The microbial diversity of mesophilic starter cultures used in cheese production

Den mikrobielle diversiteten i mesofile starterkulturer anvendt i osteproduksjon

Philosophiae Doctor (PhD) Thesis

Cyril Frantzen

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

Ås (2017)

Thesis number 2017:96
ISSN 1894-6402
ISBN 978-82-575-1483-9

Table of contents

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.

Sammendrag

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*.

I denne studien har bakterie- og bakteriofag-diversiteten i starterkulturer og meieriprøver fra tre ulike store norske ysterier blitt undersøkt ved hjelp av molekylære og DNA-sekvenseringsbaserte metoder. Bruk av et melkebasert vekstmedium (GMA) i tillegg til det tradisjonelle vekstmediet M17 var avgjørende for å øke kapasiteten til å isolere en større diversitet av bakterier fra starterkulturene, som igjen førte til et større potensial for å isolere bakteriofager fra meieriprøvene. Ved hjelp av fagtyping ble et stort antall forskjellige bakterier og bakteriofager diskriminert fra hverandre. Et interessant funn var at mange av stammene som kun vokste på det melkebaserte mediet var følsomme for bakteriofag som M17-stammene ikke var følsomme for. Et stort antall fenotypisk forskjellige starter bakterier med ulik fagfølsomhet ble hel-genom sekvensert og karakterisert ved hjelp av pan-genomiske analyser. Pan-genom analysene skilte bakteriene inn i 21 *Lactococcus lactis* subart *lactis*, 28 *Lactococcus lactis* subart *cremoris*, og 12 *Leuconostoc* spp. linjer. Analysen diskriminerte ikke *Leuconostoc mesenteroides* subart *mesenteriodes* fra

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 dominerte av *Leuconostoc mesenteroides* subsp. *cremoris* mens de to resterende kulturene var dominerte 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:

Frantzen C, Kleppen HP, Holo H. 2016. 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. International Dairy Journal **53:**45-50.

Paper II:

Frantzen CA, Kot W, Pedersen TB, Ardö YM, Broadbent JR, Neve H, Hansen LH, Dal Bello F, Østlie HM, Kleppen HP, Vogensen FK, Holo H. 2017. Genomic characterization of dairy associated *Leuconostoc* species and diversity of leuconostocs in undefined mixed mesophilic starter cultures. Frontiers in Microbiology **8**:132.

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:

1. **Porcellato D, Frantzen C, Rangberg A, Umu OC, Gabrielsen C, Nes IF, Amdam GV, Diep DB.** 2015. Draft genome sequence of *Lactobacillus kunkeei* AR114 Isolated from Honey Bee Gut. Genome Announcements **3**.

2. **Arbulu S, Frantzen C, Lohans CT, Cintas LM, Herranz C, Holo H, Diep DB, Vederas JC, Hernández PE.** 2016. Draft genome sequence of the bacteriocin-producing strain *Enterococcus faecium* M3K31, isolated from griffon vultures (*Gyps fulvus* subsp. *fulvus*). Genome Announcements **4:**e00055-00016.

3. **Arbulu S, Jimenez JJ, Borrero J, Sanchez J, Frantzen C, Herranz C, Nes IF, Cintas LM, Diep DB, Hernandez PE.** 2016. Draft genome sequence of the bacteriocinogenic strain *Enterococcus faecalis* DBH18, isolated from mallard ducks (*Anas platyrhynchos*). Genome Announcements **4**.

4. **Stamsas GA, Straume D, Ruud Winther A, Kjos M, Frantzen CA, Havarstein LS.** 2017. Identification of EloR (Spr1851) as a regulator of cell elongation in *Streptococcus pneumoniae*. Molecular Microbiology. doi:10.1111/mmi.13748.

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

[32]. 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 CO_2, 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].

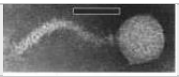
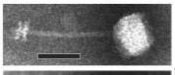
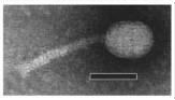
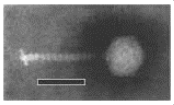
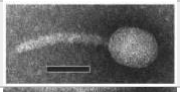
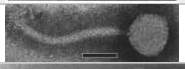
Phages may follow one of two life cycles, the lytic cycle or the lysogenic cycle. Phages with the ability for both are called temperate phages, while phages that can only reproduce by the lytic cycle are called virulent phages. In both life cycles, phage infection initiates with the interaction between 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]).

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

[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 sub-division 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.

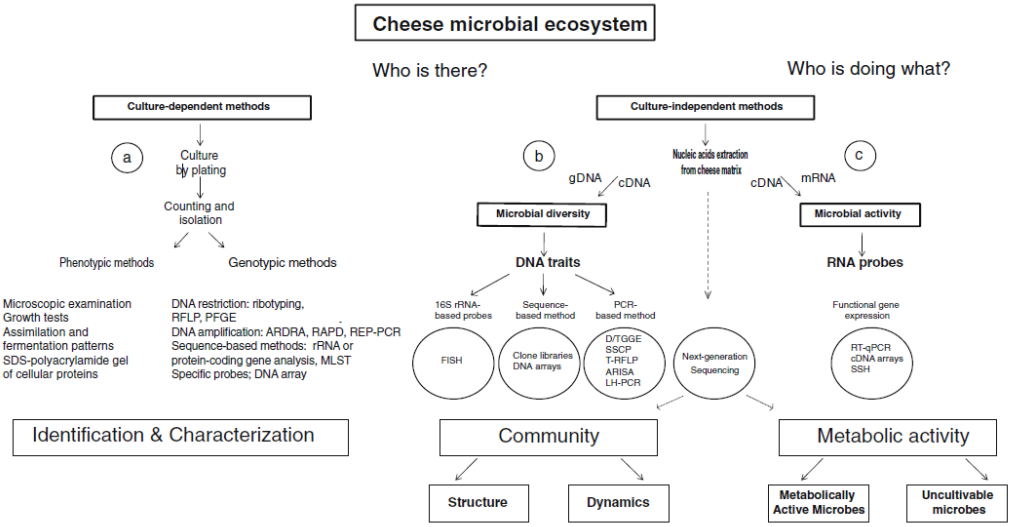


Figure 1. Overview of culture-dependent and culture-independent methodology, with examples of methods within each category. Adapted from [31].

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 fingerprinting 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]. Fingerprinting techniques refers to the genotyping methods that distinguish between samples based on non-sequence characteristics and/or pattern(s) of DNA. Commonly, fingerprinting 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.

PAPER II

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 dextransucrases. 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.

PAPER III

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 Biotechonology 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 peformed 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

1. **Evershed RP, Payne S, Sherratt AG, Copley MS, Coolidge J, Urem-Kotsu D, Kotsakis K, Ozdogan M, Ozdogan AE, Nieuwenhuyse O, Akkermans PMMG, Bailey D, Andeescu R-R, Campbell S, Farid S, Hodder I, Yalman N, Ozbasaran M, Bicakci E, Garfinkel Y, Levy T, Burton MM.** 2008. Earliest date for milk use in the Near East and southeastern Europe linked to cattle herding. Nature **455:**528-531.
2. **Stiles ME, Holzapfel WH.** 1997. Lactic acid bacteria of foods and their current taxonomy. International Journal of Food Microbiology **36:**1-29.
3. **Paul Ross R, Morgan S, Hill C.** 2002. Preservation and fermentation: past, present and future. International Journal of Food Microbiology **79:**3-16.
4. **Østlie HM, Kraggerud H, Longva AB, Abrahamsen RK.** 2016. Characterisation of the microflora during ripening of a Norwegian semi-hard cheese with adjunct culture of propionic acid bacteria. International Dairy Journal **54:**43-49.
5. **Martley FG, Crow VL.** 1993. Interactions between non-starter microorganisms during cheese manufacture and ripening. International Dairy Journal **3:**461-483.
6. **Crow VL, Coolbear T, Holland R, Pritchard GG, Martley FG.** 1993. Starters as finishers: Starter properties relevant to cheese ripening. International Dairy Journal **3:**423-460.
7. **El Soda M, Madkor SA, Tong PS.** 2000. Adjunct cultures: Recent developments and potential significance to the cheese industry. Journal of Dairy Science **83:**609-619.
8. **Conto F, Nobile MAD, Faccia M, Zambrini AV, Conte A.** 2017. Advances in Dairy Products. Wiley.
9. **Leroy F, De Vuyst L.** 2004. Lactic acid bacteria as functional starter cultures for the food fermentation industry. Trends in Food Science & Technology **15:**67-78.
10. **Spinnler HE, Gripon JC.** 2004. Surface mould-ripened cheeses, p 157-174. *In* Fox PF, McSweeney PLH, Cogan TM, Guinee TP (ed), Cheese: Chemistry, Physics and Microbiology, vol 2. Academic Press.
11. **Fox PF, Guinee TP, Cogan TM, McSweeney PLH.** 2016. Fundamentals of Cheese Science. Springer US.
12. **Bigret M, Mäyrä-Mäkinen A.** 2004. Industrial use and production of lactic acid bacteria, Lactic Acid Bacteria doi:10.1201/9780824752033.ch5. CRC Press.
13. **Fox PF, McSweeney PLH, Cogan TM, Guinee TP.** 2004. Cheese: Chemistry, Physics and Microbiology: General Aspects. Elsevier Science.
14. **Johnson ME.** 2013. Mesophilic and thermophilic cultures used in traditional cheesemaking. Microbiology Spectrum **1**.
15. **Champagne CP, Gagnon D, St-Gelais D, Vuillemard JC.** 2009. Interactions between *Lactococcus lactis* and *Streptococcus thermophilus* strains in Cheddar cheese processing conditions. International Dairy Journal **19:**669-674.
16. **Law BA, Tamime AY.** 2011. Technology of Cheesemaking. Wiley.
17. **Ravindra P.** 2015. Advances in Bioprocess Technology. Springer International Publishing.
18. **Parente E, Cogan TM.** 2004. Starter Cultures: General Aspects, p 123-147. *In* Patrick F. Fox PLHMTMC, Timothy PG (ed), Cheese: Chemistry, Physics and Microbiology, vol Volume 1. Academic Press.

19. **Smid EJ, Erkus O, Spus M, Wolkers-Rooijackers JC, Alexeeva S, Kleerebezem M.** 2014. Functional implications of the microbial community structure of undefined mesophilic starter cultures. Microbial Cell Factories **13:**S2.
20. **Erkus O, de Jager VC, Spus M, van Alen-Boerrigter IJ, van Rijswijck IM, Hazelwood L, Janssen PW, van Hijum SA, Kleerebezem M, Smid EJ.** 2013. Multifactorial diversity sustains microbial community stability. The ISME Journal **7**.
21. **Daly C.** 1983. The use of mesophilic cultures in the dairy industry. Antonie van Leeuwenhoek **49:**297-312.
22. **Ayad EHE, Verheul A, Wouters JTM, Smit G.** 2001. Population dynamics of lactococci from industrial, artisanal and non-dairy origins in defined strain starters for Gouda-type cheese. International Dairy Journal **11:**51-61.
23. **Vedamuthu ER.** 1994. The dairy *Leuconostoc*: Use in dairy products. Journal of Dairy Science **77:**2725-2737.
24. **Cavanagh D, Fitzgerald GF, McAuliffe O.** 2015. From field to fermentation: the origins of *Lactococcus lactis* and its domestication to the dairy environment. Food Microbiology **47:**45-61.
25. **Cavanagh D, Casey A, Altermann E, Cotter PD, Fitzgerald GF, McAuliffe O.** 2015. Evaluation of *Lactococcus lactis* Isolates from nondairy sources with potential dairy applications reveals extensive phenotype-genotype disparity and implications for a revised species. Applied and Environmental Microbiology **81:**3961-3972.
26. **Cock LS, Stouvenel ARd.** 2006. Lactic acid production by a strain of *Lactococcus lactis* subs lactis isolated from sugar cane plants. Electronic Journal of Biotechnology **9**.
27. **Bhanwar S, Bamnia M, Ghosh M, Ganguli A.** 2013. Use of *Lactococcus lactis* to enrich sourdough bread with gamma-aminobutyric acid. International Journal of Food Sciences and Nutrition **64:**77-81.
28. **Kelleher P, Bottacini F, Mahony J, Kilcawley KN, van Sinderen D.** 2017. Comparative and functional genomics of the *Lactococcus lactis* taxon; insights into evolution and niche adaptation. BMC Genomics **18:**267.
29. **Kütahya OE, Starrenburg MJC, Rademaker JLW, Klaassen CHW, van Hylckama Vlieg JET, Smid EJ, Kleerebezem M.** 2011. High-resolution amplified fragment length polymorphism typing of *Lactococcus lactis* strains enables identification of genetic markers for subspecies-related phenotypes. Applied and Environmental Microbiology **77:**5192-5198.
30. **Passerini D, Beltramo C, Coddeville M, Quentin Y, Ritzenthaler P, Daveran-Mingot M-L, Le Bourgeois P.** 2010. Genes but not genomes reveal bacterial domestication of *Lactococcus lactis*. PLOS ONE **5:**e15306.
31. **Ndoye B, Rasolofo EA, LaPointe G, Roy D.** 2011. A review of the molecular approaches to investigate the diversity and activity of cheese microbiota. Dairy Science & Technology **91:**495.
32. **Azizan KA, Baharum SN, Mohd Noor N.** 2012. Metabolic profiling of *Lactococcus lactis* under different culture conditions. Molecules **17:**8022-8036.
33. **Nomura M, Kobayashi M, Narita T, Kimoto-Nira H, Okamoto T.** 2006. Phenotypic and molecular characterization of *Lactococcus lactis* from milk and plants. Journal of Applied Microbiology **101:**396-405.
34. **Fernández E, Alegría Á, Delgado S, Martín MC, Mayo B.** 2011. Comparative phenotypic and molecular genetic profiling of wild *Lactococcus lactis* subsp. *lactis* 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.
35. **Kempler GM, McKay LL.** Biochemistry and genetics of citrate utilization in *Streptococcus lactis* ssp. *diacetylactis*. Journal of Dairy Science **64:**1527-1539.
36. **Verhue WM, Tjan FSB.** 1991. Study of the Citrate Metabolism of Lactococcus lactis subsp. lactis Biovar Diacetylactis by Means of (13)C Nuclear Magnetic Resonance. Applied and Environmental Microbiology **57:**3371-3377.
37. **Harvey RJ, Collins EB.** 1963. Roles of citrate and acetion in the metabolism of *Streptococcus diacetilactis*. Journal of Bacteriology **86:**1301-1307.
38. **Godon J-J, Delorme C, Ehrlich SD, Renault P.** 1992. Divergence of genomic sequences between *Lactococcus lactis* subsp. *lactis* and *Lactococcus lactis* subsp. *cremoris*. Applied and Environmental Microbiology **58:**4045-4047.
39. **Kelly WJ, Ward LJH, Leahy SC.** 2010. Chromosomal diversity in *Lactococcus lactis* and the origin of dairy starter cultures. Genome Biology and Evolution **2**.
40. **Bandell M, Lhotte ME, Marty-Teysset C, Veyrat A, Prévost H, Dartois V, Diviès C, Konings WN, Lolkema JS.** 1998. Mechanism of the citrate transporters in carbohydrate and citrate cometabolism in *Lactococcus* and *Leuconostoc* species. Applied and Environmental Microbiology **64:**1594-1600.
41. **Hache C, Cachon R, Wache Y, Belguendouz T, Riondet C, Deraedt A, Divies C.** 1999. Influence of lactose-citrate co-metabolism on the differences of growth and energetics in *Leuconostoc lactis*, *Leuconostoc mesenteroides* ssp. *mesenteroides* and *Leuconostoc mesenteroides* ssp. *cremoris*. Systematic and Applied Microbiology **22:**507-513.
42. **Ardo Y, Varming C.** 2010. Bacterial influence on characteristic flavour of cheeses made with mesophilic DL-starter. Australian Journal of Dairy Technology **65:**153-158.
43. **Porcellato D, Ostlie HM, Brede ME, Martinovic A, Skeie SB.** 2013. Dynamics of starter, adjunct non-starter lactic acid bacteria and propionic acid bacteria in low-fat and full-fat Dutch-type cheese. International Dairy Journal **33:**104-111.
44. **Hemme D, Foucaud-Scheunemann C.** 2004. *Leuconostoc*, characteristics, use in dairy technology and prospects in functional foods. International Dairy Journal **14**.
45. **Gu CT, Wang F, Li CY, Liu F, Huo GC.** 2012. *Leuconostoc mesenteroides* subsp. *suionicum* subsp. nov. International Journal of Systematic and Evolutionary Microbiology **62:**1548-1551.
46. **Cogan TM, Jordan KN.** 1994. Metabolism of *Leuconostoc* Bacteria. Journal of Dairy Science **77:**2704-2717.
47. **Thunell RK.** 1995. Taxonomy of the Leuconostocs. Journal of Dairy Science **78:**2514-2522.
48. **Lodics TA, Steenson LR.** 1990. Characterization of bacteriophages and bacteria indigenous to a mixed-strain cheese starter. Journal of Dairy Science **73:**2685-2696.
49. **Johansen E, Kibenich A.** 1992. Characterization of *Leuconostoc* isolates from commercial mixed strain mesophilic starter cultures. Journal of Dairy Science **75:**1186-1191.
50. **Olsen KN, Brockmann E, Molin S.** 2007. Quantification of *Leuconostoc* populations in mixed dairy starter cultures using fluorescence in situ hybridization. Journal of Applied Microbiology **103:**855-863.

