<td>212</td>
<td>2.183</td>
</tr>
<tr>
<td><strong>Leuconostoc pseudomesenteroides.(b)</strong></td>
<td>2.071</td>
<td>133</td>
<td>2.121</td>
</tr>
<tr>
<td><strong>Leuconostoc pseudomesenteroides.(b)</strong></td>
<td>2.061</td>
<td>85</td>
<td>2.123</td>
</tr>
<tr>
<td><strong>Leuconostoc pseudomesenteroides.(b)</strong></td>
<td>2.115</td>
<td>256</td>
<td>2.198</td>
</tr>
<tr>
<td><strong>Leuconostoc pseudomesenteroides.(b)</strong></td>
<td>1.994</td>
<td>72</td>
<td>2.006</td>
</tr>
<tr>
<td><strong>Leuconostoc pseudomesenteroides.(b)</strong></td>
<td>1.980</td>
<td>85</td>
<td>1.991</td>
</tr>
<tr>
<td><strong>Leuconostoc pseudomesenteroides.(b)</strong></td>
<td>2.073</td>
<td>222</td>
<td>2.134</td>
</tr>
<tr>
<td><strong>Leuconostoc pseudomesenteroides</strong></td>
<td>2.054</td>
<td>149</td>
<td>2.108</td>
</tr>
</tbody>
</table>
Subsystem
Acetoin, butanediol metabolism
Gram Positive Competence
Mg/Co/Ni transporter MgtE
O-acetylhomoserine sulfhydrylase (EC 2.5.1.49)
Restriction-Modification System
Choline and Betaine Uptake and Betaine Biosynthesis
Substrate-specific component STY3230 of queuosine-regulated ECF transporter
Tryptophan synthase beta chain (EC 4.2.1.20)
Dethiobiotin synthetase (EC 6.3.3.3)
capsular polysaccharide biosynthesis protein
Cold shock, CspA family of proteins
Tryptophan synthase alpha chain (EC 4.2.1.20)
ATP synthase F0 sector subunit b (EC 3.6.3.14)
Exonuclease SbcD
DNA repair, bacterial
Conserved gene cluster possibly involved in RNA metabolism
Thioredoxin-disulfide reductase
F0F1-type ATP synthase
Endoribonuclease L-PSP
D-alanine--D-alanine ligase A (EC 6.3.2.4)
preQ1-regulated inosine-uridine nucleoside hydrolase (EC 3.2.2.1)
Excinuclease ABC subunit A paralog of unknown function
Thioredoxin-disulfide reductase
Glycerol and Glycerol-3-phosphate Uptake and Utilization
Adenosylmethionine-8-amino-7-oxononanoate aminotransferase (EC 2.6.1.62)
Biogenesis of c-type cytochromes
ComF operon protein C
ATP synthase F0 sector subunit a (EC 3.6.3.14)
Adenosine deaminase (EC 3.5.4.4)
O-succinylhomoserine sulfhydrylase (EC 2.5.1.48)
Cell Wall and Capsule
Rhamnose containing glycans
Cell Wall and Capsule
Peptidoglycan Biosynthesis
D-alanine--D-alanine ligase B (EC 6.3.2.5)
Clustering-based subsystems
Bacterial Cell Division
Cell division protein FtsX
Clustering-based subsystems
EBS-176279.3.peg.1262
Hypothetical protein
Clustering-based subsystems
EBS-176279.4.peg.12694
Endonuclease L-PSP
Cofactors, Vitamins, Prothropic Groups, Pigments
Biotin biosynthesis
Cofactors, Vitamins, Prothotropic Groups, Pigments
Biotin biosynthesis
Clustering-based subsystems
Menapnone and Phyloquinone Biosynthesis
2-heptaprenyl-1,4-naphthoquinone methyltransferase (EC 2.1.1.163)
DNA Metabolism
DNA repair, UnrABC system
Exonuclease ABC subunit A of unknown function
DNA Metabolism
DNA repair, bacterial
Exonuclease SkdO
DNA Metabolism
Gram Positive Competence
ComA operon protein C
DNA Metabolism
Restriction Modification System
Type I restriction-modification system, specificity subunit 1 (EC 1.1.21.8)
DNA Metabolism
Restriction Modification System
Type II restriction-modification system methyltransferase subunit (EC 1.1.1.72)
Membrane Transport
ABC transporter oligopeptide (TC 3.A.1.5.1)
Oligopeptide transport ATP-binding protein OppD (TC 3.A.1.5.1)
Membrane Transport
Magnesium transport
Mg/Cu/Ni transporter Mgdr
Membrane Transport
EF class transporters
Substrate-specific component STY3230 of queuosine-regulated ECF transporter
Miscellaneous
Conserved gene cluster possibly involved in RNA metabolism
Serine acetylation (EC 3.1.3.10)
Miscellaneous
Phosphoglucoisiate mutase family 2
Nucleosides and Nucleotides
Purine conversions
Adenosine deaminase (EC 3.5.4.6)
Nucleosides and Nucleotides
Purine conversions
ProU regulated nicotinamide nucleoside hydrolase (EC 3.2.2.1)
Phages, Prophages, Transposable elements, Plasmids
Phage packaging machinery
Phage terminase small subunit
Phages, Prophages, Transposable elements, Plasmids
Phage tail fiber proteins
Phage tail fiber protein
Potassium metabolism
Hypersensitive potassium uptake
Potassium uptake protein TriT
RNA Metabolism
RNA processing enzymes
RNA ligase
RNA Metabolism
Ribonuclease H
Ribonuclease H (EC 3.1.26.5)
Regulation and Cell signaling
LysR-family proteins in Escherichia coli
Chromosome initiation inhibitor
Regulation and Cell signaling
LysR-family proteins in Escherichia coli
Cys region transcription activator CysB
Respiration
FIS-I type ATP synthase
ATP synthase F0 sector subunit a (EC 1.1.1.14)
Respiration
FIS-I type ATP synthase
ATP synthase F0 sector subunit b (EC 1.1.1.14)
Respiration
Bifurcations of c-type cytochromes
Cytochrome c type cytochromes protein DcmD
Secondary Metabolism
Aavin biosynthesis
Antithiole phosphate biosynthesis (EC 2.4.1.28)
Secondary Metabolism
Aavin biosynthesis
Phosphoribosylanthranilate isomerase (EC 3.1.3.24)
Secondary Metabolism
Aavin biosynthesis
Tryptophan synthase alpha chain (EC 4.2.1.20)
Secondary Metabolism
Aavin biosynthesis
Tryptophan synthase beta chain (EC 4.2.1.20)
Stress Response
Cold shock, CspA family of proteins
Cold shock protein CspA
Stress Response
Heat shock shock gene cluster extended
Xanthomonas homologous triphosphatase pyrophosphatase
Stress Response
Choline and Betaine Uptake and Betaine Biosynthesis
L-proline-glycine betaine ABC transport system permease protein ProT (EC 3.1.2.12.1)
Sulfur Metabolism
Thioredoxin-disulfide reductase
Alky hydroperoxide reductase protein C (EC 1.1.4.1)
Sulfur Metabolism
Thioredoxin-disulfide reductase
Alky hydroperoxide reductase protein C (EC 1.1.4.1)
Nucleosides and Nucleotides

Purine Utilization

Xanthine permease

2&491, 2&493; cyclic-nucleotide 2&493; phosphodiesterase (EC 3.1.4.16)

GMP reductase (EC 1.7.1.7)

Dihydropyrimidine (EC 5.1.2.3)

Uridine kinase (EC 2.7.1.48) [C1]

Phosphoribosyl pyrophosphate (EC 2.4.2.1)

Phage head maturation protease

Phage terminase, small subunit

DNA helicase, phage-associated

Phage major tail protein

Phage tail length tape-measure protein

Kup system potassium uptake protein

Ribosomal protein-L3p-arginine acetyltransferase

LeuA-Ala-GGC

Yra-Pro-CGG

Acylamino-acid-releasing enzyme (EC 3.4.19.1)