51. **Kleppen HP, Nes IF, Holo H.** 2012. Characterization of a *Leuconostoc* bacteriophage infecting flavor producers of cheese starter cultures. Applied and Environmental Microbiology **78:**6769-6772.
52. **Pedersen TB, Kot WP, Hansen LH, Sorensen SJ, Broadbent JR, Vogensen FK, Ardo Y.** 2014. Genome Sequence of *Leuconostoc mesenteroides* subsp. *cremoris* Strain T26, isolated from mesophilic undefined cheese starter. Genome Announcements **2**.
53. **Pedersen TB, Kot WP, Hansen LH, Sorensen SJ, Broadbent JR, Vogensen FK, Ardo Y.** 2014. Genome sequences of two *Leuconostoc pseudomesenteroides* strains isolated from danish dairy starter cultures. Genome Announcements **2**.
54. **Garvie EI.** 1960. The genus *Leuconostoc* and its nomenclature. Journal of Dairy Research **27:**283-292.
55. **Garvie EI.** 1967. The growth factor and amino acid requirements of species of the genus *Leuconostoc*, including *Leuconostoc paramesenteroides* (sp.nov.) and *Leuconostoc oenos*. Microbiology **48:**439-447.
56. **Garvie EI.** 1969. Lactic dehydrogenases of strains of the genus *Leuconostoc*. Microbiology **58:**85-94.
57. **Garvie EI, Zezula V, Hill VA.** 1974. Guanine plus cytosine content of the deoxyribonucleic acid of the leuconostocs and some heterofermentative lactobacilli. International Journal of Systematic and Evolutionary Microbiology **24:**248-251.
58. **Garvie EI.** 1979. Proposal of Neotype Strains for *Leuconostoc mesenteroides* (Tsenkovskii) van Tieghem, *Leuconostoc dextranicum* (Beijerinck) Hucker and Pederson, and *Leuconostoc cremoris* (Knudsen and Sørensen) Garvie. International Journal of Systematic and Evolutionary Microbiology **29:**149-151.
59. **Garvie EI.** 1983. NOTES: *Leuconostoc mesenteroides* subsp. *cremoris* (Knudsen and Sørensen) comb. nov. and *Leuconostoc mesenteroides* subsp. *dextranicum* (Beijerinck) comb. nov. International Journal of Systematic and Evolutionary Microbiology **33:**118-119.
60. **Farrow JAE, Facklam RR, Collins MD.** 1989. Nucleic acid homologies of some vancomycin-resistant leuconostocs and description of *Leuconostoc citreum* sp. nov. and *Leuconostoc pseudomesenteroides* sp. nov. International Journal of Systematic and Evolutionary Microbiology **39:**279-283.
61. **Villani F, Moschetti G, Blaiotta G, Coppola S.** 1997. Characterization of strains of *Leuconostoc mesenteroides* by analysis of soluble whole-cell protein pattern, DNA fingerprinting and restriction of ribosomal DNA. Journal of Applied Microbiology **82**.
62. **Bjorkroth KJ, Geisen R, Schillinger U, Weiss N, De Vos P, Holzapfel WH, Korkeala HJ, Vandamme P.** 2000. Characterization of *Leuconostoc gasicomitatum* sp. nov., associated with spoiled raw tomato-marinated broiler meat strips packaged under modified-atmosphere conditions. Applied and Environmental Microbiology **66:**3764-3772.
63. **Cibik R, Lepage E, Talliez P.** 2000. Molecular diversity of *Leuconostoc mesenteroides* and *Leuconostoc citreum* isolated from traditional french cheeses as revealed by RAPD fingerprinting, 16S rDNA sequencing and 16S rDNA fragment amplification. Systematic and Applied Microbiology **23**.
64. **Perez G, Cardell E, Zarate V.** 2002. Random amplified polymorphic DNA analysis for differentiation of *Leuconostoc mesenteroides* subspecies isolated from Tenerife cheese. Letters in Applied Microbiology **34:**82-85.

65. **Sanchez JI, Martinez B, Rodriguez A.** 2005. Rational selection of *Leuconostoc* strains for mixed starters based on the physiological biodiversity found in raw milk fermentations. International Journal of Food Microbiology **105:**377-387.
66. **Vihavainen EJ, Björkroth KJ.** 2009. Diversity of *Leuconostoc gasicomitatum* associated with meat spoilage. International Journal of Food Microbiology **136**.
67. **Nieto-Arribas P, Seseña S, Poveda JM, Palop L, Cabezas L.** 2010. Genotypic and technological characterization of *Leuconostoc* isolates to be used as adjunct starters in Manchego cheese manufacture. Food Microbiology **27**.
68. **Alegria A, Delgado S, Florez AB, Mayo B.** 2013. Identification, typing, and functional characterization of *Leuconostoc* spp. strains from traditional, starter-free cheeses. Dairy Science & Technology **93:**657-673.
69. **Zeller-Péronnet V, Brockmann E, Pavlovic M, Timke M, Busch U, Huber I.** 2013. Potential and limitations of MALDI-TOF MS for discrimination within the species *Leuconostoc mesenteroides* and *Leuconostoc pseudomesenteroides*. Journal für Verbraucherschutz und Lebensmittelsicherheit **8:**205-214.
70. **Dan T, Liu W, Sun Z, Lv Q, Xu H, Song Y, Zhang H.** 2014. A novel multi-locus sequence typing (MLST) protocol for *Leuconostoc lactis* isolates from traditional dairy products in China and Mongolia. BMC Microbiology **14:**1-9.
71. **Zhang W, Liu W, Song Y, Xu H, Menghe B, Zhang H, Sun Z.** 2015. Multilocus sequence typing of a dairy-associated *Leuconostoc mesenteroides* population reveals clonal structure with intragenic homologous recombination. Journal of Dairy Science **98:**2284-2293.
72. **Barrangou R, Yoon SS, Breidt Jr F, Jr., Fleming HP, Klaenhammer TR.** 2002. Characterization of six *Leuconostoc fallax* bacteriophages isolated from an industrial sauerkraut fermentation. Applied and Environmental Microbiology **68:**5452-5458.
73. **Deveau H, Labrie SJ, Chopin M-C, Moineau S.** 2006. Biodiversity and classification of lactococcal phages. Applied and Environmental Microbiology **72:**4338-4346.
74. **Ackermann HW.** 2003. Bacteriophage observations and evolution. Research in Microbiology **154:**245-251.
75. **McCabe O, Spinelli S, Farenc C, Labbe M, Tremblay D, Blangy S, Oscarson S, Moineau S, Cambillau C.** 2015. The targeted recognition of *Lactococcus lactis* phages to their polysaccharide receptors. Molecular Microbiology **96**.
76. **Feiner R, Argov T, Rabinovich L, Sigal N, Borovok I, Herskovits AA.** 2015. A new perspective on lysogeny: prophages as active regulatory switches of bacteria. Nature Reviews Microbiology **13:**641-650.
77. **Fortier LC, Sekulovic O.** 2013. Importance of prophages to evolution and virulence of bacterial pathogens. Virulence **4:**354-365.
78. **Lunde M, Blatny JM, Kaper F, Nes IF, Lillehaug D.** 2000. The life cycles of the temperate lactococcal bacteriophage phiLC3 monitored by a quantitative PCR method. FEMS Microbiology letters **192:**119-124.
79. **Lood R, Collin M.** 2011. Characterization and genome sequencing of two *Propionibacterium acnes* phages displaying pseudolysogeny. BMC Genomics **12:**198-198.
80. **Ripp S, Miller RV.** 1997. The role of pseudolysogeny in bacteriophage-host interactions in a natural freshwater environment. Microbiology **143:**2065-2070.
81. **Łoś M, Węgrzyn G.** 2012. Chapter 9 - Pseudolysogeny, p 339-349. *In* Łobocka M, Szybalski WT (ed), Advances in Virus Research, vol 82. Academic Press.

82. **Cenens W, Makumi A, Mebrhatu MT, Lavigne R, Aertsen A.** 2013. Phage–host interactions during pseudolysogeny. Bacteriophage **3:**e25029.
83. **Whitehead HR, Cox GA.** 2009. 83. Observations on sudden changes in the rate of acid formation in milk by cultures of lactic streptococci. Journal of Dairy Research **5:**197-207.
84. **Whitehead HR, Cox GA.** 2009. 130. Bacteriophage phenomena in cultures of lactic streptococci. Journal of Dairy Research **7:**55-62.
85. **Rousseau GM, Moineau S.** 2009. Evolution of *Lactococcus lactis* phages within a cheese factory. Applied and Environmental Microbiology **75:**5336-5344.
86. **Kleppen HP, Bang T, Nes IF, Holo H.** 2011. Bacteriophages in milk fermentations: Diversity fluctuations of normal and failed fermentations. International Dairy Journal **21:**592-600.
87. **Mahony J, Cambillau C, van Sinderen D.** 2017. Host recognition by lactic acid bacterial phages. FEMS Microbiology reviews **41:**S16-s26.
88. **Mahony J, van Sinderen D.** 2014. Current taxonomy of phages infecting lactic acid bacteria. Frontiers in Microbiology **5:**7.
89. **Mahony J, Oliveira J, Collins B, Hanemaaijer L, Lugli GA, Neve H, Ventura M, Kouwen TR, Cambillau C, van Sinderen D.** 2017. Genetic and functional characterisation of the lactococcal P335 phage-host interactions. BMC Genomics **18:**146.
90. **Millen AM, Romero DA.** 2016. Genetic determinants of lactococcal C2viruses for host infection and their role in phage evolution. Journal of General Virology **97:**1998-2007.
91. **Murphy J, Bottacini F, Mahony J, Kelleher P, Neve H, Zomer A, Nauta A, van Sinderen D.** 2016. Comparative genomics and functional analysis of the 936 group of lactococcal *Siphoviridae* phages. Scientific Reports **6:**21345.
92. **Dupuis M-È, Moineau S.** 2010. Genome organization and characterization of the virulent lactococcal phage 1358 and its similarities to *Listeria* phages. Applied and Environmental Microbiology **76:**1623-1632.
93. **Fortier L-C, Bransi A, Moineau S.** 2006. Genome sequence and global gene expression of Q54, a new phage species linking the 936 and c2 phage species of *Lactococcus lactis*. Journal of Bacteriology **188:**6101-6114.
94. **Villion M, Chopin MC, Deveau H, Ehrlich SD, Moineau S, Chopin A.** 2009. P087, a lactococcal phage with a morphogenesis module similar to an *Enterococcus faecalis* prophage. Virology **388:**49-56.
95. **Samson JE, Moineau S.** 2010. Characterization of *Lactococcus lactis* phage 949 and comparison with other lactococcal phages. Applied and Environmental Microbiology **76:**6843-6852.
96. **Garneau JE, Tremblay DM, Moineau S.** 2008. Characterization of 1706, a virulent phage from *Lactococcus lactis* with similarities to prophages from other Firmicutes. Virology **373:**298-309.
97. **Kotsonis SE, Powell IB, Pillidge CJ, Limsowtin GKY, Hillier AJ, Davidson BE.** 2008. Characterization and genomic analysis of Phage asccφ28, a phage of the family *Podoviridae* infecting *Lactococcus lactis*. Applied and Environmental Microbiology **74:**3453-3460.
98. **Chopin A, Deveau H, Ehrlich SD, Moineau S, Chopin M-C.** 2007. KSY1, a lactococcal phage with a T7-like transcription. Virology **365:**1-9.

99. **Stuer-Lauridsen B, Janzen T, Schnabl J, Johansen E.** 2003. Identification of the host determinant of two prolate-headed phages infecting *lactococcus lactis*. Virology **309:**10-17.
100. **Tremblay DM, Tegoni M, Spinelli S, Campanacci V, Blangy S, Huyghe C, Desmyter A, Labrie S, Moineau S, Cambillau C.** 2006. Receptor-binding protein of *Lactococcus lactis* phages: Identification and characterization of the saccharide receptor-binding site. Journal of Bacteriology **188:**2400-2410.
101. **Dupont K, Janzen T, Vogensen FK, Josephsen J, Stuer-Lauridsen B.** 2004. Identification of *Lactococcus lactis* genes required for bacteriophage adsorption. Applied and Environmental Microbiology **70:**5825-5832.
102. **Mahony J, Kot W, Murphy J, Ainsworth S, Neve H, Hansen LH, Heller KJ, Sorensen SJ, Hammer K, Cambillau C.** 2013. Investigation of the relationship between lactococcal host cell wall polysaccharide genotype and 936 phage receptor binding protein phylogeny. Applied and Environmental Microbiology **79**.
103. **Ainsworth S, Sadovskaya I, Vinogradov E, Courtin P, Guerardel Y, Mahony J, Grard T, Cambillau C, Chapot-Chartier M-P, van Sinderen D.** 2014. Differences in lactococcal cell wall polysaccharide structure are major determining factors in bacteriophage sensitivity. mBio **5**.
104. **Valyasevi R, Sandine WE, Geller BL.** 1991. A membrane protein is required for bacteriophage c2 infection of *Lactococcus lactis* subsp. *lactis* C2. Journal of Bacteriology **173:**6095-6100.
105. **Monteville MR, Ardestani B, Geller BL.** 1994. Lactococcal bacteriophages require a host cell wall carbohydrate and a plasma membrane protein for adsorption and ejection of DNA. Applied and Environmental Microbiology **60:**3204-3211.
106. **Kot W, Neve H, Heller KJ, Vogensen FK.** 2014. Bacteriophages of *Leuconostoc*, *Oenococcus*, and *Weissella*. Frontiers in Microbiology **5:**186.
107. **Sozzi T, Poulin JM, Maret R, Pousaz R.** 1978. Isolation of a bacteriophage of *Leuconostoc mesenteroides* from dairy products. Journal of Applied Bacteriology **44:**159-161.
108. **Shin C, Sato Y.** 1979. Isolation of *Leuconostoc* bacteriophages from dairy products. Nihon Chikusan Gakkaiho **50:**419-422.
109. **Pedersen TB, Vogensen FK, Ardö Y.** 2016. Effect of heterofermentative lactic acid bacteria of DL-starters in initial ripening of semi-hard cheese. International Dairy Journal **57:**72-79.
110. **Kot W, Hansen LH, Neve H, Hammer K, Jacobsen S, Pedersen PD, Sørensen SJ, Heller KJ, Vogensen FK.** 2014. Sequence and comparative analysis of *Leuconostoc* dairy bacteriophages. International Journal of Food Microbiology **176:**29-37.
111. **Pujato SA, Guglielmotti DM, Martínez-García M, Quiberoni A, Mojica FJM.** 2017. *Leuconostoc mesenteroides* and *Leuconostoc pseudomesenteroides* bacteriophages: Genomics and cross-species host ranges. International Journal of Food Microbiology **257:**128-137.
112. **Pujato SA, Mercanti DJ, Guglielmotti DM, Rousseau GM, Moineau S, Reinheimer JA, Quiberoni AdL.** 2015. Phages of dairy *Leuconostoc mesenteroides*: Genomics and factors influencing their adsorption. International Journal of Food Microbiology **201:**58-65.

113. **Lu Z, Altermann E, Breidt F, Kozyavkin S.** 2010. Sequence analysis of *Leuconostoc mesenteroides* bacteriophage Phi1-A4 isolated from an industrial vegetable fermentation. Applied and Environmental Microbiology **76:**1955-1966.
114. **Kot W, Hammer K, Neve H, Vogensen FK.** 2013. Identification of the receptor-binding protein in lytic *Leuconostoc pseudomesenteroides* bacteriophages. Applied and Environmental Microbiology **79:**3311-3314.
115. **Jang SH, Hwang MH, Chang H-I.** 2010. Complete genome sequence of ΦMH1, a *Leuconostoc* temperate phage. Archives of Virology **155:**1883-1885.
116. **Ali Y, Kot W, Atamer Z, Hinrichs J, Vogensen FK, Heller KJ, Neve H.** 2013. Classification of lytic bacteriophages attacking dairy *Leuconostoc* starter strains. Applied and Environmental Microbiology **79:**3628-3636.
117. **Cocolin L, Ercolini D.** 2007. Molecular techniques in the microbial ecology of fermented foods. Springer New York.
118. **Randazzo CL, Caggia C, Neviani E.** 2009. Application of molecular approaches to study lactic acid bacteria in artisanal cheeses. Journal of Microbiological Methods **78:**1-9.
119. **Chakravorty S, Helb D, Burday M, Connell N, Alland D.** 2007. A detailed analysis of 16S ribosomal RNA gene segments for the diagnosis of pathogenic bacteria. Journal of microbiological methods **69:**330-339.
120. **Duthoit F, Godon J-J, Montel M-C.** 2003. Bacterial community dynamics during production of registered designation of origin salers cheese as evaluated by 16S rRNA gene single-strand conformation polymorphism analysis. Applied and Environmental Microbiology **69:**3840-3848.
121. **Klijn N, Weerkamp AH, de Vos WM.** 1991. Identification of mesophilic lactic acid bacteria by using polymerase chain reaction-amplified variable regions of 16S rRNA and specific DNA probes. Applied and Environmental Microbiology **57:**3390-3393.
122. **Palys T, Nakamura LK, Cohan FM.** 1997. Discovery and classification of ecological diversity in the bacterial world: the role of DNA sequence data. International Journal of Systematic Bacteriology **47:**1145-1156.
123. **Larsen MV, Cosentino S, Rasmussen S, Friis C, Hasman H, Marvig RL, Jelsbak L, Sicheritz-Pontén T, Ussery DW, Aarestrup FM, Lund O.** 2012. Multilocus sequence typing of total-genome-sequenced bacteria. Journal of Clinical Microbiology **50:**1355-1361.
124. **Xu H, Sun Z, Liu W, Yu J, Song Y, Lv Q, Zhang J, Shao Y, Menghe B, Zhang H.** 2014. Multilocus sequence typing of *Lactococcus lactis* from naturally fermented milk foods in ethnic minority areas of China. Journal of Dairy Science **97:**2633-2645.
125. **Carraro L, Maifreni M, Bartolomeoli I, Martino ME, Novelli E, Frigo F, Marino M, Cardazzo B.** 2011. Comparison of culture-dependent and -independent methods for bacterial community monitoring during Montasio cheese manufacturing. Research in Microbiology **162:**231-239.
126. **Hugenholtz P, Goebel BM, Pace NR.** 1998. Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. Journal of Bacteriology **180:**4765-4774.
127. **Head IM, Saunders JR, Pickup RW.** 1998. Microbial evolution, diversity, and ecology: A decade of ribosomal RNA analysis of uncultivated microorganisms. Microbial Ecology **35:**1-21.

128. **Nocker A, Burr M, Camper AK.** 2007. Genotypic microbial community profiling: a critical technical review. Microbial Ecology **54:**276-289.
129. **Ward DM, Weller R, Bateson MM.** 1990. 16S rRNA sequences reveal numerous uncultured microorganisms in a natural community. Nature **345:**63-65.
130. **Besnard V, Federighi M, Cappelier JM.** 2000. Development of a direct viable count procedure for the investigation of VBNC state in *Listeria monocytogenes*. Letters in Applied Microbiology **31:**77-81.
131. **Auty MA, Gardiner GE, McBrearty SJ, O'Sullivan EO, Mulvihill DM, Collins JK, Fitzgerald GF, Stanton C, Ross RP.** 2001. Direct in situ viability assessment of bacteria in probiotic dairy products using viability staining in conjunction with confocal scanning laser microscopy. Applied and Environmental Microbiology **67:**420-425.
132. **Vogensen F. K. KT, Larsen J. J., Kringelum B., Ellekjaer D., Waagner Nielsen E. .** 1987. Improved direct differentiation between *Leuconostoc cremoris*, *Streptococcus lactis* ssp. *diacetylactis*, and *Streptococcus cremoris*/*Streptococcus lactis* on agar. . Milchwissenschaft **42:**646-648.
133. **Ercolini D, Hill PJ, Dodd CER.** 2003. Development of a fluorescence in situ hybridization method for cheese using a 16S rRNA probe. Journal of Microbiological Methods **52:**267-271.
134. **Bottari B, Ercolini D, Gatti M, Neviani E.** 2006. Application of FISH technology for microbiological analysis: current state and prospects. Applied Microbiology and Biotechnology **73:**485-494.
135. **Ercolini D.** 2013. High-throughput sequencing and metagenomics: Moving forward in the culture-independent analysis of food microbial ecology. Applied and Environmental Microbiology **79:**3148-3155.
136. **Porcellato D, Skeie SB.** 2016. Bacterial dynamics and functional analysis of microbial metagenomes during ripening of Dutch-type cheese. International Dairy Journal **61:**182-188.
137. **Escobar-Zepeda A, Sanchez-Flores A, Quirasco Baruch M.** 2016. Metagenomic analysis of a Mexican ripened cheese reveals a unique complex microbiota. Food Microbiology **57:**116-127.
138. **Verreault D, Gendron L, Rousseau GM, Veillette M, Massé D, Lindsley WG, Moineau S, Duchaine C.** 2011. Detection of airborne lactococcal bacteriophages in cheese manufacturing plants. Applied and Environmental Microbiology **77:**491-497.
139. **Neve H, Berger A, Heller KJ.** 1995. A method for detecting and enumerating airbourne virulent bacteriophage of dairy starter cultures. Kieler Milchwirtschaftliche Forschungsberichte **47:**193-207.
140. **Boucher I, Moineau S.** 2001. Phages of *Lactococcus lactis*: An ecological and economical equilibrium. Recent Research and Developments in Virology **3:**243-256.
141. **Stern A, Sorek R.** 2011. The phage-host arms-race: Shaping the evolution of microbes. Bioessays **33:**43-51.
142. **Jany J-L, Barbier G.** 2008. Culture-independent methods for identifying microbial communities in cheese. Food Microbiology **25:**839-848.

Paper I

International Dairy Journal 53 (2016) 45–50

Contents lists available at ScienceDirect

International Dairy Journal

journal homepage: www.elsevier.com/locate/idairyj

Short communication

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

Cyril Frantzen [a, *], Hans Petter Kleppen [b], Helge Holo [a, c]

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

ARTICLE INFO

Article history:
Received 17 April 2015
Received in revised form
12 September 2015
Accepted 18 September 2015
Available online 24 October 2015

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 *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 *Lactococcus* ssp., and their bacteriophages.

1. Introduction

Mesophilic mixed (LD-type) starter cultures are used in the production of Dutch-type 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 infecting *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, Sørensen, & Derkx, 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

* Corresponding author. Tel.: +47 97650815.
E-mail address: cyril.frantzen@nmbu.no (C. Frantzen).

http://dx.doi.org/10.1016/j.idairyj.2015.09.005
0958-6946/

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 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).

2.3. Bacteriophage inhibition array

Bacterial isolates were cultivated in M17-c in 96-well microtitre plates and growth inhibition by dairy samples or bacteriophage was assayed essentially as described by Kleppen et al. (2011) with the following modifications; each well contained 150 μL M17-c and 50 μL dairy sample or phage suspension. Controls contained 200 μL M17-c. The microtitre plates were inoculated with arrays of lactococcal isolates using a stainless steel 48-pin replicator (Sigma–Aldrich, Munich, Germany), incubated over night at 22 °C and the OD at 620 nm was measured for each well using a SPECTROstar Nano microtiter plate-reader (BMG LABTECH, Ortenberg, Germany). All analyses were performed in triplicate. Isolates showing OD less than 50% of the control in at least one of the three replicates were considered inhibited.

2.4. Phage typing

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 microtitre 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 10^7 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 (2.1×10^9 versus 7×10^8 cfu mL^{-1}), while counts on the two media both gave 1.3×10^9 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 produced using culture A were collected over a period of two months. The bulk starter samples

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 (ϕ-M17) and 55 GMA-derived bacteriophage isolates (ϕ-GMA).

3.4. Low overlap in phage sensitivity between subpopulations

The ϕ-M17 and ϕ-GMA bacteriophage collections were pooled (ϕ-$M17_{pA}$ and ϕ-GMA_{pA}) 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 ϕ-$M17_{pA}$ but only seven were inhibited by ϕ-GMA_{pA}. On the other hand, only eight of the GMA isolates were inhibited by ϕ-$M17_{pA}$ while 55 were inhibited by ϕ-GMA_{pA}. 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 similar to the M17 collection and 23 (68/3) GMA strains should be sensitive to ϕ-$M17_{pA}$. The finding that only eight GMA isolates were sensitive to ϕ-$M17_{pA}$ thus shows that most M17 isolates do 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 (ϕ-$M17_{pB}$), and a mixture of 53 culture B-GMA-derived bacteriophages (ϕ-GMA_{pB}), 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 ϕ-M17 and 32 were sensitive to ϕ-GMA. Of the GMA isolates 36 were sensitive to ϕ-M17 while 53 were sensitive to ϕ-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 ϕ-M17 and ϕ-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 ϕ-M17 phages inhibited the acidification process and reduced ΔpH to 0.79 whilst the presence of ϕ-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

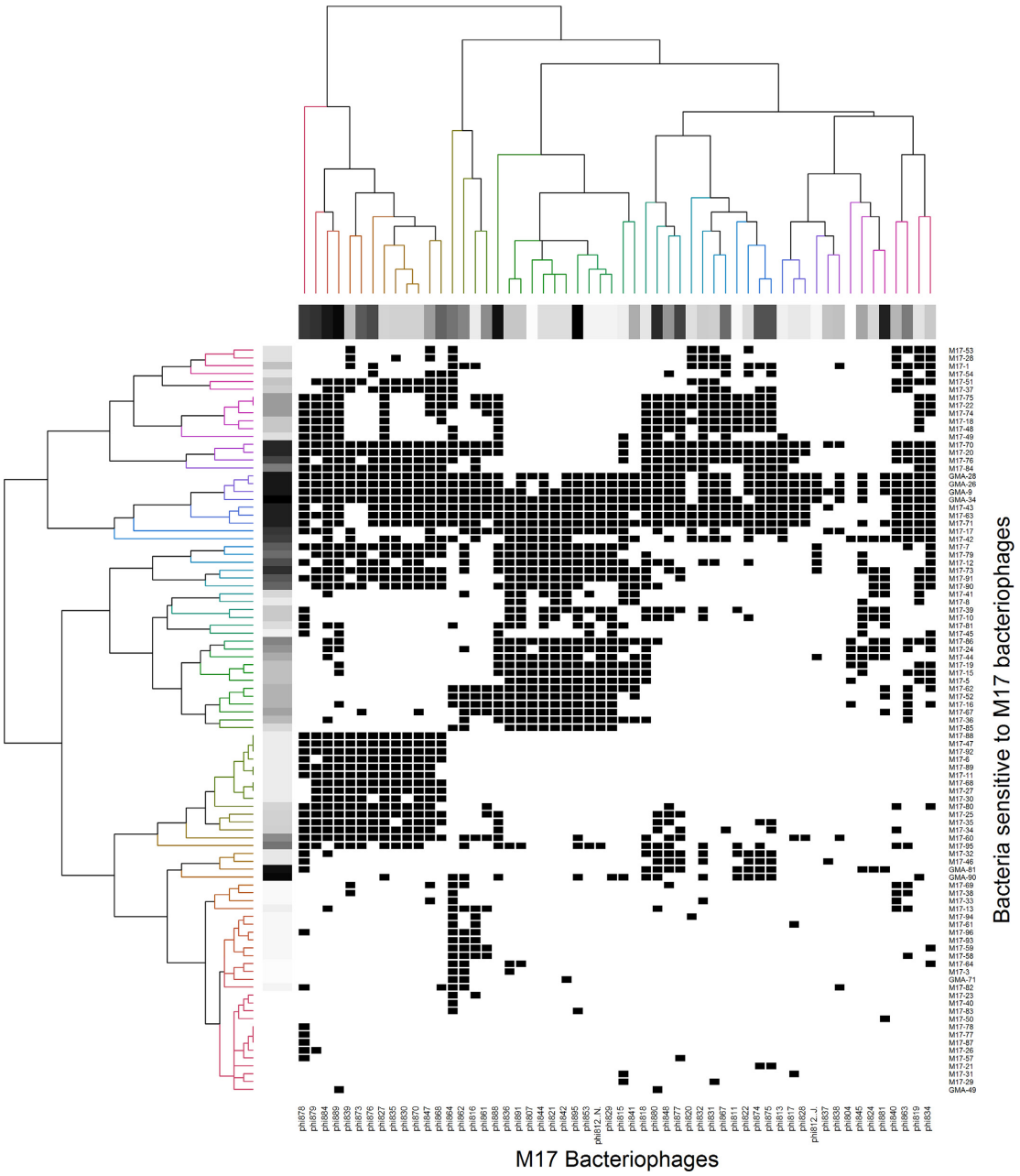


Fig. 1. Phage-host spectra for the M17 bacteriophages of culture A. All bacterial isolates inhibited by ϕ-M17 in the bacteriophage inhibition array were analysed by plaque-assays. Bacterial isolates not sensitive to any ϕ-M17 bacteriophages were omitted from the figure. The black squares indicate sensitivity 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.

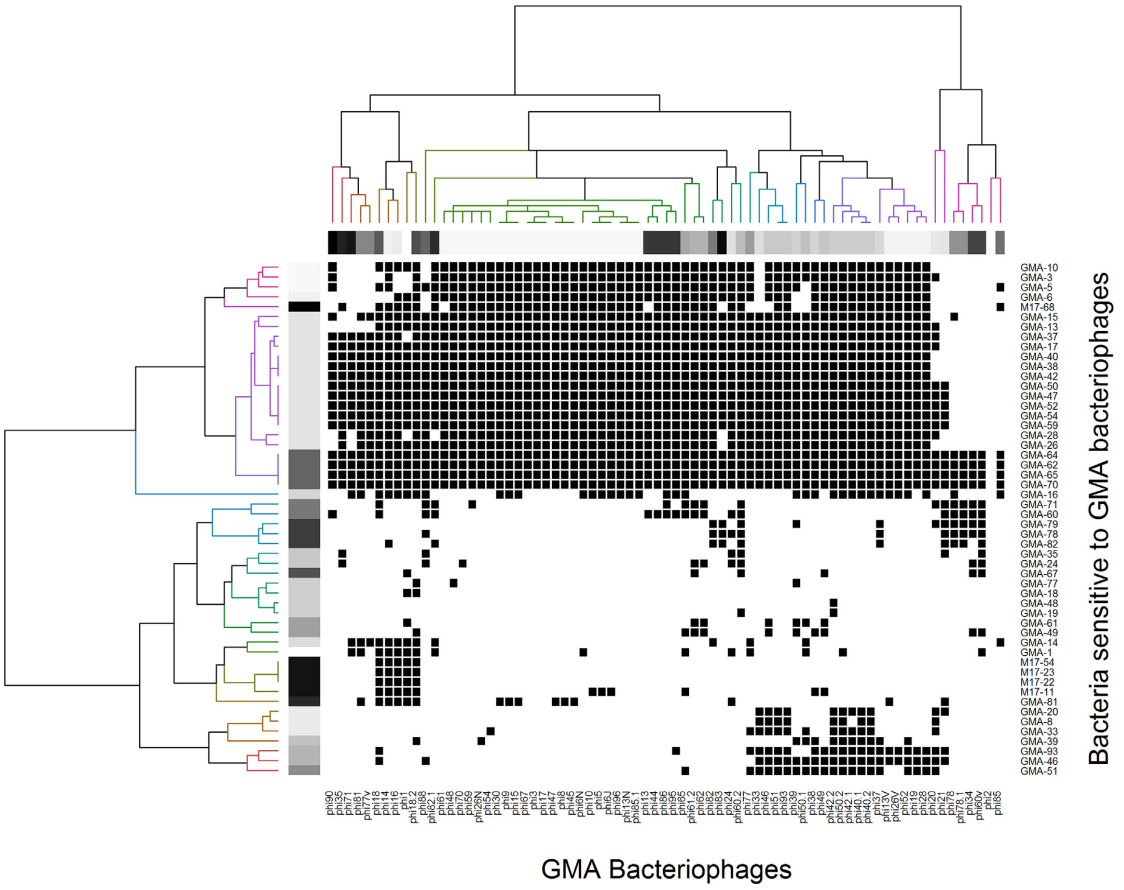


Fig. 2. Phage-host spectra for the GMA bacteriophages of culture A. All bacterial isolates inhibited by ϕ-GMA in the bacteriophage inhibition array were analysed by plaque-assays. Bacterial isolates not sensitive to any ϕ-GMA bacteriophages were omitted from the figure. The black squares indicate sensitivity 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 GMA bacteriophages produced 20 phage sensitivity profiles, and the corresponding GMA bacteriophages produced 35 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.

Table 1
Starter culture acidification activity.[a]

Sample	ΔpH (st.dev)
RSM	0.00
Culture A (no phages)	0.99 (±0.02)
Culture A + Dairy sample 1	0.71 (±0.00)
Culture A + Dairy sample 2	0.73 (±0.01)
Culture A + ϕ-$M17_{pA}$	0.79 (±0.03)
Culture A + ϕ-GMA_{pA}	0.80 (±0.02)
Culture A + ϕ-$M17_{pA}$ + ϕ-GMA_{pA}	0.67 (±0.03)

[a] Acidification activity tests were performed using culture A in the presence of bacteriophages; culture A with no phages was used as a standard. Dairy samples 1 and 2 were filtered bulk starter samples from reported slow fermentations. ϕ-$M17_{pA}$, pooled M17-derived phages; ϕ-GMA_{pA}, pooled GMA-derived phages.

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 ests 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 Högger 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

Bertani, G. (1951). Studies on lysogenesis. I. The mode of phage liberation by lysogenic Escherichia coli. *Journal of Bacteriology, 62*, 293–300.

Boucher, I., & Moineau, S. (2001). Phages of *Lactococcus lactis*: an ecological and economical equilibrium. *Recent Research and Developments in Virology, 3*, 243–256.

Deveau, H., Labrie, S. J., Chopin, M. C., & Moineau, S. (2006). Biodiversity and classification of lactococcal phages. *Applied and Environmental Microbiology, 72*, 4338–4346.

Erkus, O., de Jager, V. C., Spus, M., van Alen-Boerrigter, I. J., van Rijswijck, I. M., Hazelwood, L., et al. (2013). Multifactorial diversity sustains microbial community stability. *ISME Journal, 7*, 2126–2136.

Galili, T. (2015). *dendextend: Extending R's dendrogram functionality. R package version 0.18.3. [Computer software]*. Retrieved from http://CRAN.R-project.org/package=dendextend.

Hugenholtz, J., Splint, R., Konings, W. N., & Veldkamp, H. (1987). Selection of protease-positive and protease-negative variants of *Streptococcus cremoris*. *Applied and Environmental Microbiology, 53*, 309–314.

IDF. (2009). *ISO 26323:2009 IDF guideline — Determination of acidifying activity of dairy cultures*. Brussels, Belgium: International Dairy Federation.

Johansen, E., Øregaard, G., Sørensen, K. I., & Derkx, P. M. F. (2015). Modern approaches for isolation, selection, and improvement of bacterial strains for fermentation applications. In W. Holzapfel (Ed.), *Advances in fermented foods and beverages* (pp. 227–248). Cambridge, UK: Woodhead Publishing.

Kleppen, H. P., Bang, T., Nes, I. F., & Holo, H. (2011). Bacteriophages in milk fermentations: diversity fluctuations of normal and failed fermentations. *International Dairy Journal, 21*, 592–600.

Lillehaug, D. (1997). An improved plaque assay for poor plaque-producing temperate lactococcal bacteriophages. *Journal of Applied Microbiology, 83*, 85–90.

Mahony, J., Kot, W., Murphy, J., Ainsworth, S., Neve, H., Hansen, L. H., et al. (2013). Investigation of the relationship between lactococcal host cell wall polysaccharide genotype and 936 phage receptor binding protein phylogeny. *Applied and Environmental Microbiology, 79*, 4385–4392.

Moisan, M., & Moineau, S. (2012). Multilocus sequence typing scheme for the characterization of 936-like phages infecting *Lactococcus lactis*. *Applied and Environmental Microbiology, 78*, 4646–4653.

Nickels, C., & Leesment, H. (1964). Method for the differentiation and qualitative determination of starter bacteria. *Milchwissenschaft, 19*, 374–378.

Reddy, M. S., Vedamuthu, E. R., Washam, C. J., & Reinbold, G. W. (1972). Agar medium for differential enumeration of lactic streptococci. *Applied Microbiology, 24*, 947–952.

Rousseau, G. M., & Moineau, S. (2009). Evolution of *Lactococcus lactis* phages within a cheese factory. *Applied and Environmental Microbiology, 75*, 5336–5344.