Putative parvulin type peptidyl-prolyl isomerase, similarity with ProA foldase

Ribosomal protein-L3p-alanine acetyltransferase

26 phosphoesterase superfamily protein BCP1899

Rv12 family transcriptional regulators

Rv12 family transcriptional regulator

Rv12 family transcriptional regulator, group III

Quinone oxidoreductase (EC 1.6.5.5)

Cold shock protein CspC

Ferridoxin (EC 1.16.3.1)

Iron-binding ferritin-like antioxidant protein

Non-specific DNA-binding protein Opx

Beta-lactamase class A

Probable cadmium-transporting ATPase (EC 3.6.3.3)

Copper chaperone

ABC transporter membrane-spanning permease, Pgp export, Vex3

Phages, Prophages, Transposable Elements

Phage packaging machinery

DNA helicase, phage-associated

Phage major tail protein

Phage tail length tape-measure protein

Kup system potassium uptake protein

Ribosomal protein-L3p-arginine acetyltransferase

LeuA-Ala-GGC

Yra-Pro-CGG

Acylamino-acid-releasing enzyme (EC 3.4.19.1)

Putative parvulin type peptidyl-prolyl isomerase, similarity with ProA foldase

Ribosomal protein-L3p-alanine acetyltransferase

26 phosphoesterase superfamily protein BCP1899

Rv12 family transcriptional regulators

Rv12 family transcriptional regulator

Rv12 family transcriptional regulator, group III

Quinone oxidoreductase (EC 1.6.5.5)

Cold shock protein CspC

Ferridoxin (EC 1.16.3.1)

Iron-binding ferritin-like antioxidant protein

Non-specific DNA-binding protein Opx

Beta-lactamase class A

Probable cadmium-transporting ATPase (EC 3.6.3.3)

Copper chaperone

ABC transporter membrane-spanning permease, Pgp export, Vex3

Potassium metabolism

Potassium homeostasis

Phages, Prophages, Transposable Elements

Phage packaging machinery

DNA helicase, phage-associated

Phage major tail protein

Phage tail length tape-measure protein

Kup system potassium uptake protein

Ribosomal protein-L3p-arginine acetyltransferase

LeuA-Ala-GGC

Yra-Pro-CGG

Acylamino-acid-releasing enzyme (EC 3.4.19.1)

Putative parvulin type peptidyl-prolyl isomerase, similarity with ProA foldase

Ribosomal protein-L3p-alanine acetyltransferase

26 phosphoesterase superfamily protein BCP1899

Rv12 family transcriptional regulators

Rv12 family transcriptional regulator

Rv12 family transcriptional regulator, group III

Quinone oxidoreductase (EC 1.6.5.5)

Cold shock protein CspC

Ferridoxin (EC 1.16.3.1)

Iron-binding ferritin-like antioxidant protein

Non-specific DNA-binding protein Opx

Beta-lactamase class A

Probable cadmium-transporting ATPase (EC 3.6.3.3)

Copper chaperone

ABC transporter membrane-spanning permease, Pgp export, Vex3
Subsystem
Iron-sulfur cluster assembly
preQ1-regulated inosine-uridine nucleoside hydrolase (EC 3.2.2.1)
Acetoin (diacetyl) reductase (EC 1.1.1.304)
ATP synthase F0 sector subunit a (EC 3.6.3.14)
DNA repair, UvrABC system
Substrate-specific component FolT of folate ECF transporter
Imidazole glycerol phosphate synthase cyclase subunit (EC 4.1.3.-)
Substrate-specific component STY3230 of queuosine-regulated ECF transporter
Biogenesis of c-type cytochromes
Cysteine desulfurase (EC 2.8.1.7), SufS subfamily
Restriction-Modification System
2'-5' RNA ligase
DNA repair, bacterial
Phosphoribosyl-AMP cyclohydrolase (EC 3.5.4.19)
Galactosylceramide and Sulfatide metabolism
Acetoin, butanediol metabolism
Adenosine deaminase (EC 3.5.4.4)
O-acetylhomoserine sulfhydrylase (EC 2.5.1.49)
HoxN/HupN/NixA family nickel/cobalt transporter
Thioredoxin-disulfide reductase
Conserved gene cluster possibly involved in RNA metabolism
Thioredoxin-disulfide reductase
Glutamate N-acetyltransferase (EC 2.3.1.35)
Histidinol dehydrogenase (EC 1.1.1.23)
Iron-sulfur cluster assembly
capsular polysaccharide biosynthesis protein
Type I restriction-modification system, specificity subunit S (EC 3.1.21.3)
Cobalt-zinc-cadmium resistance protein
Acetoin, butanediol metabolism
Acetolactate synthase small subunit (EC 2.2.1.6)
Cold shock, CspA family of proteins
ATP synthase F0 sector subunit c (EC 3.6.3.14)
Cobalt-zinc-cadmium resistance protein CzcD
Role
Acetolactate synthase large subunit (EC 2.2.1.6)
Cystathionine gamma-lyase (EC 4.4.1.1)
Ribonucleotide reductase of class III (anaerobic), large subunit (EC 1.17.4.2)
Phosphomethylpyrimidine kinase (EC 2.7.4.7)
RNA processing orphans
Imidazole glycerol phosphate synthase amidotransferase subunit (EC 2.4.2.-)
Iron-sulfur cluster assembly protein SufB
Phosphoribosyl-ATP pyrophosphatase (EC 3.6.1.31)
Dethiobiotin synthetase (EC 6.3.3.3)

Virulence, Disease and Defense
Sulfur Metabolism
Sulfur Metabolism
Stress Response
Secondary Metabolism
Secondary Metabolism
Respiration
Protein Metabolism
Potassium metabolism
Nucleosides and Nucleotides
Nucleosides and Nucleotides
Nucleosides and Nucleotides
Miscellaneous
Miscellaneous
Miscellaneous
Miscellaneous
Metabolism of Aromatic Compounds
Membrane Transport
DNA Metabolism
Cofactors, Vitamins, Prosthetic Groups, Pigments
Cofactors, Vitamins, Prosthetic Groups, Pigments
Cofactors, Vitamins, Prosthetic Groups, Pigments
Cofactors, Vitamins, Prosthetic Groups, Pigments
Clustering-based subsystems
Cell Wall and Capsule
Cell Wall and Capsule
Cell Wall and Capsule
Carbohydrates
Carbohydrates
Carbohydrates
Amino Acids and Derivatives
Amino Acids and Derivatives
Amino Acids and Derivatives
Amino Acids and Derivatives
Amino Acids and Derivatives
Amino Acids and Derivatives
Amino Acids and Derivatives
Amino Acids and Derivatives
mesenteroides vs lactis
Copper homeostasis
Cobalt-zinc-cadmium resistance
Auxin biosynthesis
Auxin biosynthesis
LysR-family proteins in Escherichia coli
LysR-family proteins in Escherichia coli
Protein degradation
Phage tail fiber proteins
Ribonucleotide reduction
Purine conversions
Benzoate degradation
ECF class transporters
ECF class transporters
Transport of Nickel and Cobalt
ABC transporter oligopeptide (TC 3.A.1.5.1)
Menaquinone and Phylloquinone Biosynthesis
5-FCL-like protein
5-FCL-like protein
Biotin biosynthesis
Glycyl-tRNA synthetase containing cluster
CBSS-176299.4.peg.1996A
Peptidoglycan Biosynthesis
Teichoic and lipoteichoic acids biosynthesis
Methionine Biosynthesis
Methionine Biosynthesis
Lysine Biosynthesis DAP Pathway
Histidine Biosynthesis
Histidine Biosynthesis
Branched-Chain Amino Acid Biosynthesis
Branched-Chain Amino Acid Biosynthesis
Branched-Chain Amino Acid Biosynthesis
Chorismate: Intermediate for synthesis of Tryptophan, PAPA antibiotics, PABA, 3-hydroxyanthranilat
Chorismate: Intermediate for synthesis of Tryptophan, PAPA antibiotics, PABA, 3-hydroxyanthranilat
Chorismate: Intermediate for synthesis of Tryptophan, PAPA antibiotics, PABA, 3-hydroxyanthranilat
Arginine Biosynthesis -- gjo
Arginine Biosynthesis -- gjo
Alanine biosynthesis
Beta-galactosidase large subunit (EC 3.2.1.23)
Glutamate--cysteine ligase (EC 6.3.2.2)
Phosphoglycerate mutase family 5
Oligopeptide transport ATP-binding protein OppD (TC 3.A.1.5.1)
Ferrochelatase, protoheme ferro-lyase (EC 4.99.1.1)
DNA double-strand break repair Rad50 ATPase
Glycerophosphoryl diester phosphodiesterase, periplasmic (EC 3.1.4.46)
N-Acetyl-D-glucosamine ABC transport system, permease protein 2
N-acetyl-L,L-diaminopimelate aminotransferase (EC 2.6.1.-)
Threonine dehydratase (EC 4.3.1.19)
3-isopropylmalate dehydrogenase (EC 1.1.1.85)
3-isopropylmalate dehydratase large subunit (EC 4.2.1.33)
Phosphoribosylformimino-5-aminoimidazole carboxamide ribotide isomerase (EC 5.3.1.16)
Anthranilate synthase, amidotransferase component (EC 4.1.3.27)
Paper III
Diversity of *Lactococcus lactis* in undefined mixed dairy starter cultures revealed by comparative genome analyses and targeted amplicon sequencing of *epsD*