RStudio. (2012). *RStudio: Integrated development environment for R (Version 0.98.1103) [Computer software]*. Retrieved from http://www.rstudio.org (Boston, MA).

Salvador, S., & Chan, P. (2004). Determining the number of clusters/segments in hierarchical clustering/segmentation algorithms. In *Proceedings of the 16th IEEE International Conference on Tools with Artificial Intelligence* (pp. 576–584). IEEE Computer Society.

Terzaghi, B. E., & Sandine, W. E. (1975). Improved medium for lactic streptococci and their bacteriophages. *Applied Microbiology, 29*, 807–813.

Warnes, G. R., Bolker, B., Bonebakker, L. R. G., Liaw, W. H. A., Lumley, T., Maechler, M., et al. (2015). *gplots: Various R programming tools for plotting data. R package version 2.16.0. [Computer software]*. Retrieved from http://CRAN.R-project.org/package=gplots.

Table S1

TIFN7-genotyping results.

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

Table S2

CWPS-genotyping results.

M17 isolates					GMA isolates				
Name	IL-KF	MG-SK	UC- CV	No-band	Name	IL-KF	MG-SK	UC- CV	No-band
M17-1		+			GMA-1	+			
M17-2	+				GMA-2		+		
M17-3		+			GMA-3	+			
M17-4				+	GMA-4	+			
M17-5		+			GMA-5	+			
M17-6	+				GMA-6	+			
M17-7	+				GMA-8		+		
M17-8		+			GMA-9	+			
M17-9	+				GMA-10	+			
M17-10	+				GMA-13	+			
M17-11	+				GMA-14	+			
M17-12		+			GMA-15	+			
M17-13	+				GMA-16	+			
M17-14	+				GMA-17	+	+		
M17-15	+				GMA-18		+		
M17-16		+			GMA-19		+		
M17-17		+			GMA-20		+		
M17-18	+				GMA-21		+		
M17-19		+			GMA-24	+			
M17-20		+			GMA-26	+			
M17-21	+				GMA-28	+			
M17-22	+				GMA-32	+	+		
M17-23				+	GMA-33		+		
M17-24	+				GMA-34		+		
M17-25	+				GMA-35		+		
M17-26		+			GMA-37	+			
M17-27	+				GMA-39				+
M17-28		+			GMA-41		+		
M17-29	+				GMA-44		+		
M17-30	+				GMA-46		+		
M17-31	+	+			GMA-47		+		
M17-32	+				GMA-48		+		
M17-33	+	+			GMA-49	+			
M17-34	+				GMA-51		+		
M17-35	+				GMA-52	+			
M17-36	+				GMA-60		+		
M17-37	+				GMA-61		+		
M17-38		+			GMA-64	+			
M17-39	+				GMA-65		+		
M17-40	+				GMA-67		+		
M17-41		+			GMA-70		+		
M17-42	+	+			GMA-71		+		
M17-43	+	+			GMA-77	+			
M17-44	+				GMA-78		+		
M17-45			+		GMA-79		+		
M17-46		+			GMA-81	+			
M17-47	+				GMA-82		+		
M17-48	+				GMA-93	+			
M17-49				+	GMA-95	+			
M17-50				+					
M17-51	+								
M17-52		+							
M17-53		+							
M17-54	+	+							

M17-55	+		
M17-56	+		
M17-57	+		
M17-58		+	
M17-59		+	
M17-60	+		
M17-61	+		
M17-62		+	
M17-63	+		
M17-64	+		
M17-65	+		
M17-66		+	
M17-67		+	
M17-68	+		
M17-69			
M17-70	+		
M17-71	+		
M17-72	+		
M17-73	+		
M17-74	+		
M17-75	+		
M17-76	+		
M17-77	+		
M17-78	+		
M17-79	+		
M17-80	+		
M17-81			+
M17-82		+	
M17-83	+	+	
M17-84	+		
M17-85		+	
M17-86		+	
M17-87		+	
M17-88	+		
M17-89	+		
M17-90	+		
M17-91	+		
M17-92	+		
M17-93	+		
M17-94		+	
M17-95	+		
M17-96		+	

Supplementary Figure 1: Phage-host spectra for the M17 and GMA bacteriophages

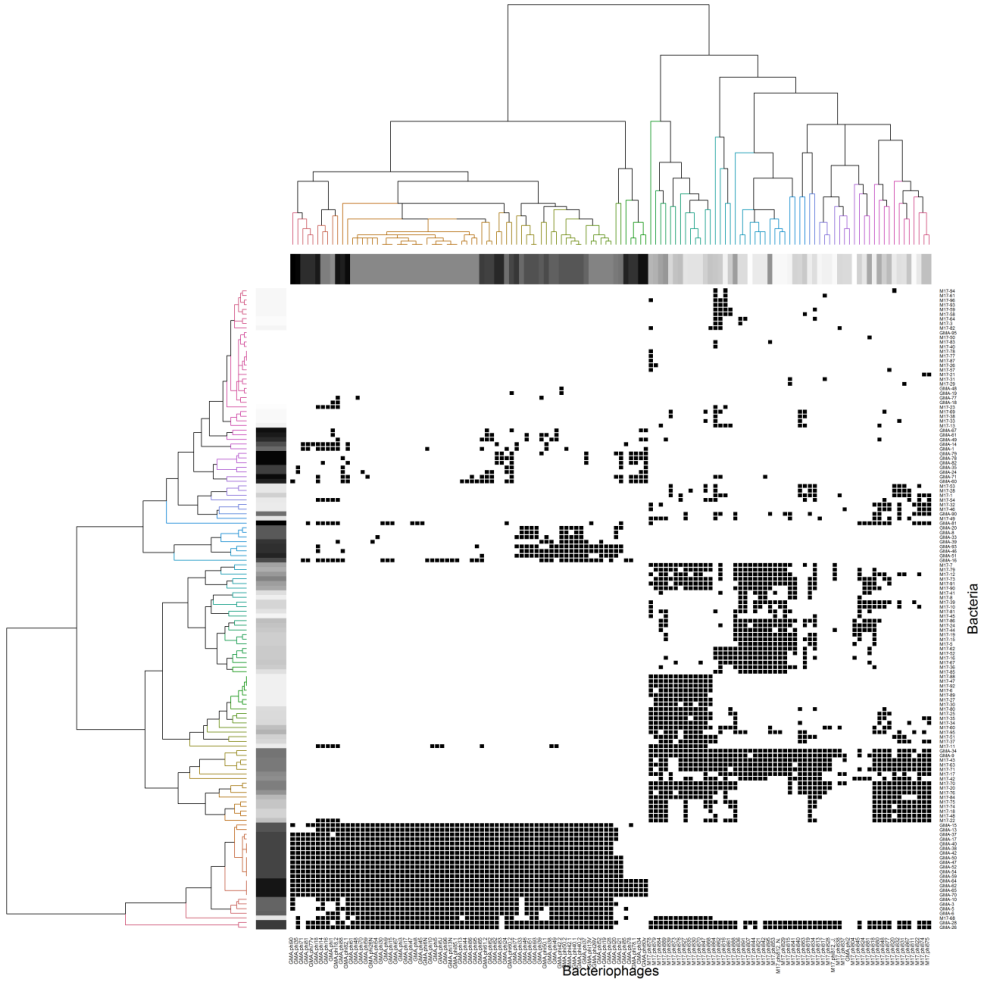


Fig. S1. Combined phage-host spectra for the M17 and GMA bacteriophages of culture A. All bacterial isolates inhibited in the bacteriophage inhibition array were analysed by plaque-assays. Bacterial isolates not sensitive to any bacteriophages were omitted from the figure. The black squares indicate 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

frontiers in Microbiology

ORIGINAL RESEARCH
published: 03 February 2017
doi: 10.3389/fmicb.2017.00132

Genomic Characterization of Dairy Associated *Leuconostoc* Species and Diversity of Leuconostocs in Undefined Mixed Mesophilic Starter Cultures

*Cyril A. Frantzen[1], Witold Kot[2], Thomas B. Pedersen[3], Ylva M. Ardö[3], Jeff R. Broadbent[4], Horst Neve[5], Lars H. Hansen[2], Fabio Dal Bello[6], Hilde M. Østlie[1], Hans P. Kleppen[1,7], Finn K. Vogensen[3] and Helge Holo[1,8]**

[1] Laboratory of Microbial Gene Technology and Food Microbiology, Department of Chemistry, Biotechnology and Food Science, Norwegian University of Life Sciences, Ås, Norway, [2] Department of Environmental Science, Aarhus University, Roskilde, Denmark, [3] Department of Food Science, University of Copenhagen, Copenhagen, Denmark, [4] Department of Nutrition, Dietetics and Food Sciences, Utah State University, Logan, UT, USA, [5] Department of Microbiology and Biotechnology, Max Rubner-Institut, Kiel, Germany, [6] Sacco Srl, Cordorago, Italy, [7] ACD Pharmaceuticals AS, Leknes, Norway, [8] TINE SA, Oslo, Norway

OPEN ACCESS

Edited by:
Sandra Torriani,
University of Verona, Italy

Reviewed by:
Beatriz Martínez,
Consejo Superior de Investigaciones Cientificas (CSIC), Spain
Daniel M. Linares,
Teagasc-The Irish Agriculture and Food Development Authority, Ireland

***Correspondence:**
Helge Holo
helge.holo@nmbu.no

Specialty section:
This article was submitted to Food Microbiology, a section of the journal Frontiers in Microbiology

Received: 17 November 2016
Accepted: 18 January 2017
Published: 03 February 2017

Citation:
Frantzen CA, Kot W, Pedersen TB, Ardö YM, Broadbent JR, Neve H, Hansen LH, Dal Bello F, Østlie HM, Kleppen HP, Vogensen FK and Holo H (2017) Genomic Characterization of Dairy Associated Leuconostoc Species and Diversity of Leuconostocs in Undefined Mixed Mesophilic Starter Cultures. Front. Microbiol. 8:132. doi: 10.3389/fmicb.2017.00132

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* subsp. *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* subspecies *cremoris*. *Leuconostoc pseudomesenteroides* dominated in two of the

cultures while *Leuconostoc lactis*, reported to be a major constituent in fermented dairy products, was only present in low amounts in one of the cultures. This is the first in-depth study of *Leuconostoc* genomics and diversity in dairy starter cultures. The results and the techniques presented may be of great value for the dairy industry.

Keywords: dairy, cheese, *leuconostoc*, comparative, genomics, diversity analysis, starter cultures, differentiation

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 CO_2, 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 CO_2 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 ripening 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 mesenteroides* 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 Kibenich, 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 $FeSO_4$, 0.2 g/L $MgSO_4$, 0.05 g/L $MnSO_4$, 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 (OD_{600}) 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 erronous 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/). Hierarchal 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-linage 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 OD_{600} 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′-AACACGAAGCTGTTGAATTGCGTG-3′), and Eno-R (5′-GCAAATCCACCTTCATCACCAACTGA-3′). Forward (5′-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG-) and

reverse (5′GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG-) 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 publically 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 hierarchal 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* subsp.) include significantly smaller genomes for *Ln. cremoris* and *Ln. lactis* (1.6–1.8 Mb) compared to *Ln. mesenteroides*, *Ln. dextranicum*,

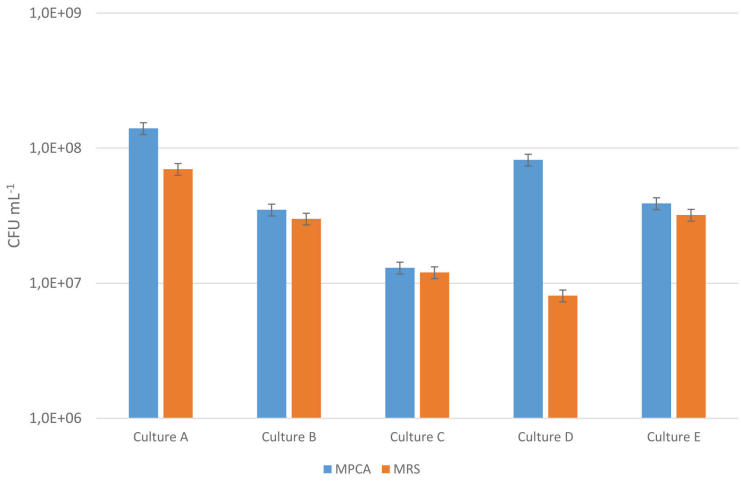


FIGURE 1 | Bacterial counts for five starter cultures A–E on MRS and MPCA supplemented with vancomycin to select for *Leuconostoc*. The counts are the mean of three separate extractions made from the same culture batch and the error bar indicates the standard deviation. The blue bars represent the bacterial 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.

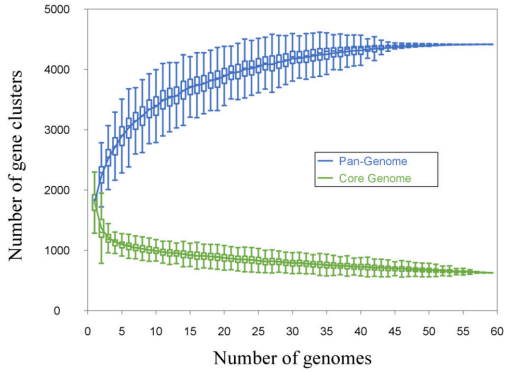


FIGURE 2 | Pan- and core-genome estimation. The estimation is made by including genomes one by one, matching the genetic content from each genome, with the growing pan- and the decreasing core-genome. Homologous genes are clustered together in orthologous groups. If the genomes included in the estimation are sufficiently distant from each other with regards to phylogeny, more than one orthologous group can exist for the same gene. The cut-off for this is set by the inflation value in the Markov Cluster Algorithm (MCL), for our dataset the inflation value was set to 1.5. The genetic content was curated for significantly divergent singletons, likely to be the product of erroneous assembly or annotation. The final pan-genome was estimated at 4415 orthologous groups, while the core-genome was estimated at 638 orthologous groups.

and *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 *Ln. cremoris* and *Ln. lactis*. Analysis of functional genomics indicated a closer relationship between *Ln. lactis* and *Ln. pseudomesenteroides*, than that of *Ln. mesenteroides*. Comparison of genetic potential within and between the *Ln. mesenteroides* subspecies showed only minor differences between *Ln. mesenteroides* and *Ln. dextranicum*. Rather, as shown in **Figure 3**, the variation between the isolates was much greater than the difference between *Ln. mesenteroides* and *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 *Ln. lactis*, where *Ln. lactis* 91922, isolated from kimchi was clearly distinguishable from LN19 and LN24 isolated from dairy. Comparison of *Ln. cremoris* and other *Ln. mesenteroides* subspecies isolates revealed that a range of genetic elements found in these species that were missing in *Ln. cremoris*. Apart from some enzymes encoding for rhamnose-containing glucans, *Ln. cremoris* isolates did not have any genetic functionality absent in *Ln. mesenteroides* or *Ln. dextranicum*. Moreover, several truncated genes and deletions were found in *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 *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) *Ln. cremoris* Lineages

Comparison of the genetic content for *Ln. cremoris* lineages showed that *Ln. cremoris* C1, C2, and C3 were highly similar and differentiated from each other mostly because of sequence variation in shared OGs. *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 *rmlA, rmlB, rmlC,* and *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) *Ln. mesenteroides* and *Ln. dextranicum* Lineages

Comparison of the genetic content showed a large variance between and within the *Ln. mesenteroides* lineages. Interestingly, no major difference between subspecies *Ln. mesenteroides* and *Ln. dextranicum* was found. *Ln. dextranicum* 20484 is grouped together with *Ln. mesenteroides* isolates LN32 and LN34, while *Ln. dextranicum* LbE16 is grouped together with *Ln. mesenteroides* LbE15 and LN08. This subspecies segregation of *Ln. dextranicum* and *Ln. mesenteroides* was based on the phenotypical ability to produce dextran from sucrose. Dextransucrase, 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 *Ln. mesenteroides* isolates included in this study, among them several genes encoding for dextransucrases with 40–67% amino acid identity to each other. Genotypically, the potential for dextran production exists within many if not all *Ln. mesenteroides* isolates, and does not differentiate *Ln. mesenteroides* from *Ln. dextranicum*. This finding was manifest by the separation of *Ln. mesenteroides* and *Ln. dextranicum* isolates into four lineages. 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*. In all of the strains in lineages M3 and M4, both the *cit* operon and the *lacLM* genes were present, while strains in lineages M1 and M2 were lacking the *cit* operon, and half of them also lacked the *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 LbT16 and LN05 also contained the deletion in the *lacZ* gene which is commonly identified in *Ln. cremoris* type strains. A genetic potential for proteolysis of casein (*prtP*) was identified in *Ln. mesenteroides* lineages M1 and M4, but not in M2 or M3.

(III) *Ln. lactis* Lineages

The pan-genomic differentiation grouped all the *Ln. lactis* isolates into one lineage. However, differences in genetic potential were found between the kimchi isolate *Ln. lactis* 91922 and

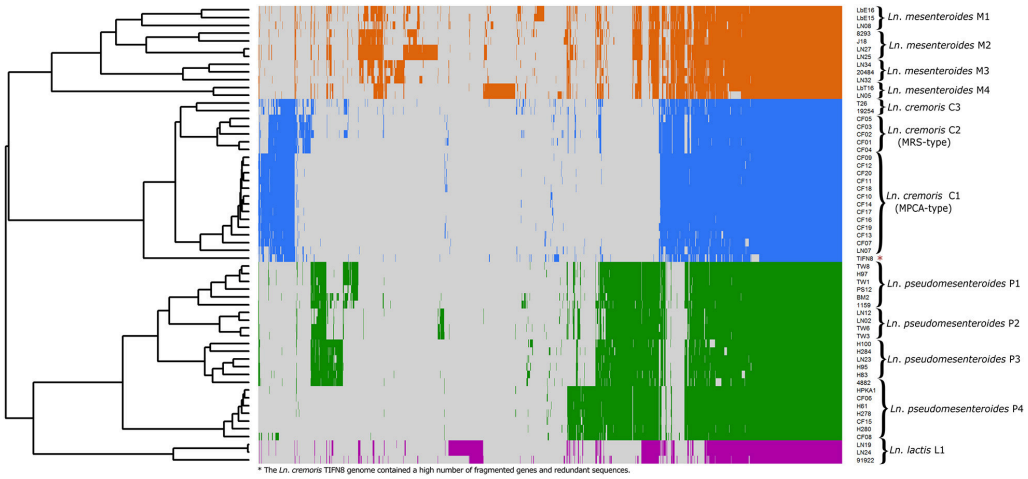


FIGURE 3 | Differentiation of 59 *Leuconostoc* genomes using the pan-genome of 4415 OGs. Hierarchal clustering of genomes clearly separated *Leuconostoc* species and subspecies. Moreover, the high sensitivity of the method produced twelve robust *Leuconostoc* lineages annotated on the right side of the figure. Four lineages of *Ln. mesenteroides* (colored orange), three lineages of *Ln. cremoris* (colored blue), four lineages of *Ln. pseudomesenteroides* (colored green), and one linage of *Ln. lactis* (colored purple) are shown. (*) The *Ln. cremoris* TIFN8 genome was excluded from further analysis because the genome data contained a high number of fragmented genes and redundant sequences. The heatmap was generated with R using the heatmap.2 function included in the Gplots package supplemented by the Dendextend package.

dairy isolates LN19 and LN24. *Ln. 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 *Ln. lactis* isolates.

(IV) *Ln. pseudomesenteroides* Lineages

Despite the significant pan-genomic differences and the sequence variation in shared OGs, the functional differences between lineages of *Ln. pseudomesenteroides* were surprisingly few. *Ln. pseudomesenteroides* P4 was different from the other three lineages with regards to genome synteny and genetic potential. Genetic functionality in the category of methionine biosynthesis, β-glucoside metabolism, sucrose metabolism, as well as an additional lactate dehydrogenase was identified in *Ln. 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 *Ln. cremoris* and *Ln. mesenteroides* subspecies carried the genetic potential to produce a wide range of amino acids while *Ln. lactis* and *Ln. 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 *Ln. 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 *Ln. 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. *Ln. 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

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

Profile name	Average genome size (Mb)	Average CDS
Ln. cremoris C1 *(MPCA-type)*	1.680 (±5)	1760 (±20)
Ln. cremoris C2 *(MRS-type)*	1.741 (±40)	1822 (±30)
Ln. cremoris C3	1.765 (±124)	1956 (±198)
Ln. mesenteroides M1	1.869 (±19)	1851 (±7)
Ln. mesenteroides M2	2.150 (±123)	2212 (±162)
Ln. mesenteroides M3	2.014 (±19)	2074 (±18)
Ln. mesenteroides M4	2.061 (±219)	2101 (±173)
Ln. pseudomesenteroides P1	2.028 (±47)	2081 (±61)
Ln. pseudomesenteroides P2	1.921 (±25)	1925 (±46)
Ln. pseudomesenteroides P3	2.063 (±44)	2133 (±60)
Ln. pseudomesenteroides P4	2.032 (±61)	2046 (±60)
Ln. lactis L1	1.718 (±26)	1700 (±43)

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

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

Amino acid pathway	*Ln. cremoris*	*Ln. mesenteroides*	*Ln. lactis*	*Ln. pseudomesenteroides*
Alanine	+	+	+	+
Arginine	+	+	+	+
Aspartate	+	+	+	+
Cysteine	+	+	–	+
Glutamine	–	–	+	+
Glutamic acid	–	–	–	–
Glycine	+	+	+	+
Histidine	+	+	–	–
Isoleucine	–	–	–	–
Leucine	–	–	–	–
Lysine	+	+	+	–
Methionine	+	+	–	–
Phenylalanine	+	+	+	+
Proline	+	+	+	+
Serine	+	+	+	+
Threonine	+	+	+	+
Tryptophan	+	+	–	–
Tyrosine	+	+	+	+
Valine	–	–	–	–

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

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 *Ln. lactis* and *Ln. pseudomesenteroides*. This could indicate a potential for synthesis of fructose-1,6-bisphosphate and glyceraldehyde-3-phosphate through fructose-1-phosphate, and hence homofermentative breakdown of fructose in *Ln. lactis* and *Ln. pseudomesenteroides*.

Comparative analysis of genes related to carbohydrate metabolism revealed big differences between the species (**Table 3**). All leuconostocs in this study encode beta-galactosidase, enabling utilization of lactose. Interestingly, the dairy *Ln. mesenteroides* have two different beta-galactosidases, *lacZ* and the plasmid-encoded *lacLM* (Obst et al., 1995), while the non-dairy isolates only contain *lacZ*. In *Ln. cremoris*, *lacZ* contains a large central deletion of 1200 bp between positions 740–1940. The *Ln. lactis* isolates only encode beta-galactosidase through *lacZ*, while the *Ln. pseudomesenteroides* isolates only encode beta-galactosidase through *lacLM*. In *Leuconostoc*, lactose is taken up by the lactose-specific transporter LacS, which couples lactose uptake to the secretion of galactose. LacS contains a C-terminal EIIAGlc-like domain and in *S. thermophilus* it has been shown that this domain can be phosphorylated, causing an increased lactose uptake rate (Gunnewijk and Poolman, 2000). All *Leuconostoc* isolates have this gene, but in *Ln. cremoris* *lacS* is truncated and lacks the C-terminal domain, possibly affecting lactose uptake and hence, growth rate on lactose. Alignment of all *lacS* sequences from this study revealed a close relationship between *Ln. pseudomesenteroides, Ln. lactis*, and *Ln. mesenteroides* isolates of non-dairy origin. In fact, *lacS* of non-dairy associated *Ln. mesenteroides* is more similar to the *lacS* from *Ln. lactis* and *Ln. pseudomesenteroides* (>75% identity) than that of dairy-associated *Ln. mesenteroides* or *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. pseudomesenteroides* 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.

	Ln. cremoris			*Ln. mesenteroides*				*Ln. pseudomesenteroides*				*Ln. lactis*
Gene(s)	**C1 ($n = 13$)**	**C2 ($n = 5$)**	**C3 ($n = 2$)**	**M1 ($n = 3$)**	**M2 ($n = 4$)**	**M3 ($n = 3$)**	**M4 ($n = 2$)**	**P1 ($n = 6$)**	**P2 ($n = 4$)**	**P3 ($n = 5$)**	**P4 ($n = 8$)**	**L1 ($n = 3$)**
araBAD	–	–	–	+	+	–	–	–	–	–	–	+(33%)
malP	–	–	–	#	+	+	–	+	+	+	+	+
malEFG	–	–	–	–	–	–	–	#	#	#	#	+
malX	–	–	–	–	–	–	–	–	–	–	–	+
malL	–	–	–	+	+	+	–	+	+	+	+	+
malR	–	–	–	+	+	+	–	+	+	+	+	+
lacL	+	+	+	+(66%)	+(50%)	+	+	+	+	+	+	–
lacM	+	+	+	+(66%)	+(50%)	+	+	+	+	+	+	–
lacZ	#	#	#	+	+	#	#	–	–	–	–	+
lacS	#	#	#	+	+	+	+	+	+	+	+	+
galEKT	+	+	+	+	+(75%)	+	+	+	+	+	+	+
manXYZ	+	+	+	+	+	+	+	+	+	+	+	+
manA	+	+	+	+	+	+	+	+	+	+	+	+
scrB	–	–	–	+	+	+	+	–	–	–	+	+
xylABG	#	#	#	+	+	#	#	+	+	+	+	+
treA	–	–	–	+	+	+	+	–	–	–	–	#(66%)
trePP	–	–	–	–	–	–	–	+	+	+	+	+(33%)
bglA	#	#	#	#	+	#	#	+	+	+	+	+
fruA	–	–	–	–	–	–	–	–	–	–	–	+
levE	–	–	–	–	+	+	+	+	+	+	+	–
frk	#	#	#	+	+	+	+	+	+	+	+	+
citCDEFGOS	+	+	+	+	+(50%)	–	+	+	+	+	+	+(66%)
fba	–	–	–	–	–	–	–	+	+	+	+	+

+, 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; malP, maltose phosphorylase; malEFG, maltose transport genes; malX, maltose/maltodextrin binding precursor; malL, sucrose-isomaltose; malR, maltose operon regulatory gene; lacL, beta-galactosidase, big subunit; lacM, beta-galactosidase, small subunit; lacZ, beta-galactosidase; lacS, lactose permease; galEKT, galactose metabolism; manXYZ, mannose transport genes; manA, mannose-6-phosphate isomerase; scrB, sucrose-6-phosphate hydrolase; xylABG, xylose isomerase, xylose kinase, xylose transport protein; treA, trehalose-6-phosphate hydrolase; trePP, trehalose-6-phosphate phosphorylase; bglA, 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 *cimH* 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 non-dairy 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

TABLE 4 | Genetic potential for proteolytic activity.

	Ln. cremoris			*Ln. mesenteroides*				*Ln. pseudomesenteroides*				*Ln. lactis*
Gene(s)	**C1 ($n = 13$)**	**C2 ($n = 5$)**	**C3 ($n = 2$)**	**M1 ($n = 3$)**	**M2 ($n = 4$)**	**M3 ($n = 3$)**	**M4 ($n = 2$)**	**P1 ($n = 6$)**	**P2 ($n = 4$)**	**P3 ($n = 5$)**	**P4 ($n = 8$)**	**L1 ($n = 3$)**
prtP	–	–	–	+(33%)	–	–	+	+	+	+	+	+(66%)
pepA	+	+	+	+	+	+	+	+	+	+	+	+
pepC	+	+	+	+	+	+	+	+	+	+	+	–
pepF	+	+	+	+	+	+	+	+	+	+	+	+
pepN	–	–	–	+	+	+	+	+	+	+	+	+
pepO	+	+	+	+	+	+	+	+	+	+	+	+
pepQ	+	+	+	+	+	+	+	+	+	+	+	+
pepS	+	+	+	+	+	+	+	+	+	+	+	+
pepT	+	+	+	+	+	+	+	+	+	+	+	+
pepV	–	–	+	–	–	–	–	+	+	+	+	–
pepX	#	#	#	+	+	+	+	+	+	+	+	–
oppABCDF	#	#	#	+	+	+	+	+	+	+	+	+

+, *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, aminopeptidase N; pepO, neutral endopeptidase; pepS, aminopeptidase; pepT, peptidase T; pepV, beta-ala-xaa dipeptidase; pepX, xaa-pro dipeptidyl-peptidase; oppABCDF, peptide ABC transporter operon.*

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 dipeptidyl aminopeptidase was truncated in *Ln. cremoris* (534 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-ala-dipeptidase. This dipeptidase has been shown to cleave dipeptides with an N-terminal β-Ala or D-ala 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 91922 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, D 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* C1) 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

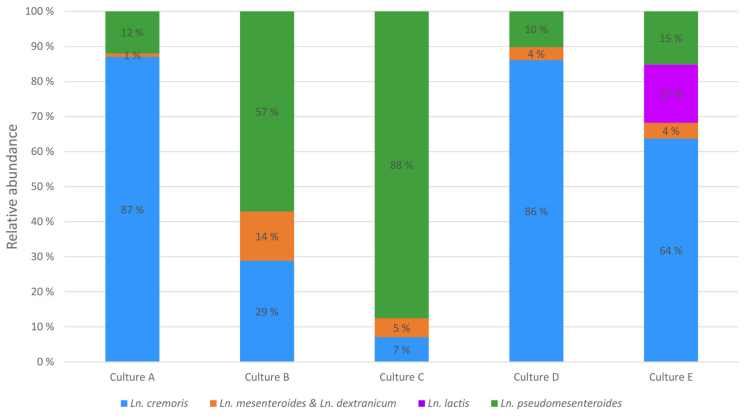


FIGURE 4 | Composition of leuconostocs in five starter cultures A–E using targeted-amplicon sequencing of the *eno* gene.

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-Teysset 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 *Ln. cremoris* and *Ln. pseudomesenteroides* was significantly different from the highly conserved and likely to be plasmid-encoded *cit* operon found in *Ln. lactis* and *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 *Leuconostoc* populations in cheeses starter cultures. Three of the starter cultures (A, D, and E) were dominated by *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 *Ln. pseudomesenteroides*. Interestingly, the cultures dominated by *Ln. cremoris* also contain *Ln. pseudomesenteroides* strains. *Ln. pseudomesenteroides* growth rates in pure culture are significantly higher than that of *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 *Ln. cremoris* and *Ln. lactis*, but in this study, *Ln. lactis* was detected only in culture E, which was dominated by *Ln. cremoris*. In two of the starter cultures studies in this work, *Ln. pseudomesenteroides* was the dominating *Leuconostoc*, which shows that they are highly relevant in the production of cheese. This is also reflected by recent studies, where the presence of *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 *Ln. pseudomesenteroides*.

The differences between the starter cultures could have an impact on the characteristics of the cheese product. *Ln. cremoris* lacks a wide range of genes involved in carbohydrate metabolism and proteolytic activity, and studies have shown that *Ln. cremoris* and *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 *Ln. pseudomesenteroides* was negligible. This was supported by our data, which showed that the *Ln. pseudomesenteroides* P4 isolates lack the genes necessary for reduction of diacetyl to acetoin and 2,3-butandiol. In addition, these isolates lacked the genes *ilvB* and *ilvH* encoding acetolactate synthetase large and small subunits, which is found in all *Ln. mesenteroides* subspecies isolates. However, a different gene *alsS*, encoding the same function, was found in all leuconostocs, including *Ln. pseudomesenteroides*. Studies on α-acetolactate synthase (ALS) and α-acetolactate decarboxylase (ALDC) activity in *Ln. mesenteroides* subspecies and *Ln. lactis* showed that the activity of both ALS and ALDC was higher for *Ln. lactis* (which does not have the *ilv* or *leu* operon) than that of *Ln. cremoris* (which does have part of these two operons) (Monnet et al., 1994). For comparison, the ALS activity of *Lc. lactis* biovar *diacetylactis* was comparable or in some cases even higher than that of *Ln. lactis*. *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 *Ln. pseudomesenteroides* compared to mock starters containing *Ln. cremoris* (Pedersen et al., 2016). This observation could be attributed to the rapid growth rate of *Ln. pseudomesenteroides* when compared to that of *Ln. cremoris*. The presence of the degenerated *ilv* and *leu* operons could somehow be negative to *Ln. cremoris* growth rate. Indeed, when cloning of the *ilv* operon into *Escherichia coli*, the presence of *Leuconostoc ilvB* was strongly detrimental to growth, while recombinant strains with an insertion in the *Leuconostoc ilvB* genes displayed normal growth. Their hypothesis was that expression of *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 *Ln. pseudomesenteroides* (Cardamone et al., 2011), and we identified genetic potential for caseinolytic activity in *Ln. pseudomesenteroides* in our data. This would enable *Ln. pseudomesenteroides* to grow better in milk than *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 *Ln. pseudomesenteroides* the better bacteria of the two based on these tidbits of information alone. However, both *Ln. cremoris* and *Ln. pseudomesenteroides* have shown to be significant to the production of cheeses. It is difficult to conclude which *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 derivates in products have been shown to diverge significantly, depending on which *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 addition to all amplicon sequencing), analyzed the data, wrote the R-scripts, devised the methods and wrote the manuscript. FB, HØ, TP, HK, and HN provided bacterial isolates for a larger diversity. WK and LH performed the sequencing of isolates in Denmark. Supervision of danish activities was provided by FV. Supervision of Norwegian activities was provided by HK, HØ, and HH. All co-authors were involved in reviewing and commenting on the manuscript prior to its submission. A large contribution to final editing was made by HN and JB.

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

REFERENCES

Alegria, A., Delgado, S., Florez, A. B., and Mayo, B. (2013). Identification, typing, and functional characterization of *Leuconostoc* spp. strains from traditional, starter-free cheeses. *Dairy Sci. Technol.* 93, 657–673. doi: 10.1007/s13594-013-0128-3

Ardö, Y., and Varming, C. (2010). Bacterial influence on characteristic flavour of cheeses made with mesophilic DL-starter. *Aust. J. Dairy Technol.* 65, 153–158.

Auty, M. A., Gardiner, G. E., McBrearty, S. J., O'Sullivan, E. O., Mulvihill, D. M., Collins, J. K., et al. (2001). Direct in situ viability assessment of bacteria in probiotic dairy products using viability staining in conjunction with confocal scanning laser microscopy. *Appl. Environ. Microbiol.* 67, 420–425. doi: 10.1128/AEM.67.1.420-425.2001

Bandell, M., Lhotte, M. E., Marty-Teysset, C., Veyrat, A., Prévost, H., Dartois, V., et al. (1998). Mechanism of the citrate transporters in carbohydrate and citrate cometabolism in *Lactococcus* and *Leuconostoc* species. *Appl. Environ. Microbiol.* 64, 1594–1600.