Cyril A. Frantzen¹, Hans Petter Kleppen¹,², Helge Holo¹,³,*

¹Laboratory of Microbial Gene Technology and Food Microbiology, Department of Chemistry, Biotechnology and Food Science, P.O. Box 5003, N-1432 Ås Norwegian University of Life Sciences, Norway

²ACD Pharmaceuticals AS, Leknes, Norway

³TINE SA, Oslo, Norway

*Correspondence:*

Helge Holo

helge.holo@nmbu.no
Abstract

Undefined mesophilic mixed (DL) starter cultures are used in the production of continental cheeses and contain unknown strain mixtures of Lactococcus lactis and leuconsstocs. The choice of starter culture affects the taste, aroma and quality of the final product. To gain insight into the diversity of Lactococcus lactis strains in starter cultures, we whole-genome sequenced 97 isolates from three different starter cultures. Pan-genomic analyses, which included 30 publically available complete genomes, grouped the strains into 21 subsp. lactis and 28 subsp. cremoris lineages. Only one of the 97 isolates grouped with previously sequenced strains, and the three starter cultures showed no overlap in lineage distribution. Culture diversity was assessed by targeted amplicon-sequencing using purR, a core gene, and epsD, present in 95 of the 97 starter culture isolates, but absent in most of the reference strains. This allowed unprecedented discrimination of starter culture Lactococcus lactis, and revealed significant differences between the three starter cultures and compositional shifts during cultivation of cultures in milk.
Importance

In contemporary cheese production, standardized frozen seed stock starter cultures are used to ensure production stability, reproducibility, and quality control of the product. The dairy industry experiences significant disruption of cheese production due to phage attack and one commonly used countermeasure to phage attack is to employ a starter rotation strategy, in which two or more starters with minimal overlap in phage sensitivity are used alternately. Culture-independent analysis of the lactococcal diversity in complex undefined starter cultures revealed large differences between the three starter cultures, and temporal shifts in lactococcal composition during production of bulk starters. A better understanding of the lactococcal diversity in starter cultures will enable the development of more robust starter cultures, and assist in maintaining the efficiency and stability of the production process by ensuring the presence of key bacteria that are important to the characteristics of the product.
Introduction

Mesophilic mixed starter cultures (DL cultures) used in the production of continental cheeses are composed of undefined mixtures of *Lactococcus lactis* subsp. *lactis*, *Lactococcus lactis* subsp. *cremoris*, *Lactococcus lactis* subsp. *lactis* biovar. *diacetylactis* and *Leuconostoc* spp. The latter two provide aroma and texture to the cheese product (1), while *L. lactis* subsp. *lactis* and *L. lactis* subsp. *cremoris* are the major contributors in the acidification process through fermentation of lactose. Typically, contemporary starter cultures originate from traditional dairy farm cheese production based on back-slopping starter bacteria from one production to the next. Back-slopping facilitated the co-evolution of unknown numbers of strains and their bacteriophages, giving each dairy farm culture its distinct microbial composition, inherently withstanding phage attack (2).

In industrialized cheese production, standardized starter cultures are used to ensure reproducible technical and sensory properties of the product. To preserve their microbial composition, commercial starter cultures are manufactured from frozen seed stock cultures, and care is taken to minimize composition change during the production process. Even though the starter cultures are standardized, little is known about the microbial diversity and community interactions of the culture (3). Bacteriophages infecting *L. lactis* subsp. *lactis* and *L. lactis* subsp. *cremoris* are ubiquitous in dairies and can negatively affect the production process and the quality of the final product (4, 5). Starter cultures originating from traditional cheese farms are considered more robust against phage attack compared to defined cultures (2), a characteristic gained from their large number of strains with diverse phage sensitivity (6). Because industrial cheese production is dependent on predictable starter culture performance, the use of frozen batch inoculum is often preferred to back slopping. This effectively halts the lactococcal
evolution, while giving phages the advantage of evolving freely in the dairy environment (5).

Thus, the dairy industry experiences significant disruption of cheese production due to phage attack.

One countermeasure to phage attack is to employ a starter rotation strategy, in which two or more starters with minimal overlap in phage sensitivity are used alternately. However, the choice of starter culture may affect taste, aroma, and quality of the final product. Since very little knowledge exists on genetic diversity of the bacteria or the microbial composition constituting undefined DL-starters, it is difficult to decide which starters to use in a rotation strategy (7).

Bacteriophages are frequently found in the dairy environment, often in very high titers (4, 8, 9). However, in fermentation failures with DL starter cultures, the diversity of phages rather than their quantity appears to be more important (4).

Knowledge on the microbial diversity of starter cultures is limited, and the complexity and diversity of DL starter cultures beyond sub-species is unknown (2). In order to better predict production performance and advise functional culture rotation strategies it is of the utmost importance to characterize the strain diversity of DL and other undefined starter cultures.

Moreover, identification of key starter culture strains important to the character of the product will drastically improve the ability to assess the impact of phage attack. With the advances in high-throughput DNA sequencing technology in the recent years, and the significant increase in lactococcal genomic data available to the scientific community, new opportunities have emerged to achieve this. Here, we present pan-genomic differentiation of lactococci obtained from DL starter cultures and show significant differences in the lactococcal diversity between DL starter cultures using targeted-amplicon sequencing.
**Method and Materials**

**Cultivation and isolation of strains**

All bacterial strains used in this study are listed in supplementary table S1. The media used for cultivation were M17 (10) supplemented with 0.5% (w/v) lactose (Merck, Kenilworth, New Jersey, USA) or 10% (w/v) skimmed milk powder (TINE SA, Oslo, Norway) supplemented with 50 mM β-glycerophosphate (Sigma-Aldrich, Munich, Germany) (GM) as proposed by Hugenholtz (11). Bulk starters were produced by incubating commercial starter cultures in 10% (w/v) skim milk at 22°C for 14 hours in triplicate. Commercial starter cultures were suspended in GM to an optical density at 600nm (OD600) of 1.0, serially diluted in 10% (w/v) skim milk and spread plated on M17- and GM-agar plates in triplicate. The plates were incubated at 22 °C for 5 days before colonies were picked. Isolates were transferred to M17 and GM broth media respectively, and cultivated at 22°C for two passages before aliquots were added 15 % (w/v) glycerol (Sigma-Aldrich) and stored at -70 °C.

**Genome sequencing, assembly and annotation**

Genomic DNA from lactococcal isolates was extracted from 1mL of overnight culture using Qiagen DNeasy Blood & Tissue Kit (Qiagen, Hilden, Germany). The cells were lysed with 40 mg/mL lysozyme (Qiagen, Hilden, Germany) prior to column purification. DNA libraries were constructed using the Nextera XT DNA Sample Prep kit (Illumina, San Diego, California, USA) according to manufacturer instructions and sequenced on an Illumina MiSeq (Illumina, San Diego, California, USA) platform using V3 chemistry. Raw sequences were adapter trimmed,
quality filtered (Q>20), *de novo* assembled using SPAdes V3.10.1 (12) and annotated using the Prokka v1.12 pipeline (13). Contigs shorter than 1000bp or with less than 5x coverage were removed from each assembly prior to gene annotation. This whole genome project has been deposited at DDBJ/ENA/GenBank under the BioProject PRJNA392995. In addition, 30 publically available complete *L. lactis* subsp. genomes were acquired from the National Center for Biotechnology Information (NCBI) genomes database (Supplementary table S1) (14-27). These genomes were re-annotated using the Prokka v1.12 pipeline.

**Pan-/Core-genomic analysis**

The protein coding sequences of all isolates were compared by an all-against-all approach using BLASTP (28) and grouped into orthologous clusters using GET_HOMOLOGUES v2.0.10 (29). Pan- and core-genome sizes were estimated using the pan-genomic analysis tool PanGP v1.0.1 (30). Orthologous groups (OGs) were identified via the Markov Cluster Algorithm (MCL) with an inflation value of 2.5 (31) and intersected using the compare_clusters.pl script provided with GET_HOMOLOGUES. The orthologous clusters were curated to exclude significantly divergent singletons, which are likely to be the result of erroneous assembly or annotation. A pan-genomic presence/absence matrix was constructed, including each gene cluster and each genome. Hierarchal single-linkage clustering analysis of this matrix was performed in R (http://www.r-project.org/) to construct a pan-genome heatmap overview using the heatmap.2 function included in the Gplots package v2.16 (32) supplemented by the Dendextend package v0.18.3 (33). Genes were divided into three categories, core-genes, which are present in all genomes, softcore-genes, which are present in above 95% of genomes, and pan-genes, which are all the genes present in one or more genomes. Core-genes were included in a multi-locus multiple alignment scheme to
determine phylogenetic distances between genomes and to construct a WPGMA phylogenetic
supertree using the sequence alignment metric functions in the Decipher v2.0 (34) and MASS
v7.3-47 (35) packages in R. A distance cut-off for the number of clusters was determined using
the knee of the curve approach (36), binning the isolates into genomic lineages.

Relative quantification of the microbial community in starter cultures
Compositional analysis in starter cultures was performed in triplicates on total DNA extracted
from the starter cultures using 1 mL of starter culture diluted to an OD$_{600}$ of 1. The samples were
treated with 20 mg/mL lysozyme (Sigma-Aldrich) and 3 U/L mutanolysin (Sigma-Aldrich),
mechanically lysed using FastPrep (MP Biomedicals) with 0.5 g acid-washed glass beads (<106
µm) (Sigma-Aldrich) and purified using the Qiagen DNeasy Blood & Tissue Kit (Qiagen). A
suitable amplicon target was identified by screening the softcore-genes for nucleotide sequence
variation using the sequence alignment metrics functions in the DECIPHER package v1.16.1
(34). Genes without flanking consensus regions within a <500bp variable region adequate for
differentiation, or which did not provide sufficient discrimination between lineages, were
discarded. The loci purR and epsD, and the v2-v3 region of 16S rDNA were amplified by PCR
using the KAPA HiFi PCR Kit (KAPA Biosystems, Wilmington, Massachusetts, USA) with the
following primers: purR-324F (5’-YACTCCATCAAATCTTCGTAAAAT-3’), purR-811R (5’-
TGTCACTAAATATATTCTTCCCAATTGAACA-3’), epsD-138F (5’-
KCTTATYGCGGCTGCATT-3’), epsD-604R (5’-GATARTARAGTTCTAAATCTGCTCGT-
3’), 16S-44F (5’- GCCTGCTAATACATGCAAGTYGA-3’), 16S-536R (5’-
CTGCTGGCACGTAKTTAGCCGTCC-3’). Forward (5’-
TCGTCGGCAGCGTCAGATGTGATATAAGACAG-3) and reverse
Illumina adapter overhangs were added to the 5’ end of the primers to allow for Nextera XT DNA indexing of the PCR-products. The libraries were sequenced on the Illumina MiSeq platform using V3 (2x300bp) reagents. The resulting data were paired-end-merged and quality filtered using PEAR (37) and clustered using VSEARCH v2.4.3 (38) with error-minimization from USEARCH v10.0.240 (39). When quantifying at the species and subspecies level, the 16S rDNA and purR amplicon data was clustered using the common identity level threshold of 97% (40, 41). When quantifying at the level of genetic lineages, the purR and epsD data was clustered by a similarity threshold of 99.5%, corresponding to a nucleotide difference of two single-nucleotide polymorphisms. For taxonomic classification, the resulting Operational Taxonomic Unit (OTU) was matched against a local BLAST-database produced using the lactococcal genomes sequenced in this study, as well as the lactococcal genomes available on the NCBI.
Results

Isolation and whole-genome sequencing of bacteria

The microbial diversity of three commercially available DL starter cultures (A, B, and C) was assessed, mainly focusing on culture A. The starter cultures were acquired from three different culture manufacturers. To increase the likelihood of high diversity representation, two different growth media and phage typing was used (42). We focused on culture A, selecting sixty-six isolates from starter culture “A”, and complemented those with 15 isolates from culture “B” and 14 isolates from culture “C”. The 97 lactococcal isolates were whole-genome sequenced on an Illumina MiSeq platform. Thirty complete Lactococcus lactis genome sequences acquired from the National Center for Biotechnology Information (NCBI) were also included in the study as reference genomes.

Pan-/Core-genome analysis

All the coding sequences (CDS) in the genomes were compared by a blast all-against-all approach to identify orthologous gene groups (OGs) and construct pan- and core-matrices. The pan- and core-genome sizes were determined at 8064 OGs, and 551 OGs, respectively (Figure 1). Pan-genomic differentiation of isolates using hierarchal clustering on the pan-matrix clearly separated subsp. lactis from subsp. cremoris (Figure 2), as did the core-genome analysis using 551 genes to construct a phylogenetic supertree (Figure 3). Analysis of the 127 Lactococcus lactis genomes (Supplementary table S1) showed that 64 of these belonged to subspecies cremoris, and 63 to subspecies lactis. Interestingly, analysis of 16S rDNA revealed that a number of isolates (CF103, CF117, CF128, CF129, CF207, CF223, CF229), all identified as
subsp. cremoris in the pan- and core-genome analysis, contain a novel and unique 16S rDNA sequence more closely related to subsp. lactis type than subsp. cremoris type (Supplementary Figure S1). Analysis of the 16S rDNA sequence confirmed that all 16S rDNA copies in the genomes are of this novel variant. Discrepancies in subspecies identification of lactococci using 16S rDNA have also been reported in previous studies (43, 44).