Barrangou, R., Yoon, S. S., Breidt, F. Jr., Fleming, H. P., and Klaenhammer, T. R. (2002). Characterization of six *Leuconostoc fallax* bacteriophages isolated from an industrial sauerkraut fermentation. *Appl. Environ. Microbiol.* 68, 5452–5458. doi: 10.1128/AEM.68.11.5452-5458.2002

Björkroth, K. J., Geisen, R., Schillinger, U., Weiss, N., De Vos, P., Holzapfel, W. H., et al. (2000). Characterization of *Leuconostoc gasicomitatum* sp. nov., associated with spoiled raw tomato-marinated broiler meat strips packaged under modified-atmosphere conditions. *Appl. Environ. Microbiol.* 66, 3764–3772. doi: 10.1128/AEM.66.9.3764-3772.2000

Callon, C., Millet, L., and Montel, M. C. (2004). Diversity of lactic acid bacteria isolated from AOC Salers cheese. *J. Dairy Res.* 71, 231–244. doi: 10.1017/S0022029904000159

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., et al. (2009). BLAST+: architecture and applications. *BMC Bioinformatics* 10:421. doi: 10.1186/1471-2105-10-421

Campedelli, I., Flórez, A. B., Salvetti, E., Delgado, S., Orrù, L., Cattivelli, L., et al. (2015). Draft genome sequence of three antibiotic-resistant *Leuconostoc mesenteroides* strains of dairy origin. *Genome Announc.* 3:e01018-15. doi: 10.1128/genomeA.01018-15

Cardamone, L., Quiberoni, A., Mercanti, D. J., Fornasari, M. E., Reinheimer, J. A., and Guglielmotti, D. M. (2011). Adventitious dairy *Leuconostoc* strains with interesting technological and biological properties useful for adjunct starters. *Dairy Sci. Technol.* 91, 457–470. doi: 10.1007/s13594-011-0022-9

Cavin, J. F., Dartois, V., Labarre, C., and Diviès, C. (1999). Cloning of branched chain amino acid biosynthesis genes and assays of α-acetolactate synthase activities in *Leuconostoc mesenteroides* subsp. *cremoris*. *Res. Microbiol.* 150, 189–198. doi: 10.1016/S0923-2508(99)80035-7

Cibik, R., Lepage, E., and Talliez, P. (2000). Molecular diversity of *Leuconostoc mesenteroides* and *Leuconostoc citreum* isolated from traditional french cheeses as revealed by RAPD fingerprinting, 16S rDNA sequencing and 16S rDNA fragment amplification. *Syst. Appl. Microbiol.* 23, 267–278. doi: 10.1016/S0723-2020(00)80014-4

Cogan, T. M., and Jordan, K. N. (1994). Metabolism of *Leuconostoc* bacteria. *J. Dairy Sci.* 77, 2704–2717. doi: 10.3168/jds.S0022-0302(94)77213-1

Conesa, A., Götz, S., García-Gómez, J. M., Terol, J., Talón, M., and Robles, M. (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21, 3674–3676. doi: 10.1093/bioinformatics/bti610

Contreras-Moreira, B., and Vinuesa, P. (2013). GET_HOMOLOGUES, a versatile software package for scalable and robust microbial pangenome analysis. *Appl. Environ. Microbiol.* 79, 7696–7701. doi: 10.1128/AEM.02411-13

Dan, T., Liu, W., Sun, Z., Lv, Q., Xu, H., Song, Y., et al. (2014). A novel multilocus sequence typing (MLST) protocol for *Leuconostoc lactis* isolates from traditional dairy products in China and Mongolia. *BMC Microbiol.* 14:150. doi: 10.1186/1471-2180-14-150

Drider, D., Bekal, S., and Prévost, H. (2004). Genetic organization and expression of citrate permease in lactic acid bacteria. *Genet. Mol. Res.* 3, 273–281.

Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26, 2460–2461. doi: 10.1093/bioinformatics/btq461

Edgar, R. C. (2013). UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nat. Methods* 10, 996–998. doi: 10.1038/nmeth.2604

Enright, A. J., Van Dongen, S., and Ouzounis, C. A. (2002). An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res.* 30, 1575–1584. doi: 10.1093/nar/30.7.1575

Erkus, O., de Jager, V. C., Spus, M., van Alen-Boerrigter, I. J., van Rijswijck, I. M., Hazelwood, L., et al. (2013). Multifactorial diversity sustains microbial community stability. *ISME J.* 7, 2126–2136. doi: 10.1038/ismej.2013.108

Farrow, J. A. E., Facklam, R. R., and Collins, M. D. (1989). Nucleic acid homologies of some vancomycin-resistant leuconostocs and description of *Leuconostoc citreum* sp. nov. and *Leuconostoc pseudomesenteroides* sp. nov. *Int. J. Syst. Evol. Microbiol.* 39, 279–283.

Galili, T. (2015). *Dendextend: Extending R's dendrogram functionality*. R package version 0.18.3. [Computer software]. Available online at: http://CRAN.R-project.org/package=dendextend

Garvie, E. I. (1960). The genus *Leuconostoc* and its nomenclature. *J. Dairy Res.* 27, 283–292. doi: 10.1017/S0022029900010359

Garvie, E. I. (1967). The growth factor and amino acid requirements of species of the genus *Leuconostoc*, including *Leuconostoc paramesenteroides* (sp.nov.) and *Leuconostoc* oenos. *Microbiology* 48, 439–447. doi: 10.1099/00221287-48-3-439

Garvie, E. I. (1969). Lactic dehydrogenases of strains of the genus *Leuconostoc*. *Microbiology* 58, 85–94. doi: 10.1099/00221287-58-1-85

Garvie, E. I. (1979). Proposal of neotype strains for *Leuconostoc mesenteroides* (Tsenkovskii) van Tieghem, *Leuconostoc dextranicum* (Beijerinck) Hucker and Pederson, and *Leuconostoc cremoris* (Knudsen and Sørensen) Garvie. *Int. J. Syst. Evol. Microbiol.* 29, 149–151.

Garvie, E. I. (1983). NOTES: *Leuconostoc mesenteroides* subsp. *cremoris* (Knudsen and Sørensen) comb. nov. and *Leuconostoc mesenteroides* subsp. *dextranicum* (Beijerinck) comb. nov. *Int. J. Syst. Evol. Microbiol.* 33, 118–119.

Garvie, E. I., Zezula, V., and Hill, V. A. (1974). Guanine plus cytosine content of the deoxyribonucleic acid of the *leuconostocs* and some heterofermentative lactobacilli. *Int. J. Syst. Evol. Microbiol.* 24, 248–251. doi: 10.1099/00207713-24-2-248

Godon, J. J., Delorme, C., Bardowski, J., Chopin, M. C., Ehrlich, S. D., and Renault, P. (1993). Gene inactivation in Lactococcus lactis: branched-chain amino acid biosynthesis. *J. Bacteriol.* 175, 4383–4390. doi: 10.1128/jb.175.14.4383-4390.1993

Grissa, I., Vergnaud, G., and Pourcel, C. (2007). CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Res.* 35, W52–W57. doi: 10.1093/nar/gkm360

Gu, C. T., Wang, F., Li, C. Y., Liu, F., and Huo, G. C. (2012). *Leuconostoc mesenteroides* subsp. *suionicum* subsp. nov. *Int. J. Syst. Evol. Microbiol.* 62, 1548–1551. doi: 10.1099/ijs.0.031203-0

Gunnewijk, M. G., and Poolman, B. (2000). HPr(His approximately P)-mediated phosphorylation differently affects counterflow and proton motive force-driven uptake via the lactose transport protein of *Streptococcus thermophilus*. *J. Biol. Chem.* 275, 34080–34085. doi: 10.1074/jbc.M003513200

Hache, C., Cachon, R., Wache, Y., Belguendouz, T., Riondet, C., Deraedt, A., et al. (1999). Influence of lactose-citrate co-metabolism on the differences of growth and energetics in *Leuconostoc lactis*, *Leuconostoc mesenteroides* ssp. *mesenteroides* and *Leuconostoc mesenteroides* ssp. *cremoris*. *Syst. Appl. Microbiol.* 22, 507–513. doi: 10.1016/S0723-2020(99)80002-2

Hemme, D., and Foucaud-Scheunemann, C. (2004). *Leuconostoc*, characteristics, use in dairy technology and prospects in functional foods. *Int. Dairy J.* 14, 467–494. doi: 10.1016/j.idairyj.2003.10.005

Hugenholtz, J. (1993). Citrate metabolism in lactic acid bacteria. *FEMS Microbiol. Rev.* 12, 165–178. doi: 10.1111/j.1574-6976.1993.tb00017.x

Johansen, E., and Kibenich, A. (1992). Characterization of *Leuconostoc* isolates from commercial mixed strain mesophilic starter cultures. *J. Dairy Sci.* 75, 1186–1191. doi: 10.3168/jds.S0022-0302(92)77865-5

Jung, J. Y., Lee, S. H., Lee, S. H., and Jeon, C. O. (2012). Complete genome sequence of *Leuconostoc mesenteroides* subsp. *mesenteroides* strain J18, isolated from kimchi. *J. Bacteriol.* 194, 730–731. doi: 10.1128/JB.06498-11

Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28, 27–30. doi: 10.1093/nar/28.1.27

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., et al. (2012). Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28, 1647–1649. doi: 10.1093/bioinformatics/bts199

Kleppen, H. P., Nes, I. F., and Holo, H. (2012). Characterization of a *Leuconostoc* bacteriophage Infecting flavor producers of cheese starter cultures. *Appl. Environ. Microbiol.* 78, 6769–6772. doi: 10.1128/AEM.00562-12

Lin, J., Schmitt, P., and Diviès, C. (1991). Characterization of a citrate-negative mutant of *Leuconostoc mesenteroides* subsp. *mesenteroides*: metabolic and plasmidic properties. *Appl. Microbiol. Biotechnol.* 34, 628–631.

Lodics, T. A., and Steenson, L. R. (1990). Characterization of bacteriophages and bacteria indigenous to a mixed-strain cheese starter. *J. Dairy Sci.* 73, 2685–2696. doi: 10.3168/jds.S0022-0302(90)78953-9

Magni, C., de Felipe, F. L., Sesma, F., López, P., and de Mendoza, D. (1994). Citrate transport in *Lactococcus lactis* biovar *diacetylactis*: expression of the plasmid-borne citrate permease p. *FEMS Microbiol. Lett.* 118, 75–82. doi: 10.1111/j.1574-6968.1994.tb06806.x

Marty-Teysset, C., Posthuma, C., Lolkema, J. S., Schmitt, P., Divies, C., and Konings, W. N. (1996). Proton motive force generation by citrolactic fermentation in *Leuconostoc mesenteroides*. *J. Bacteriol.* 178, 2178–2185. doi: 10.1128/jb.178.8.2178-2185.1996

Meslier, V., Loux, V., and Renault, P. (2012). Genome sequence of *Leuconostoc pseudomesenteroides* strain 4882, isolated from a dairy starter culture. *J. Bacteriol.* 194, 6637. doi: 10.1128/jb.01696-12

Monnet, C., Phalip, V., Schmitt, P., and Diviès, C. (1994). Comparison of α-acetolactate synthase and α-acetolactate decarboxylase in *Lactococcus* spp. and *Leuconostoc* spp. *Biotechnol. Lett.* 16, 257–262. doi: 10.1007/BF00134622

Nieto-Arribas, P., Seseña, S., Poveda, J. M., Palop, L., and Cabezas, L. (2010). Genotypic and technological characterization of *Leuconostoc* isolates to be used as adjunct starters in Manchego cheese manufacture. *Food Microbiol.* 27, 85–93. doi: 10.1016/j.fm.2009.08.006

Nurk, S., Bankevich, A., Antipov, D., Gurevich, A. A., Korobeynikov, A., Lapidus, A., et al. (2013). Assembling single-cell genomes and mini-metagenomes from chimeric MDA products. *J. Comput. Biol.* 20, 714–737. doi: 10.1089/cmb.2013.0084

Obst, M., Meding, E. R., Vogel, R. F., and Hammes, W. P. (1995). Two genes encoding the β-galactosidase of *Lactobacillus sakei*. *Microbiology* 141(Pt 12), 3059–3066. doi: 10.1099/13500872-141-12-3059

Olsen, K. N., Brockmann, E., and Molin, S. (2007). Quantification of *Leuconostoc* populations in mixed dairy starter cultures using fluorescence in situ hybridization. *J. Appl. Microbiol.* 103, 855–863. doi: 10.1111/j.1365-2672.2007.03298.x

Østlie, H. M., Kraggerud, H., Longva, A. B., and Abrahamsen, R. K. (2016). Characterisation of the microflora during ripening of a Norwegian semi-hard cheese with adjunct culture of propionic acid bacteria. *Int. Dairy J.* 54, 43–49. doi: 10.1016/j.idairyj.2015.10.005

Overbeek, R., Olson, R., Pusch, G. D., Olsen, G. J., Davis, J. J., Disz, T., et al. (2014). The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Res.* 42, D206–D214. doi: 10.1093/nar/gkt1226

Pedersen, T. B., Kot, W. P., Hansen, L. H., Sørensen, S. J., Broadbent, J. R., Vogensen, F. K., et al. (2014a). Genome sequence of *Leuconostoc mesenteroides* subsp. *cremoris* Strain T26, isolated from mesophilic undefined cheese starter. *Genome Announc.* 2:e00485-14. doi: 10.1128/genomeA.00485-14

Pedersen, T. B., Kot, W. P., Hansen, L. H., Sørensen, S. J., Broadbent, J. R., Vogensen, F. K., et al. (2014b). Genome sequences of two *Leuconostoc pseudomesenteroides* strains isolated from danish dairy starter cultures. *Genome Announc.* 2:e00484-14. doi: 10.1128/genomeA.00484-14

Pedersen, T. B., Vogensen, F. K., and Ardo, Y. (2016). Effect of heterofermentative lactic acid bacteria of DL-starters in initial ripening of semi-hard cheese. *Int. Dairy J.* 57, 72–79. doi: 10.1016/j.idairyj.2016.02.041

Pérez, G., Cardell, E., and Zárate, V. (2002). Random amplified polymorphic DNA analysis for differentiation of *Leuconostoc mesenteroides* subspecies isolated from Tenerife cheese. *Lett. Appl. Microbiol.* 34, 82–85. doi: 10.1046/j.1472-765x.2002.01050.x

Porcellato, D., Ostlie, H. M., Brede, M. E., Martinovic, A., and Skeie, S. B. (2013). Dynamics of starter, adjunct non-starter lactic acid bacteria and propionic acid

bacteria in low-fat and full-fat Dutch-type cheese. *Int. Dairy J.* 33, 104–111. doi: 10.1016/j.idairyj.2013.01.007

Porcellato, D., and Skeie, S. B. (2016). Bacterial dynamics and functional analysis of microbial metagenomes during ripening of Dutch-type cheese. *Int. Dairy J.* 61, 182–188. doi: 10.1016/j.idairyj.2016.05.005

Salvador, S., and Chan, P. (2004). "Determining the number of clusters/segments in hierarchical clustering/segmentation algorithms," in *Proceedings of the 16th IEEE International Conference on Tools with Artificial Intelligence* (Boca Raton, FL: IEEE Computer Society).

Sánchez, J. I., Martínez, B., and Rodríguez, A. (2005). Rational selection of *Leuconostoc* strains for mixed starters based on the physiological biodiversity found in raw milk fermentations. *Int. J. Food Microbiol.* 105, 377–387. doi: 10.1016/j.ijfoodmicro.2005.04.025

Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30, 2068–2069. doi: 10.1093/bioinformatics/btu153

Thunell, R. K. (1995). Taxonomy of the *Leuconostocs*. *J. Dairy Sci.* 78, 2514–2522. doi: 10.3168/jds.S0022-0302(95)76881-3

Vaughan, E. E., David, S., Harrington, A., Daly, C., Fitzgerald, G. F., and De Vos, W. M. (1995). Characterization of plasmid-encoded citrate permease (citP) genes from *Leuconostoc* species reveals high sequence conservation with the *Lactococcus lactis* citP gene. *Appl. Environ. Microbiol.* 61, 3172–3176.

Vedamuthu, E. R. (1994). The dairy *Leuconostoc*: use in dairy products. *J. Dairy Sci.* 77, 2725–2737. doi: 10.3168/jds.S0022-0302(94)77215-5

Vihavainen, E. J., and Björkroth, K. J. (2009). Diversity of *Leuconostoc gasicomitatum* associated with meat spoilage. *Int. J. Food Microbiol.* 136, 32–36. doi: 10.1016/j.ijfoodmicro.2009.09.010

Villani, F., Moschetti, G., Blaiotta, G., and Coppola, S. (1997). Characterization of strains of *Leuconostoc mesenteroides* by analysis of soluble whole-cell protein pattern, DNA fingerprinting and restriction of ribosomal DNA. *J. Appl. Microbiol.* 82, 578–588. doi: 10.1111/j.1365-2672.1997.tb03588.x

Vogensen, F. K., Karst, T., Larsen, J. J., Kringelum, B., Ellekjaer, D., and Waagner Nielsen, E. (1987). Improved direct differentiation between *Leuconostoc cremoris*, *Streptococcus lactis* ssp. diacetylactis, and *Streptococcus cremoris*/*Streptococcus lactis* on agar. *Milchwissenschaft* 42, 646–648.

Vongerichten, K. F., Klein, J. R., Matern, H., and Plapp, R. (1994). Cloning and nucleotide sequence analysis of pepV, a carnosinase gene from *Lactobacillus delbrueckii* subsp. *lactis* DSM 7290, and partial characterization of the enzyme. *Microbiology* 140, 2591–2600. doi: 10.1099/00221287-140-10-2591

Ward, D. M., Weller, R., and Bateson, M. M. (1990). 16S rRNA sequences reveal numerous uncultured microorganisms in a natural community. *Nature* 345, 63–65.

Warnes, G. R., Bolker, B., Bonebakker, L., R. G., Liaw, W. H. A., Lumley, T., Maechler, M., et al. (2015). *gplots: Various R Programming Tools for Plotting Data*. R package version 2.16.0. [Computer software]. Available online at: http://CRAN.R-project.org/package=gplots

Wright, E. (2015). *DECIPHER: Database Enabled Code for Ideal Probe Hybridization Employing*. R. package v1.16.1.

Zeller-Péronnet, V., Brockmann, E., Pavlovic, M., Timke, M., Busch, U., and Huber, I. (2013). Potential and limitations of MALDI-TOF MS for discrimination within the species *Leuconostoc mesenteroides* and *Leuconostoc pseudomesenteroides*. *J. für Verbraucherschutz und Lebensmittelsicherheit* 8, 205–214. doi: 10.1007/s00003-013-0826-z

Zhang, J., Kobert, K., Flouri, T., and Stamatakis, A. (2014). PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics* 30, 614–620. doi: 10.1093/bioinformatics/btt593

Zhang, W., Liu, W., Song, Y., Xu, H., Menghe, B., Zhang, H., et al. (2015). Multilocus sequence typing of a dairy-associated *Leuconostoc mesenteroides* population reveals clonal structure with intragenic homologous recombination. *J. Dairy Sci.* 98, 2284–2293. doi: 10.3168/jds.2014-9227

Zhao, Y., Jia, X., Yang, J., Ling, Y., Zhang, Z., Yu, J., et al. (2014). PanGP: a tool for quickly analyzing bacterial pan-genome profile. *Bioinformatics* 30, 1297–1299. doi: 10.1093/bioinformatics/btu017

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.