**Differentiation and clustering of genomes**

Robust genotypic discrimination was achieved through analysis of the pan-genome in combination with nucleotide variation in core-genes. This provided high-resolution differentiation of isolates beyond subspecies (Figure 2). The 63 L. lactis subsp. lactis isolates clustered into 21 genetic lineages (L1-L21), while the 64 L. lactis subsp. cremoris isolates clustered into 28 genetic lineages (C1-C28) (Supplementary table S2). The L. lactis subsp. lactis isolates from our starter cultures fell into 11 of the 21 lineages (Table 2), while the reference genomes occupied the other ten. Notably, the lineages appear culture specific, as no lineage was represented in more than one culture. The reference strains IL1403, 229, and UC77, all isolated from dairy, belong to the same clade as the starter culture isolates, while the other reference subsp. lactis strains showed a more distant relationship to the strains in our starter cultures. The L. lactis subsp. cremoris isolates from our starter cultures clustered into 21 of the 28 lineages. With one exception, we also observed a culture specific lineage distribution for these isolates (Table 3). One isolate from starter culture B clustered with the reference strains 158, UC509.9, and UC109. As shown in Figure 3, most of the reference strains and all of our starter culture isolates grouped into two clades. Only the reference strains MG1363, NZ9000 and KW2 did not fall into these clades.
Identification of amplicon targets for strain differentiation

In order to devise a scheme for differentiation and quantification of the microbial diversity in each of the starter cultures by amplicon sequencing, core-genes and softcore-genes were screened for sequence variation reflecting the genomic differentiation. After curation of targets, the core-gene purR encoding a purine biosynthesis repressor (45), and the softcore-gene epsD, part of the eps capsular polysaccharide biosynthesis operon (46, 47) were selected as amplicon targets. Among the core-genes, purR was the candidate with the largest number of unique amplicons, with 25 variants (Supplementary Figure S2). The topology of the phylogenetic tree made using the purR amplicon corresponds to the core-genomic supertree, neither of which provide a resolution sufficient to reflect the genetic lineages defined by the pan-genome analysis. Importantly, discrimination between subspecies using the purR amplicon coincided to the subspecies classification made by the pan- and core-genome analyses. An even larger number of variants among our starter isolates was identified in the softcore-gene epsD. This gene was present in all except two of our isolates (CF124 and CF223) but only in nine out of the 30 reference strains and presented with 33 variants (Supplementary Figure S3). Altogether, 26 epsD variants were found in the sequenced strain collection from our starter cultures with a sequence distribution corresponding to the pan genomic lineages. No lineage was represented by more than one epsD sequence variant, but a few lineages (L7 and L12, C2 and C7, C5 and C9 and C16, C6 and C23) shared epsD sequences.

Microbial diversity in the starter cultures

Assessment of the microbial diversity in three starter cultures was performed by targeted-amplicon sequencing of three loci, the V2-V3 region of 16S rDNA, the purR gene (pos. 324-
Quantification of microbial diversity was performed on frozen starter cultures, and on bulk starters grown at 22°C for 14 hours. The results revealed big differences between the starter cultures, as well as significant shifts in the microbial composition during bulk starter manufacture. The amplicon data for 16S rDNA showed significant differences in the microbial composition between the starter cultures (Table 1). All cultures were dominated by *L. lactis* subsp. *cremoris*, although this was most prominent in culture B with more than 70% *L. lactis* subsp. *cremoris*. A small decrease in *L. lactis* subsp. *cremoris* was shown in cultivation of the bulk starter for all three cultures. The content of leuconostocs varied from <1% in culture B to 24.6% in culture A, and 29.4% in culture C. Relative quantification of lactococcal subspecies was performed using the *purR* amplicon data as well using the commonly used 97% clustering threshold. By comparing the *purR* and 16S rDNA amplicon data, a significant underestimation of *L. lactis* subsp. *cremoris* was identified in the 16S rDNA data (Figure 4). The discrepancy varied from 4.5% in the bulk starter of culture C to 15.5% in the frozen culture of culture B. This demonstrates the impact of strains containing the 16S rDNA sequences which clutter subspecies identification as described earlier. Moreover, this shows that such sequences are not unique to culture A, but present in all three cultures.

**Large strain diversity**

To assess the genetic diversity in the three starter cultures, amplicon-targeted sequencing of *purR* and *epsD* was performed. Using a 99.5% similarity threshold to cluster the amplicon data into OTUs, significant differences between the genetic diversity of the starter cultures were revealed. Moreover, a number of OTUs were found to be specific to their culture showing that a large proportion of the strains did not overlap between the starter cultures.
The *purR* amplicon sequences clustered into 17 OTUs (Table 2) and enabled relative quantification corresponding to the core-genomic differentiation of strains as shown in Figure 3. The results show considerable differences in the *purR* diversity in the three starter cultures and their corresponding bulk starters (Figure 5). Of the 17 distinct *purR* OTUs, 10 were found in Culture A, 8 in culture B, and 13 in culture C. Two OTUs unique to culture A, one OTU unique to culture B, and two OTUs unique to culture C were identified. The culture specific OTUs accounted for a substantial proportion in cultures A and C, amounting to 21.7% and 34.3%, respectively, in frozen cultures, declining significantly during bulk starter cultivation to 13.4% and 20.3%, respectively. Culture A and B were dominated by Otu2, corresponding to several genetic lineages. The same OTU was also abundant in culture C. A noteworthy difference between the cultures was observed for Otu1, a subsp. *lactis* type OTU reflecting the higher abundance of subsp. *lactis* in culture C compared to culture A and B. The remaining *purR* OTUs were detected in all three starter cultures, OTU5, 6, 9, 12, 13 in considerable amounts, and OTU10, 11, 14, 15 and 16 in trace amounts (Table 2). Five of the 17 OTUs were novel variants not found in any of our genomes.

The *epsD* amplicon sequences clustered into 52 OTUs (Table 3), enabling high-resolution quantification of the genetic diversity among *eps* positive strains present in the starter cultures. The results show substantial differences in *epsD* diversity between the three starter cultures and their corresponding bulk starters (Figure 6). Of these 52 OTUs, 31 were found in culture A, 28 in culture B, and 18 in culture C. Most of these *epsD* OTUs, 13 in culture A, 9 in culture B, and 11 in culture C, were culture specific. The specific OTUs amounted to a large proportion of the total population. The OTUs unique to culture A (OTU15, OTU20, OTU24, OTU26, OTU31, OTU36, OTU38, OTU40, OTU41, OTU43, OTU44, OTU48, and OTU49) amounted to 18.9% of the
population in the frozen starter, and 32.6% in the bulk starter. Culture B specific OTUs (OTU1, OTU8, OTU14, OTU21, OTU25, OTU33, OTU42, OTU50, and OTU52) amounted to 54.0% of the population in the frozen starter, and 52.5% of the population in the bulk starter. Lastly, OTUs unique to culture C (OTU6, OTU7, OTU9, OTU12, OTU17, OTU22, OTU32, OTU35, OTU37, OTU39, and OTU47) amounted to 71.9% of the population in the frozen starter, and 65.8% of the population in the bulk starter. This showed that a substantial proportion of the genetic diversity did not overlap between the starter cultures. The remaining 19 OTUs were not culture specific, but were highly variable with regards to their abundances and degree of overlap between the starter cultures. Six of the OTUs (OTU2, OTU3, OTU4, OTU5, OTU10, and OTU11) were found in significantly higher abundances in one of the cultures compared to the other two. OTU2 was abundant in culture A and B, but not detected at all in culture C. OTU3 was detected in all cultures, although was significantly more abundant in culture B, compared to culture A or C. OTU4, OTU5 and OTU11 were detected in all cultures, but was significantly more abundant in culture A, than in the other two. Lastly, OTU10 was detected in culture B and C, but not A, and was significantly more abundant in culture C compared to culture B. The remaining 13 OTUs (OTU13, OTU16, OTU18, OTU19, OTU23, OTU27, OTU28, OTU29, OTU30, OTU34, OTU45, OTU46, OTU51) were more evenly distributed between the starter cultures. However, they all presented with abundances of ~2% or lower. The epsD OTUs were all assessed using BLAST to identify closely related sequences. Nineteen of the 52 distinct epsD OTUs were a > 99.5% match with our isolates from starter cultures, while the remaining 33 epsD OTUs were new variants. Interestingly, these 33 epsD OTUs were not higher than 99.4% identity to any sequences included on the NCBI either, showing that they are indeed novel variants.
Discussion

*Lactococcus lactis* is predominantly associated with cheese production and has been subject to extensive research regarding both phenotypic traits and genetic diversity. While suggested to have originated from the plant environment (43), the genetic content of dairy-associated *L. lactis* is easily distinguished from that of its non-dairy counterpart. Evidence of genome decay in the process of adapting to the dairy environment has been accentuated in both *L. lactis* subspecies, but to a larger extent in *L. lactis* subsp. *cremoris* (27). The distinction between subspecies *lactis* and *cremoris* was initially based on phenotypic features. Since then, detailed studies on the genetic relatedness of the subspecies has shown that phenotypic features alone are inadequate to identify subspecies (48). Moreover, there is a discrepancy between the subspecies identification determined by phenotypic features with genotypic identification determined using 16S rDNA sequences (49). Strains of *L. lactis* identified as subspecies *cremoris* by genotype have been reported to show a subspecies *lactis* phenotype, and vice versa, making accurate identification and differentiation of isolates a difficult task (49, 50). Using a wide range of molecular fingerprinting methods and sequencing schemes, a large genetic diversity of *L. lactis* has been shown to exist within the dairy environment (27, 51, 52).

Our analyses of 127 *L. lactis* genomes clearly showed a large genetic diversity among dairy strains. The high-resolution of the pan-genome analysis enabled differentiation beyond subspecies, distributing the *L. lactis* subsp. *lactis* isolates into 21 genetic lineages, and the *L. lactis* subsp. *cremoris* isolates into 28 genetic lineages. Phylogenetic analysis of 551 core-genes clearly distinguished between dairy and non-dairy lactococci, and also separated between isolates from DL starter cultures and isolates from other dairy sources. Moreover, most of the lactococci from our DL starter cultures were found to fall into culture specific genetic lineages, reflecting...
spatially separated evolution of strains. Previously, the overlap in sensitivity to bacteriophages
between starter cultures A, B and C has been shown to be minimal (42), corroborating this
finding.

The lactococcal population of an undefined mesophilic starter culture has previously been
divided into seven groups (TIFN1-7) based on AFLP (51), that were quantified in a metagenome
dataset using group specific gene markers (3). None of our isolates contained the gene markers
specific for TIFN1-6. However, 19 of our subspecies cremoris isolates did contain the gene
marker specific to TIFN7. These include isolates from both media and were scattered amongst
several pan-genomic lineages comprised of 36 isolates. Interestingly, all the isolates belonging to
lineages C1, C3, C5, C9, C27 and C28 did not contain the TIFN7 gene marker. This shows that
the method of Erkus is not applicable to cheese cultures in general, but was specific to their
culture. Moreover, it highlights the limitations of using unique loci as genetic markers, compared
to using the sequence variation in conserved genes in culture-independent analysis of complex
microbial communities.

During propagation by back-slopping regimes, the microbial community of complex
starter cultures is sustained (2). However, the composition of the culture may change
significantly over shorter time periods depending on growth conditions and phage predation (3).
The dairy industry depends on reliable and reproducible culture performance, and avoid day-to-
day variations by using frozen seed stock cultures, effectively resetting the microbial
composition every day of production. Our analyses showed that starter cultures are indeed
complex and our cultures showed very little overlapping diversity. We found significant
differences in the lactococcal composition of three starter cultures acquired from three different
culture manufacturers, and showed that they changed during propagation in milk. Moreover, the
cultures are significantly different in their content of leuconostocs. In a previous study we showed large differences in *Leuconostoc* diversity between the same cultures (53). Fluctuations in the community during manufacture have an effect on the functionality of the starter such as acidification or flavor formation (54). Composition analysis of the microbial community is an important tool in the work to ensure maintenance of culture diversity, assessing the effects of phage attack, and monitoring the performance of the culture. More reproducible starter compositions can be obtained by adjusting the culture parameters.

Using targeted-amplicon sequencing, the downstream data analysis cluster the sequences into OTUs. The OTU assignments are dependent on the DNA sequence similarity threshold, which has typically been set at 97% in studies involving 16S rDNA (55). Several studies have pointed out that this threshold is excessively low, and suggest the use of a higher threshold (56-58). Recently, the use of SNP distances or so-called zero-radius OTUs (zOTUs) have become common and computer programs have been developed to accommodate this (57, 58). The advantage of increasing the threshold is a higher-resolution OTU assignment, and a significant reduction in the inflation of OTU abundances by false positives (56). In a review of molecular fingerprinting and culture-independent methods, (59) concluded that a sufficient analytical resolution could only be achieved through identification of a conserved, but highly variable locus for strain discrimination. The DNA-sequences of protein-coding genes have been shown to be more effective than 16S rDNA when distinguishing between very closely related bacteria (59, 60). Typically, housekeeping genes are the preferred targets when differentiating between strains. By these criteria, *purR* was the best candidate and enabled differentiation of clades beyond the sub-species level, as well as differentiation of subspecies superior to that of 16S rDNA. In comparison with our *purR* analyses, a significant underestimation of *L. lactis* subsp.
cremoris by 16S rDNA was demonstrated. This highlights the advantage of species-specific
amplicon targets compared to that of 16S rDNA. However, the sequence variation within the
purR amplicon was insufficient to differentiate between many of the genetic lineages. Thus, the
variance within the amplicons found among our core genes, is not high enough to expose the
complexity of DL starter cultures.

By expanding the amplicon search to include softcore-genes represented in at least 95%
of the genomes, an amplicon able to differentiate the genetic lineages from each other was found
in epsD. The pan-genome analyses discerned 33 epsD variants, 27 of which were found in our
starter culture isolates. Using this amplicon, an unprecedented resolution of differentiation
between genetic lineages was achieved. Interestingly, the phylogenetic analysis of epsD did not
separate subspecies lactis from subspecies cremoris at the root of the tree like purR and 16S
rDNA. Rather, subspecies separation was made on branches further out on the tree, a strong
indication of horizontal gene transfer. The analysis also identified new epsD sequence variants
present in low abundances. The results showed a low, but not zero overlap in epsD variants
between the starter cultures. Part of this overlap emerges from culture-specific genetic lineages
clearly separated in the pan-genome analysis, but which all contain the same epsD variant and
can not be distinguished from each other in the amplicon analysis. Most of the overlapping
OTUs were low abundance OTUs, and a significant proportion of the culture population is
composed of culture-specific OTUs.

The discovery of epsD as a suitable target for strain differentiation was surprising, as the
gene was only present in 9 of the 30 reference strains. The eps operon has been found both on
plasmids (61, 62) and to be chromosomally located (46). The epsD gene was highly represented
among the starter culture strains, missing in only two of our 97 starter culture isolates. Apart
from the missing *eps* operon, we were unable to distinguish the two isolates CF124 and CF223 from their nearest pan- and core-genomic neighbors. In the laboratory, strains harboring *eps* plasmids have been cured of their *eps* positive phenotype by serial transfers (61), and no evidence exists that suggests a chromosomal locality confers higher stability over multiple transfers (46). The high degree of sequence variation in the *eps* operon, and more specifically the sequence variation in the *epsD* amplicon represents evolutionary diversification indicating a history of selection pressure. Typically, lactococcal strains with different phage sensitivities also contain different EPS, and strains that do not produce EPS have been demonstrated to exhibit phage sensitivities different from strains which do produce EPS (61). Moreover, the production of EPS has been shown to confer resistance to phages (62, 63). Regardless of what might be the cause of the high degree of sequence variation in the *epsD* gene, its applied use in discrimination and quantification of lactococcal diversity provides culture-independent, robust, and reproducible data. Moreover, it provides the means to monitor temporal shifts in lactococcal diversity, as well as comparing the genetic diversity of *Lactococcus lactis* between starter cultures and starter culture batches.