Isolate name	Accession No.	Isolation source	Reference
19254	NZ_ACKV00000000	Homo sapiens	NCBI Genomes
T26	NZ_JAUJ00000000	dairy starter culture	Pedersen, T. B. (2016)
LbT16	NZ_LAYV00000000	Taleggio cheese	Campedelli, I. (2015)
TIFN8	NZ_ATAZ00000000	dairy starter culture	Erkus, O. (2013)
LbE15	NZ_LAYN00000000	Taleggio cheese	Campedelli, I. (2015)
LbE16	NZ_LAYU00000000	Taleggio cheese	Campedelli, I. (2015)
20484	GCA_001047695	Cheese	Park, G. (2015)
J18	GCA_000234825	Kimchi	Jung (2012)
KACC 91922	NZ_JMEA00000000	Kimchi	NCBI Genomes
1159	NZ_JAUI00000000	Dairy starter culture	Pedersen, T. B. (2014)
PS12	NZ_JDVA00000000	Dairy starter culture	Pedersen, T. B. (2014)
4882	NZ_CAKV00000000	Dairy starter culture	Meslier, V. (2012)
LN02	MPLG00000000	Dairy starter culture	This study(c)
LN05	MPLH00000000	Dairy starter culture	This study(c)
LN07	MPLI00000000	Dairy starter culture	This study(c)
LN08	MPLJ00000000	Culture collection	This study(c)
LN12	MPLK00000000	Culture collection	This study(c)
LN19	MPLL00000000	Whey isolate	This study(c)
LN23	MPLM00000000	Dairy starter culture	This study(c)
LN24	MPLN00000000	Culture collection	This study(c)
LN25	MPLO00000000	Cheese isolate	This study(c)
LN27	MPLP00000000	Culture collection	This study(c)
LN32	MPLQ00000000	Dairy starter culture	This study(c)
LN34	MPLR00000000	Dairy starter culture	This study(c)
TW1	MPBC00000000	Twarog	This study
TW3	MPBD00000000	Twarog	This study
TW6	MPBA00000000	Twarog	This study
TW8	MPBB00000000	Twarog	This study
CF01	MPAM00000000	Dairy starter culture A	This study
CF02	MPAN00000000	Dairy starter culture A	This study
CF03	MPAO00000000	Dairy starter culture A	This study
CF04	MPAP00000000	Dairy starter culture A	This study
CF05	MPAQ00000000	Dairy starter culture A	This study
CF06	MPAR00000000	Dairy starter culture A	This study
CF07	MPAS00000000	Dairy starter culture A	This study
CF08	MPLS00000000	Dairy starter culture A	This study
CF09	MPAT00000000	Dairy starter culture A	This study
CF10	MPAU00000000	Dairy starter culture A	This study
CF11	MPEA00000000	Dairy starter culture A	This study
CF12	MPEB00000000	Dairy starter culture A	This study
CF13	MPEC00000000	Dairy starter culture A	This study
CF14	MPED00000000	Dairy starter culture A	This study
CF15	MPEE00000000	Dairy starter culture A	This study
CF16	MPEF00000000	Dairy starter culture A	This study
CF17	MPEG00000000	Dairy starter culture A	This study

CF18	MPEH00000000	Dairy starter culture A	This study
CF19	MPEI00000000	Dairy starter culture A	This study
CF20	MPEJ00000000	Dairy starter culture A	This study
HPKA1	MPAL00000000	Dairy starter culture C	This study
H61	MPHN00000000	Cheese	Østlie, M, H. (2016)
H83	MPHO00000000	Cheese	Østlie, M, H. (2016)
H95	MPHP00000000	Cheese	Østlie, M, H. (2016)
H97	MPHQ00000000	Cheese	Østlie, M, H. (2016)
H100	MPHR00000000	Cheese	Østlie, M, H. (2016)
H278	MPHS00000000	Cheese	Østlie, M, H. (2016)
H280	MPHT00000000	Cheese	Østlie, M, H. (2016)
H284	MPHU00000000	Cheese	Østlie, M, H. (2016)
BM2	MPLF00000000	Butter Milk	This study(d)

(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.

Organism	Genome size (Mb)	Contigs	CDS
Leuconostoc mesenteroides subsp. cremoris	1.638	126	1.720
Leuconostoc mesenteroides subsp. cremoris	1.832	123	1.931
Leuconostoc mesenteroides subsp. mesenteroides(a)	1.906	65	1.979
Leuconostoc mesenteroides subsp. cremoris	1.686	143	1.929
Leuconostoc mesenteroides subsp. dextranicum	2.007	63	2.059
Leuconostoc mesenteroides subsp. mesenteroides	2.036	85	2.094
Leuconostoc mesenteroides subsp. dextranicum	1.818	1	1.858
Leuconostoc mesenteroides subsp. mesenteroides	1.900	1	1.946
Leuconostoc lactis	1.683	30	1.626
Leuconostoc pseudomesenteroides	2.038	100	2.075
Leuconostoc pseudomesenteroides	1.934	91	1.963
Leuconostoc pseudomesenteroides	2.008	1	2.135
Leuconostoc pseudomesenteroides	1.936	123	1.929
Leuconostoc mesenteroides subsp. mesenteroides	2.212	252	2.094
Leuconostoc mesenteroides subsp. cremoris	1.636	215	1.683
Leuconostoc mesenteroides subsp. mesenteroides	1.998	166	2.068
Leuconostoc pseudomesenteroides	1.884	127	1.860
Leuconostoc lactis	1.737	70	1.728
Leuconostoc pseudomesenteroides	2.011	157	2.028
Leuconostoc lactis	1.730	69	1.721
Leuconostoc mesenteroides subsp. mesenteroides	2.277	68	1.705
Leuconostoc mesenteroides subsp. mesenteroides	2.231	69	2.229
Leuconostoc mesenteroides subsp. mesenteroides	1.860	94	1.843
Leuconostoc mesenteroides subsp. mesenteroides	1.891	151	1.852
Leuconostoc pseudomesenteroides	2.028	81	2.097
Leuconostoc pseudomesenteroides	1.938	56	1.969
Leuconostoc pseudomesenteroides	1.928	67	1.944
Leuconostoc pseudomesenteroides	2.054	56	2.125
Leuconostoc mesenteroides subsp. cremoris	1.776	146	1.867
Leuconostoc mesenteroides subsp. cremoris	1.785	80	1.928
Leuconostoc mesenteroides subsp. cremoris	1.743	92	1.856
Leuconostoc mesenteroides subsp. cremoris	1.702	88	1.790
Leuconostoc mesenteroides subsp. cremoris	1.699	88	1.809
Leuconostoc pseudomesenteroides	2.005	68	2.009
Leuconostoc mesenteroides subsp. cremoris	1.683	186	1.751
Leuconostoc pseudomesenteroides	2.146	163	2.148
Leuconostoc mesenteroides subsp. cremoris	1.675	81	1.753
Leuconostoc mesenteroides subsp. cremoris	1.679	83	1.762
Leuconostoc mesenteroides subsp. cremoris	1.678	89	1.757
Leuconostoc mesenteroides subsp. cremoris	1.682	115	1,76
Leuconostoc mesenteroides subsp. cremoris	1.669	223	1.736
Leuconostoc mesenteroides subsp. cremoris	1.685	74	1.773
Leuconostoc pseudomesenteroides	1.974	59	2.00
Leuconostoc mesenteroides subsp. cremoris	1.680	97	1.767
Leuconostoc mesenteroides subsp. cremoris	1.678	81	1.751

Leuconostoc mesenteroides subsp. cremoris	1.686	71	1.775
Leuconostoc mesenteroides subsp. cremoris	1.694	94	1.787
Leuconostoc mesenteroides subsp. cremoris	1.685	74	1.773
Leuconostoc pseudomesenteroides	2.009	74	2,021
Leuconostoc pseudomesenteroides.(b)	2.009	62	2.02
Leuconostoc pseudomesenteroides.(b)	2.100	212	2.183
Leuconostoc pseudomesenteroides.(b)	2.071	133	2.121
Leuconostoc pseudomesenteroides.(b)	2.061	85	2.123
Leuconostoc pseudomesenteroides.(b)	2.115	256	2.198
Leuconostoc pseudomesenteroides.(b)	1.994	72	2.006
Leuconostoc pseudomesenteroides.(b)	1.980	85	1.991
Leuconostoc pseudomesenteroides.(b)	2.073	222	2.134
Leuconostoc pseudomesenteroides	2.054	149	2.108

Pseudomesenteroides vs Lactis

Category	Subsystem	Role
Amino Acids and Derivatives	Alanine biosynthesis	**Cysteine desulfurase (EC 2.8.1.7), SufS subfamily**
Amino Acids and Derivatives	Arginine Biosynthesis -- gjo	**Glutamate N-acetyltransferase (EC 2.3.1.35)**
Amino Acids and Derivatives	Arginine Biosynthesis -- gjo	**N-acetylglutamate synthase (EC 2.3.1.1)**
Amino Acids and Derivatives	Chorismate Synthesis	**Chorismate mutase I (EC 5.4.99.5)**
Amino Acids and Derivatives	Branched-Chain Amino Acid Biosynthesis	**2-isopropylmalate synthase (EC 2.3.3.13)**
Amino Acids and Derivatives	Branched-Chain Amino Acid Biosynthesis	**3-isopropylmalate dehydratase large subunit (EC 4.2.1.33)**
Amino Acids and Derivatives	Branched-Chain Amino Acid Biosynthesis	**3-isopropylmalate dehydratase small subunit (EC 4.2.1.33)**
Amino Acids and Derivatives	Branched-Chain Amino Acid Biosynthesis	**3-isopropylmalate dehydrogenase (EC 1.1.1.85)**
Amino Acids and Derivatives	Histidine Biosynthesis	**ATP phosphoribosyltransferase (EC 2.4.2.17)**
Amino Acids and Derivatives	Histidine Biosynthesis	**Histidinol dehydrogenase (EC 1.1.1.23)**
Amino Acids and Derivatives	Histidine Biosynthesis	**Imidazole glycerol phosphate synthase amidotransferase subunit (EC 2.4.2.-)**
Amino Acids and Derivatives	Histidine Biosynthesis	**Imidazole glycerol phosphate synthase cyclase subunit (EC 4.1.3.-)**
Amino Acids and Derivatives	Histidine Biosynthesis	**Imidazoleglycerol-phosphate dehydratase (EC 4.2.1.19)**
Amino Acids and Derivatives	Histidine Biosynthesis	**Phosphoribosyl-AMP cyclohydrolase (EC 3.5.4.19)**
Amino Acids and Derivatives	Histidine Biosynthesis	**Phosphoribosyl-ATP pyrophosphatase (EC 3.6.1.31)**
Amino Acids and Derivatives	Histidine Biosynthesis	**Phosphoribosylformimino-5-aminoimidazole carboxamide ribotide isomerase (EC 5.3.1.16)**
Amino Acids and Derivatives	Methionine Biosynthesis	**5-methyltetrahydrofolate--homocysteine methyltransferase (EC 2.1.1.13)**
Amino Acids and Derivatives	Methionine Biosynthesis	**5-methyltetrahydropteroyltriglutamate--homocysteine methyltransferase (EC 2.1.1.14)**
Amino Acids and Derivatives	Methionine Biosynthesis	**Cystathionine gamma-lyase (EC 4.4.1.1)**
Carbohydrates	Chitin and N-acetylglucosamine utilization	**Chitinase (EC 3.2.1.14)**
Carbohydrates	Chitin and N-acetylglucosamine utilization	**N-Acetyl-D-glucosamine ABC transport system, permease protein 1**
Carbohydrates	Chitin and N-acetylglucosamine utilization	**N-Acetyl-D-glucosamine ABC transport system, permease protein 2**
Carbohydrates	Chitin and N-acetylglucosamine utilization	**PTS system, N-acetylglucosamine-specific IIB component (EC 2.7.1.69)**
Carbohydrates	Pyruvate metabolism I: anaplerotic reactions, PEP	**Malate permease**
Carbohydrates	Beta-Glucoside Metabolism	**Beta-glucosidase (EC 3.2.1.21)**
Carbohydrates	Beta-Glucoside Metabolism	**Outer surface protein of unknown function, cellobiose operon**
Carbohydrates	Maltose and Maltodextrin Utilization	**Maltose/maltodextrin transport ATP-binding protein MalK (EC 3.6.3.19)**
Carbohydrates	Sucrose utilization	**Sucrose permease, major facilitator superfamily**
Carbohydrates	Trehalose Uptake and Utilization	**Trehalose 6-phosphate phosphorylase (EC 2.4.1.216)**
Carbohydrates	D-gluconate and ketogluconates metabolism	**2-dehydro-3-deoxygluconate kinase (EC 2.7.1.45)**
Carbohydrates	D-gluconate and ketogluconates metabolism	**Glucose 1-dehydrogenase (EC 1.1.1.47)**
Carbohydrates	D-gluconate and ketogluconates metabolism	**L-idonate 5-dehydrogenase (EC 1.1.1.264)**
Carbohydrates	D-ribose utilization	**Ribose operon repressor**
Carbohydrates	Deoxyribose and Deoxynucleoside Catabolism	**Deoxyribonucleoside regulator DeoR (transcriptional repressor)**
Carbohydrates	Deoxyribose and Deoxynucleoside Catabolism	**Phosphopentomutase (EC 5.4.2.7)**
Carbohydrates	L-ascorbate utilization (and related gene clusters)	**3-keto-L-gulonate 6-phosphate decarboxylase**
Carbohydrates	L-ascorbate utilization (and related gene clusters)	**Ascorbate-specific PTS system, EIIA component (EC 2.7.1.-)**
Carbohydrates	L-ascorbate utilization (and related gene clusters)	**Ascorbate-specific PTS system, EIIB component (EC 2.7.1.69)**
Carbohydrates	L-ascorbate utilization (and related gene clusters)	**Ascorbate-specific PTS system, EIIC component**
Carbohydrates	L-ascorbate utilization (and related gene clusters)	**L-xylulose 5-phosphate 3-epimerase (EC 5.1.3.-)**
Carbohydrates	L-ascorbate utilization (and related gene clusters)	**Probable L-ascorbate-6-phosphate lactonase UlaG (EC 3.1.1.-) (L-ascorbate utilization protein G)**
Carbohydrates	L-ascorbate utilization (and related gene clusters)	**Transcriptional antiterminator with PTS regulation domain, SPy0181 ortholog**
Carbohydrates	Glycerol and Glycerol-3-phosphate Uptake and Utilization	**Glycerol-3-phosphate ABC transporter, periplasmic glycerol-3-phosphate-binding protein (TC 3.A.1.1.3)**
Carbohydrates	Glycerol and Glycerol-3-phosphate Uptake and Utilization	**Glycerol-3-phosphate transporter**
Cell Wall and Capsule	Teichoic and lipoteichoic acids biosynthesis	**CDP-glycerol: N-acetyl-beta-D-mannosaminyl-1,4-N-acetyl-D-glucosaminyldiphosphoundecaprenyl glycerophosphotransferase**
Cell Wall and Capsule	Teichoic and lipoteichoic acids biosynthesis	**Putative polyribitolphosphotransferase**
Cell Wall and Capsule	Teichoic and lipoteichoic acids biosynthesis	**Teichoic acid glycosylation protein**
Clustering-based subsystems	Divergent RNA modification related clusters	**Predicted nucleoside phosphatase**
Clustering-based subsystems	Glycyl-tRNA synthetase containing cluster	**Zinc uptake regulation protein ZUR**
Clustering-based subsystems	Peptidyl-prolyl cis-trans isomerase containing cluster	**FIG056164: rhomboid family serine protease**
Cofactors, Vitamins, Prosthetic Gr	Biotin biosynthesis	**Biotin-protein ligase (EC 6.3.4.15)**
Cofactors, Vitamins, Prosthetic Gr	Biotin biosynthesis Experimental	**Competence protein F homolog, phosphoribosyltransferase domain**
Cofactors, Vitamins, Prosthetic Gr	5-FCL-like protein	**5,10-methylenetetrahydrofolate reductase (EC 1.5.1.20)**
Cofactors, Vitamins, Prosthetic Gr	5-FCL-like protein	**Phosphomethylpyrimidine kinase (EC 2.7.4.7)**
Cofactors, Vitamins, Prosthetic Gr	5-FCL-like protein	**Thiaminase II (EC 3.5.99.2)**
Cofactors, Vitamins, Prosthetic Gr	NAD and NADP cofactor biosynthesis global	**Amidases related to nicotinamidase**
Cofactors, Vitamins, Prosthetic Gr	NAD and NADP cofactor biosynthesis global	**Predicted N-ribosylNicotinamide CRP-like regulator**
Cofactors, Vitamins, Prosthetic Gr	Heme and Siroheme Biosynthesis	**Ferrochelatase, protoheme ferro-lyase (EC 4.99.1.1)**
DNA Metabolism	CRISPRs	**CRISPR-associated protein Cas1**
DNA Metabolism	CRISPRs	**CRISPR-associated protein Cas2**
DNA Metabolism	CRISPRs	**CRISPR-associated protein, Csn1 family**
DNA Metabolism	DNA repair, bacterial photolyase	**Deoxyribodipyrimidine photolyase (EC 4.1.99.3)**
DNA Metabolism	DNA replication strays	**DNA polymerase III polC-type (EC 2.7.7.7)**
DNA Metabolism	Restriction-Modification System	**Type I restriction-modification system, DNA-methyltransferase subunit M (EC 2.1.1.72)**
Membrane Transport	Transport of Nickel and Cobalt	**HoxN/HupN/NixA family nickel/cobalt transporter**
Membrane Transport	ECF class transporters	**Substrate-specific component FolT of folate ECF transporter**
Membrane Transport	ECF class transporters	**Substrate-specific component PdxU of predicted pyridoxine ECF transporter**
Membrane Transport	ECF class transporters	**Substrate-specific component ThiT of thiamin ECF transporter**
Metabolism of Aromatic Compour	Benzoate degradation	**benzoate MFS transporter BenK**
Metabolism of Aromatic Compour	Biphenyl Degradation	**Biphenyl-2,3-diol 1,2-dioxygenase (EC 1.13.11.39)**
Miscellaneous	Iron-sulfur cluster assembly	**Iron-sulfur cluster assembly ATPase protein SufC**
Miscellaneous	Iron-sulfur cluster assembly	**Iron-sulfur cluster assembly protein SufB**
Miscellaneous	Iron-sulfur cluster assembly	**Iron-sulfur cluster assembly protein SufD**
Miscellaneous	Iron-sulfur cluster assembly	**PaaD-like protein (DUF59) involved in Fe-S cluster assembly**
Miscellaneous	Iron-sulfur cluster assembly	**Putative iron-sulfur cluster assembly scaffold protein for SUF system, SufE2**
Miscellaneous	Muconate lactonizing enzyme family	**L-alanine-DL-glutamate epimerase**
Miscellaneous	Phosphoglycerate mutase protein family	**Phosphoglycerate mutase family 5**
Nucleosides and Nucleotides	Housecleaning nucleoside triphosphate pyrophosphatases	**5'-nucleotidase YjjG (EC 3.1.3.5)**
Nucleosides and Nucleotides	Purine Utilization	**Xanthine permease**
Nucleosides and Nucleotides	Purine conversions	**GMP reductase (EC 1.7.1.7)**
Nucleosides and Nucleotides	Purine conversions	**Nucleotide pyrophosphatase (EC 3.6.1.9)**
Nucleosides and Nucleotides	Adenosyl nucleosidases	**Purine nucleoside phosphorylase (EC 2.4.2.1)**
Nucleosides and Nucleotides	Ribonucleotide reduction	**Ribonucleotide reductase of class III (anaerobic), activating protein (EC 1.97.1.4)**
Nucleosides and Nucleotides	Ribonucleotide reduction	**Ribonucleotide reductase of class III (anaerobic), large subunit (EC 1.17.4.2)**
Phages, Prophages, Transposable	Phage capsid proteins	**Phage head maturation protease**
Phages, Prophages, Transposable	Phage replication	**DNA helicase, phage-associated**
Phages, Prophages, Transposable	Phage replication	**DNA primase/helicase, phage-associated**
Phages, Prophages, Transposable	Phage tail proteins	**Phage major tail protein**