The great rate of advancement in next-generation sequencing technologies over the past decade has been accompanied by a rapid development of bioinformatics applications. The reduced cost of sequencing has promoted whole-genome sequencing of bacterial isolates, and the vast improvements to the downstream analysis of genomic data has taken comparative analysis to a completely new level. Pan-genomic analysis of several hundred genomes enables characterization and differentiation of bacteria, and facilitates the development of rapid and robust methods such as targeted-amplicon sequencing of discriminatory loci. Dairy starter cultures are simple compared to the complexity of other environmental samples such as soil or
mammalian gut, and could be a good model for the development of groundbreaking methods for
differentiating bacteria. Our method of comparative genome analyses of whole-genome
sequenced isolates provides a robust method of discovering intra-species gene markers for
targeted-amplicon sequencing, and could be applicable to other microbial niches. The use of
purR and epsD as gene markers for *Lactococcus lactis*, enables intra-species differentiation of
genetic lineages in O, L, D and LD starter cultures. The application of the analysis to a
completely new starter culture should be prefaced by initial amplicon sequencing of the culture
to assess the culture diversity, and possibly complemented by whole-genome sequencing of
isolates to ensure the validity and continuity of the analysis.

In conclusion, our comparative genomic analysis enabled discrimination of 127
*Lactococcus lactis* genomes in to 38 genetic lineages. Significant compositional differences were
revealed between starter cultures and temporal shifts in the lactococcal population during
cultivation using amplicon-targeted sequencing of *epsD*. The EPS genotype is highly conserved,
yet *epsD* displays high sequence variability which enables culture-independent identification and
quantification of *Lactococcus lactis*. Using high-resolution culture-independent methods such as
targeted-amplicon sequencing of *epsD* and *purR*, a better understanding of the microbial
composition of starter cultures can be achieved. This will enable development of more robust
starter cultures, and assist in maintaining the stability of the culture by ensuring the presence of
key bacteria that are important to the characteristics of the product.
Conflict of Interest

The authors declare no conflict of interest.

Acknowledgments

We are grateful to TINE SA for providing culture samples. This work was funded by the Norwegian Research Council and TINE SA.
References

of Lactococcus lactis IO-1, a lactic acid bacterium that utilizes xylose and produces high levels of L-lactic acid. J Bacteriol 194:2102-2103.


34. Wright E. 2016. Using DECIPHER v2.0 to Analyze Big Biological Sequence Data in R. The R Journal 8:352-359.


Figure 1: Pan- and core-genome sizes for 127 *Lactococcus lactis* isolates using PanGP and the traverse all approach was estimated at 8064, and 551, respectively.
Figure 2: Heatmap representation of the pan-genome content of 127 *Lactococcus lactis* isolates. The black regions indicate orthologous groups that are present, while the grey regions indicate orthologous groups that are absent. Using hierarchal single-linkage clustering, with a distance cut-off determined using the knee of the curve approach, 63 subsp. *lactis* isolates clustered into 21 genetic lineages, and the remaining 64 subsp. *cremoris* isolates clustered into 28 genetic lineages.
Figure 3: Differentiation of *Lactococcus lactis* using 551 core-genes.
Figure 4: Comparison of the quantification of *Lactococcus lactis* subsp. in starter cultures using the 16S rDNA and *purR* loci. Compared to *purR*, 16S underreports the relative amount of subsp. *cremoris* compared to subsp. *lactis*. 
Figure 5: *Lactococcus lactis* diversity in three commercial starter cultures A, B, C using targeted-amplicon sequencing of *purR*. OTU1, 3, 10 and 14 were identified as *L. lactis* subspecies *lactis*, while OTU2, 4, 5, 6, 7, 8, 9, 11, 13, 15, 16, 17 were identified as *L. lactis* subspecies *cremoris*. 
Figure 6: *Lactococcus lactis* diversity and relative abundance in starter cultures using targeted-amplicon sequencing of *epsD*. The OTUs identified as *L. lactis* subspecies *lactis* were OTU1, 3, 4, 7, 8, 15, 20, 22, 24, 27 and 50. The OTUs identified as *L. lactis* subspecies *cremoris* were OTU2, 5, 11, 13, 14, 16, 18, 19, 23, 25, 29, 31, 34, 35, 36, 38, 41, 45, 46 and 51. OTU6, 9, 10, 12, 17, 21, 26, 28, 30, 32, 33, 37, 39, 40, 42, 43, 44, 47, 48, 49, and 52 could not be assigned subspecies.
Table 1: Microbial diversity and relative abundances (percentage) in starter cultures A, B, and C. Analysis was performed by amplicon sequencing of the V2-V3 region of 16S rDNA, clustered at 97% using vsearch.

<table>
<thead>
<tr>
<th></th>
<th>Culture A</th>
<th></th>
<th>Culture B</th>
<th></th>
<th>Culture C</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Frozen</td>
<td>Bulk</td>
<td>Frozen</td>
<td>Bulk</td>
<td>Frozen</td>
<td>Bulk</td>
</tr>
<tr>
<td><em>L. lactis</em> subsp. <em>cremoris</em></td>
<td>58.8</td>
<td>47.6</td>
<td>77.9</td>
<td>73.8</td>
<td>47.0</td>
<td>33.4</td>
</tr>
<tr>
<td><em>L. lactis</em> subsp. <em>lactis</em></td>
<td>24.7</td>
<td>27.8</td>
<td>21.4</td>
<td>25.7</td>
<td>34.6</td>
<td>37.2</td>
</tr>
<tr>
<td><em>Leuconostoc</em> spp.</td>
<td>16.6</td>
<td>24.6</td>
<td>0.8</td>
<td>0.5</td>
<td>18.4</td>
<td>29.4</td>
</tr>
</tbody>
</table>
Table 2: Genetic diversity and relative abundances (percentage) of *Lactococcus lactis* OTUs in starter cultures A, B, and C using targeted-amplicon sequencing of *purR*. The OTUs were generated by clustering *purR* sequences at a 99.5% similarity threshold.

<table>
<thead>
<tr>
<th>#OTU ID</th>
<th>Culture A</th>
<th></th>
<th>Culture B</th>
<th></th>
<th>Culture C</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Frozen</td>
<td>Bulk</td>
<td>Frozen</td>
<td>Bulk</td>
<td>Frozen</td>
<td>Bulk</td>
</tr>
<tr>
<td>OTU1&lt;sup&gt;a&lt;/sup&gt;</td>
<td>3</td>
<td>5,5</td>
<td>2,3</td>
<td>6,2</td>
<td>16,9</td>
<td>30,5</td>
</tr>
<tr>
<td>OTU2&lt;sup&gt;b&lt;/sup&gt;</td>
<td>52</td>
<td>50,4</td>
<td>81,4</td>
<td>75</td>
<td>24,6</td>
<td>13</td>
</tr>
<tr>
<td>OTU3&lt;sup&gt;a&lt;/sup&gt;</td>
<td>6,1</td>
<td>9,7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>OTU4&lt;sup&gt;a&lt;/sup&gt;</td>
<td>15,6</td>
<td>3,7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>OTU5&lt;sup&gt;b&lt;/sup&gt;</td>
<td>2,2</td>
<td>1,8</td>
<td>3,8</td>
<td>5,4</td>
<td>2</td>
<td>6,1</td>
</tr>
<tr>
<td>OTU6&lt;sup&gt;a&lt;/sup&gt;</td>
<td>11,8</td>
<td>14,8</td>
<td>3,4</td>
<td>2,3</td>
<td>3,3</td>
<td>6,4</td>
</tr>
<tr>
<td>OTU7&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>19,1</td>
<td>17,4</td>
</tr>
<tr>
<td>OTU8&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0</td>
<td>0</td>
<td>0,9</td>
<td>0</td>
<td>15,2</td>
<td>2,9</td>
</tr>
<tr>
<td>OTU9&lt;sup&gt;a&lt;/sup&gt;</td>
<td>2,5</td>
<td>6,2</td>
<td>0</td>
<td>1</td>
<td>5,1</td>
<td>4,4</td>
</tr>
<tr>
<td>OTU10&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1,3</td>
<td>1,9</td>
</tr>
<tr>
<td>OTU11&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1,6</td>
</tr>
<tr>
<td>OTU12&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0,9</td>
<td>1,7</td>
<td>0</td>
<td>1,8</td>
<td>0</td>
<td>8,8</td>
</tr>
<tr>
<td>OTU13&lt;sup&gt;a&lt;/sup&gt;</td>
<td>1,2</td>
<td>1,3</td>
<td>0</td>
<td>1,2</td>
<td>1,8</td>
<td>2,5</td>
</tr>
<tr>
<td>OTU14&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1,2</td>
<td>0</td>
</tr>
<tr>
<td>OTU15&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1,1</td>
</tr>
<tr>
<td>OTU16&lt;sup&gt;b&lt;/sup&gt;</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>OTU17&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0</td>
<td>0</td>
<td>4,3</td>
<td>2,5</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

<sup>a</sup> OTUs identified as *L. lactis* subspecies *lactis*.

<sup>b</sup> OTUs identified as *L. lactis* subspecies *cremoris*. 
Table 3: Genetic diversity and relative abundance (percentage) of *Lactococcus lactis* OTUs in starter cultures A, B, and C using targeted-amplicon sequencing of the *epsD* gene. OTUs were generated by clustering *epsD* sequences at 99.5% similarity threshold.

<table>
<thead>
<tr>
<th>#OTU ID</th>
<th>Culture A</th>
<th>Culture B</th>
<th>Culture C</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Frozen</td>
<td>Bulk</td>
<td>Frozen</td>
</tr>
<tr>
<td>OTU01(^a)</td>
<td>0.0</td>
<td>0.0</td>
<td>28.6</td>
</tr>
<tr>
<td>OTU02(^b)</td>
<td>24.0</td>
<td>8.6</td>
<td>9.8</td>
</tr>
<tr>
<td>OTU03(^b)</td>
<td>3.3</td>
<td>5.6</td>
<td>17.7</td>
</tr>
<tr>
<td>OTU04(^b)</td>
<td>18.6</td>
<td>8.6</td>
<td>1.3</td>
</tr>
<tr>
<td>OTU05(^b)</td>
<td>13.0</td>
<td>20.7</td>
<td>0.8</td>
</tr>
<tr>
<td>OTU06(^c)</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>OTU07(^b)</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>OTU08(^b)</td>
<td>0.0</td>
<td>0.0</td>
<td>13.3</td>
</tr>
<tr>
<td>OTU09(^b)</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>OTU10(^b)</td>
<td>0.0</td>
<td>0.0</td>
<td>0.8</td>
</tr>
<tr>
<td>OTU11(^b)</td>
<td>8.2</td>
<td>3.7</td>
<td>2.2</td>
</tr>
<tr>
<td>OTU12(^b)</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>OTU13(^b)</td>
<td>1.0</td>
<td>1.4</td>
<td>0.4</td>
</tr>
<tr>
<td>OTU14(^b)</td>
<td>0.0</td>
<td>0.0</td>
<td>3.8</td>
</tr>
<tr>
<td>OTU15(^b)</td>
<td>4.3</td>
<td>8.3</td>
<td>0.0</td>
</tr>
<tr>
<td>OTU16(^b)</td>
<td>1.9</td>
<td>3.2</td>
<td>2.4</td>
</tr>
<tr>
<td>OTU17(^b)</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>OTU18(^b)</td>
<td>1.1</td>
<td>2.2</td>
<td>0.5</td>
</tr>
<tr>
<td>OTU19(^b)</td>
<td>0.5</td>
<td>1.4</td>
<td>3.7</td>
</tr>
<tr>
<td>OTU20(^b)</td>
<td>1.7</td>
<td>4.3</td>
<td>0.0</td>
</tr>
<tr>
<td>OTU21(^b)</td>
<td>0.0</td>
<td>0.0</td>
<td>2.4</td>
</tr>
<tr>
<td>OTU22(^b)</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>OTU23(^b)</td>
<td>1.5</td>
<td>1.1</td>
<td>1.3</td>
</tr>
<tr>
<td>OTU24(^b)</td>
<td>2.5</td>
<td>4.7</td>
<td>0.0</td>
</tr>
<tr>
<td>OTU25(^b)</td>
<td>0.0</td>
<td>0.5</td>
<td>2.0</td>
</tr>
<tr>
<td>OTU26(^b)</td>
<td>0.9</td>
<td>2.8</td>
<td>0.0</td>
</tr>
<tr>
<td>OTU27(^b)</td>
<td>1.1</td>
<td>1.9</td>
<td>1.0</td>
</tr>
<tr>
<td>OTU28(^b)</td>
<td>2.0</td>
<td>2.7</td>
<td>0.0</td>
</tr>
<tr>
<td>OTU29(^b)</td>
<td>0.0</td>
<td>0.5</td>
<td>1.5</td>
</tr>
<tr>
<td>OTU30(^b)</td>
<td>1.4</td>
<td>2.3</td>
<td>0.3</td>
</tr>
<tr>
<td>OTU31(^b)</td>
<td>2.1</td>
<td>0.4</td>
<td>0.0</td>
</tr>
<tr>
<td>OTU32(^b)</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>OTU33(^b)</td>
<td>0.0</td>
<td>0.0</td>
<td>1.9</td>
</tr>
<tr>
<td>OTU34(^b)</td>
<td>2.5</td>
<td>1.1</td>
<td>0.5</td>
</tr>
<tr>
<td>OTU35(^b)</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>OTU36(^b)</td>
<td>1.8</td>
<td>1.8</td>
<td>0.0</td>
</tr>
<tr>
<td>OTU37(^b)</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>OTU</td>
<td>1.0</td>
<td>1.4</td>
<td>0.0</td>
</tr>
<tr>
<td>--------</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
</tr>
<tr>
<td>OTU38^b</td>
<td>1.0</td>
<td>1.4</td>
<td>0.0</td>
</tr>
<tr>
<td>OTU39c</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>OTU40^b</td>
<td>0.4</td>
<td>1.4</td>
<td>0.0</td>
</tr>
<tr>
<td>OTU41^b</td>
<td>1.3</td>
<td>0.9</td>
<td>0.0</td>
</tr>
<tr>
<td>OTU42^c</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>OTU43^c</td>
<td>1.2</td>
<td>2.3</td>
<td>0.0</td>
</tr>
<tr>
<td>OTU44^c</td>
<td>0.0</td>
<td>0.6</td>
<td>0.0</td>
</tr>
<tr>
<td>OTU45^b</td>
<td>0.9</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>OTU46^b</td>
<td>0.0</td>
<td>0.4</td>
<td>0.0</td>
</tr>
<tr>
<td>OTU47^c</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>OTU48^c</td>
<td>0.7</td>
<td>0.6</td>
<td>0.0</td>
</tr>
<tr>
<td>OTU49c</td>
<td>1.3</td>
<td>3.0</td>
<td>0.0</td>
</tr>
<tr>
<td>OTU50^a</td>
<td>0.0</td>
<td>0.0</td>
<td>1.1</td>
</tr>
<tr>
<td>OTU51^b</td>
<td>0.0</td>
<td>0.7</td>
<td>0.7</td>
</tr>
<tr>
<td>OTU52^c</td>
<td>0.0</td>
<td>0.0</td>
<td>0.8</td>
</tr>
</tbody>
</table>

^a OTUs identified as *L. lactis* subspecies *lactis*

^b OTUs identified as *L. lactis* subspecies *cremoris*.

^c OTUs that could not be assigned to a subspecies.