Lactis vs Pseudomesenteroides

Category	Subsystem	Role
Amino Acids and Derivatives	Chorismate: Intermediate for synthesis of Tryptophan, PAPA antibiotics, PABA, 3-hydroxyanth	**Aminodeoxychorismate lyase (EC 4.1.3.38)**
Amino Acids and Derivatives	Chorismate: Intermediate for synthesis of Tryptophan, PAPA antibiotics, PABA, 3-hydroxyanth	**Isochorismatase (EC 3.3.2.1)**
Amino Acids and Derivatives	Chorismate: Intermediate for synthesis of Tryptophan, PAPA antibiotics, PABA, 3-hydroxyanth	**Para-aminobenzoate synthase, amidotransferase component (EC 2.6.1.85)**
Amino Acids and Derivatives	Chorismate: Intermediate for synthesis of Tryptophan, PAPA antibiotics, PABA, 3-hydroxyanth	**Para-aminobenzoate synthase, aminase component (EC 2.6.1.85)**
Amino Acids and Derivatives	Chorismate: Intermediate for synthesis of Tryptophan, PAPA antibiotics, PABA, 3-hydroxyanth	**Phosphoribosylanthranilate isomerase (EC 5.3.1.24)**
Carbohydrates	Pyruvate Alanine Serine Interconversions	**D-serine/D-alanine/glycine transporter**
Carbohydrates	Pyruvate metabolism I: anaplerotic reactions, PEP	**Malolactic regulator**
Carbohydrates	Pyruvate metabolism II: acetyl-CoA, acetogenesis from pyruvate	**Acylphosphate phosphohydrolase (EC 3.6.1.7), putative**
Carbohydrates	Trehalose Uptake and Utilization	**PTS system, trehalose-specific IIA component (EC 2.7.1.69)**
Carbohydrates	Trehalose Uptake and Utilization	**Trehalose-6-phosphate hydrolase (EC 3.2.1.93)**
Carbohydrates	Acetoin, butanediol metabolism	**2,3-butanediol dehydrogenase, S-alcohol forming, (R)-acetoin-specific (EC 1.1.1.4)**
Carbohydrates	Fructose utilization	**1-phosphofructokinase (EC 2.7.1.56)**
Carbohydrates	Fructose utilization	**PTS system, fructose-specific IIA component (EC 2.7.1.69)**
Carbohydrates	Fructose utilization	**PTS system, fructose-specific IIB component (EC 2.7.1.69)**
Carbohydrates	Fructose utilization	**PTS system, fructose-specific IIC component (EC 2.7.1.69)**
Carbohydrates	Fructose utilization	**Transcriptional repressor of the fructose operon, DeoR family**
Carbohydrates	Alpha-Amylase locus in Streptocococcus	**putative esterase**
Cell Wall and Capsule	Exopolysaccharide Biosynthesis	**Manganese-dependent protein-tyrosine phosphatase (EC 3.1.3.48)**
Cell Wall and Capsule	Exopolysaccharide Biosynthesis	**Tyrosine-protein kinase EpsD (EC 2.7.10.2)**
Cell Wall and Capsule	Exopolysaccharide Biosynthesis	**Tyrosine-protein kinase transmembrane modulator EpsC**
Cell Wall and Capsule	Murein Hydrolases	**Membrane-bound lytic murein transglycosylase D precursor (EC 3.2.1.-)**
Clustering-based subsystems	CBSS-262719.3.peg.410	**Replicative DNA helicase (EC 3.6.1.-) [SA14-24]**
Clustering-based subsystems	CBSS-1313.3.peg.391	**lojap protein**
Clustering-based subsystems	CBSS-176279.3.peg.1262	**Hypothetical protein SAV1839**
Cofactors, Vitamins, Prosthetic Groups, Pigmer	5-FCL-like protein	**Substrate-specific component ThiW of predicted thiazole ECF transporter**
Cofactors, Vitamins, Prosthetic Groups, Pigmer	Folate Biosynthesis	**2-amino-4-hydroxy-6-hydroxymethyldihydropteridine pyrophosphokinase (EC 2.7.6.3)**
Cofactors, Vitamins, Prosthetic Groups, Pigmer	Folate Biosynthesis	**Dihydroneopterin aldolase (EC 4.1.2.25)**
Cofactors, Vitamins, Prosthetic Groups, Pigmer	Folate Biosynthesis	**Dihydropteroate synthase (EC 2.5.1.15)**
Cofactors, Vitamins, Prosthetic Groups, Pigmer	Folate Biosynthesis	**GTP cyclohydrolase I (EC 3.5.4.16) type 1**
Cofactors, Vitamins, Prosthetic Groups, Pigmer	Pyridoxin (Vitamin B6) Biosynthesis	**Pyridoxine biosynthesis glutamine amidotransferase, glutaminase subunit (EC 2.4.2.-)**
Cofactors, Vitamins, Prosthetic Groups, Pigmer	Pyridoxin (Vitamin B6) Biosynthesis	**Pyridoxine biosynthesis glutamine amidotransferase, synthase subunit (EC 2.4.2.-)**
Cofactors, Vitamins, Prosthetic Groups, Pigmer	Riboflavin, FMN and FAD metabolism	**5-amino-6-(5-phosphoribosylamino)uracil reductase (EC 1.1.1.193)**
Cofactors, Vitamins, Prosthetic Groups, Pigmer	Riboflavin, FMN and FAD metabolism	**6,7-dimethyl-8-ribityllumazine synthase (EC 2.5.1.78)**
Cofactors, Vitamins, Prosthetic Groups, Pigmer	Riboflavin, FMN and FAD metabolism	**Diaminohydroxyphosphoribosylaminopyrimidine deaminase (EC 3.5.4.26)**
Cofactors, Vitamins, Prosthetic Groups, Pigmer	Riboflavin, FMN and FAD metabolism	**GTP cyclohydrolase II (EC 3.5.4.25)**
Cofactors, Vitamins, Prosthetic Groups, Pigmer	Riboflavin, FMN and FAD metabolism	**Riboflavin synthase eubacterial/eukaryotic (EC 2.5.1.9)**
DNA Metabolism	DNA repair, bacterial	**Exonuclease SbcC**
DNA Metabolism	Gram Positive Competence	**ComF operon protein C**
DNA Metabolism	Restriction-Modification System	**Type III restriction-modification system methylation subunit (EC 2.1.1.72)**
Fatty Acids, Lipids, and Isoprenoids	Fatty Acid Biosynthesis FASII	**Enoyl-[acyl-carrier-protein] reductase [NADPH] (EC 1.3.1.10)**
Iron acquisition and metabolism	Encapsulating protein for DyP-type peroxidase and ferritin-like protein oligomers	**Predicted dye-decolorizing peroxidase (DyP), encapsulated subgroup**
Nucleosides and Nucleotides	Purine Utilization	**Guanine deaminase (EC 3.5.4.3)**
Phages, Prophages, Transposable elements, Pla	Phage capsid proteins	**Phage capsid and scaffold**
Phages, Prophages, Transposable elements, Pla	Phage capsid proteins	**Phage major capsid protein**
Phages, Prophages, Transposable elements, Pla	Phage packaging machinery	**Phage terminase small subunit**
Phages, Prophages, Transposable elements, Pla	Phage replication	**DNA replication protein, phage-associated**
Potassium metabolism	Potassium homeostasis	**Potassium-transporting ATPase A chain (EC 3.6.3.12) (TC 3.A.3.7.1)**
Potassium metabolism	Potassium homeostasis	**Potassium-transporting ATPase B chain (EC 3.6.3.12) (TC 3.A.3.7.1)**
Potassium metabolism	Potassium homeostasis	**Potassium-transporting ATPase C chain (EC 3.6.3.12) (TC 3.A.3.7.1)**
Protein Metabolism	tRNAs	**tRNA-Ser-GGA**
RNA Metabolism	Ribonuclease H	**Ribonuclease HI (EC 3.1.26.4)**
Regulation and Cell signaling	LysR-family proteins in Salmonella enterica Typhimurium	**LysR family transcriptional regulator STM2281**
Respiration	Respiratory dehydrogenases 1	**Methanol dehydrogenase large subunit protein (EC 1.1.99.8)**
Stress Response	Heat shock dnaK gene cluster extended	**Xanthosine/inosine triphosphate pyrophosphatase**
Stress Response	Choline and Betaine Uptake and Betaine Biosynthesis	**L-proline glycine betaine ABC transport system permease protein ProV (TC 3.A.1.12.1)**
Virulence, Disease and Defense	Multidrug Resistance Efflux Pumps	**Multi antimicrobial extrusion protein (Na(+)/drug antiporter), MATE family of MDR efflux pumps**

Mesenteroides vs pseudomesenteroides

Category	Subsystem	Role
Amino Acids and Derivatives	Chorismate: Intermediate for synthesis of Tryptophan, PAPA antibiotics, PABA, ‡	Anthranilate synthase, amidotransferase component (EC 4.1.3.27)
Amino Acids and Derivatives	Chorismate: Intermediate for synthesis of Tryptophan, PAPA antibiotics, PABA, ‡	Anthranilate synthase, aminase component (EC 4.1.3.27)
Amino Acids and Derivatives	Chorismate: Intermediate for synthesis of Tryptophan, PAPA antibiotics, PABA, ‡	Indole-3-glycerol phosphate synthase (EC 4.1.1.48)
Amino Acids and Derivatives	Chorismate: Intermediate for synthesis of Tryptophan, PAPA antibiotics, PABA, ‡	Para-aminobenzoate synthase, amidotransferase component (EC 2.6.1.85)
Amino Acids and Derivatives	Branched-Chain Amino Acid Biosynthesis	Dihydroxy-acid dehydratase (EC 4.2.1.9)
Amino Acids and Derivatives	Branched-Chain Amino Acid Biosynthesis	Ketol-acid reductoisomerase (EC 1.1.1.86)
Amino Acids and Derivatives	Branched-Chain Amino Acid Biosynthesis	Threonine dehydratase (EC 4.3.1.19)
Amino Acids and Derivatives	Histidine Biosynthesis	ATP phosphoribosyltransferase regulatory subunit (EC 2.4.2.17)
Amino Acids and Derivatives	Lysine Biosynthesis DAP Pathway	N-acetyl-L,L-diaminopimelate aminotransferase (EC 2.6.1.-)
Amino Acids and Derivatives	Methionine Biosynthesis	Cystathionine beta-synthase (EC 4.2.1.22)
Amino Acids and Derivatives	Methionine Biosynthesis	O-acetylhomoserine sulfhydrylase (EC 2.5.1.49)
Amino Acids and Derivatives	Methionine Biosynthesis	O-succinylhomoserine sulfhydrylase (EC 2.5.1.48)
Carbohydrates	Acetoin, butanediol metabolism	2,3-butanediol dehydrogenase, S-alcohol forming, (R)-acetoin-specific (EC 1.1.1.4)
Carbohydrates	Acetoin, butanediol metabolism	Acetoin (diacetyl) reductase (EC 1.1.1.304)
Carbohydrates	Acetoin, butanediol metabolism	Acetolactate synthase large subunit (EC 2.2.1.6)
Carbohydrates	Acetoin, butanediol metabolism	Acetolactate synthase small subunit (EC 2.2.1.6)
Carbohydrates	Glycerol and Glycerol-3-phosphate Uptake and Utilization	Glycerophosphoryl diester phosphodiesterase, periplasmic (EC 3.1.4.46)
Cell Wall and Capsule	Rhamnose containing glycans	capsular polysaccharide biosynthesis protein
Cell Wall and Capsule	Peptidoglycan Biosynthesis	D-alanine--D-alanine ligase A (EC 6.3.2.4)
Clustering-based subsystems	Bacterial Cell Division	Cell division protein FtsX
Clustering-based subsystems	CBSS-176279.3.peg.1262	Hypothetical protein SAV1839
Clustering-based subsystems	CBSS-176299.4.peg.1996A	Endoribonuclease L-PSP
Cofactors, Vitamins, Prosthetic Groups, Pigments	Biotin biosynthesis	Adenosylmethionine-8-amino-7-oxononanoate aminotransferase (EC 2.6.1.62)
Cofactors, Vitamins, Prosthetic Groups, Pigments	Biotin biosynthesis	Dethiobiotin synthetase (EC 6.3.3.3)
Cofactors, Vitamins, Prosthetic Groups, Pigments	Menaquinone and Phylloquinone Biosynthesis	2-heptaprenyl-1,4-naphthoquinone methyltransferase (EC 2.1.1.163)
DNA Metabolism	DNA repair, UvrABC system	Excinuclease ABC subunit A paralog of unknown function
DNA Metabolism	DNA repair, bacterial	Exonuclease SbcD
DNA Metabolism	Gram Positive Competence	ComF operon protein C
DNA Metabolism	Restriction-Modification System	Type I restriction-modification system, specificity subunit S (EC 3.1.21.3)
DNA Metabolism	Restriction-Modification System	Type III restriction-modification system methylation subunit (EC 2.1.1.72)
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/Co/Ni transporter MgtE
Membrane Transport	ECF class transporters	Substrate-specific component STY3230 of queuosine-regulated ECF transporter
Miscellaneous	Conserved gene cluster possibly involved in RNA metabolism	Serine acetyltransferase (EC 2.3.1.30)
Miscellaneous	Phosphoglycerate mutase protein family	Phosphoglycerate mutase family 2
Nucleosides and Nucleotides	Purine conversions	Adenosine deaminase (EC 3.5.4.4)
Nucleosides and Nucleotides	Purine conversions	preQ1-regulated inosine-uridine nucleoside hydrolase (EC 3.2.2.1)
Phages, Prophages, Transposable elements, Plasm	Phage packaging machinery	Phage terminase small subunit
Phages, Prophages, Transposable elements, Plasm	Phage tail fiber proteins	Phage tail fiber protein
Potassium metabolism	Hyperosmotic potassium uptake	Potassium uptake protein TrkH
RNA Metabolism	RNA processing orphans	2'-5' RNA ligase
RNA Metabolism	Ribonuclease H	Ribonuclease HI (EC 3.1.26.4)
Regulation and Cell signaling	LysR-family proteins in Escherichia coli	Chromosome initiation inhibitor
Regulation and Cell signaling	LysR-family proteins in Escherichia coli	Cys regulon transcriptional activator CysB
Respiration	F0F1-type ATP synthase	ATP synthase F0 sector subunit a (EC 3.6.3.14)
Respiration	F0F1-type ATP synthase	ATP synthase F0 sector subunit b (EC 3.6.3.14)
Respiration	Biogenesis of c-type cytochromes	Cytochrome c-type biogenesis protein DsbD, protein-disulfide reductase (EC 1.8.1.8)
Secondary Metabolism	Auxin biosynthesis	Anthranilate phosphoribosyltransferase (EC 2.4.2.18)
Secondary Metabolism	Auxin biosynthesis	Phosphoribosylanthranilate isomerase (EC 5.3.1.24)
Secondary Metabolism	Auxin biosynthesis	Tryptophan synthase alpha chain (EC 4.2.1.20)
Secondary Metabolism	Auxin biosynthesis	Tryptophan synthase beta chain (EC 4.2.1.20)
Stress Response	Cold shock, CspA family of proteins	Cold shock protein CspG
Stress Response	Heat shock dnaK gene cluster extended	Xanthosine/inosine triphosphate pyrophosphatase
Stress Response	Choline and Betaine Uptake and Betaine Biosynthesis	L-proline glycine betaine ABC transport system permease protein ProV (TC 3.A.1.12.1)
Sulfur Metabolism	Thioredoxin-disulfide reductase	Alkyl hydroperoxide reductase protein C (EC 1.6.4.-)
Sulfur Metabolism	Thioredoxin-disulfide reductase	Alkyl hydroperoxide reductase protein F (EC 1.6.4.-)

Pseudomesenteriodes vs mesenteroides

Category	Subsystem	Role
Amino Acids and Derivatives	Chorismate Synthesis	**Chorismate mutase I (EC 5.4.99.5)**
Amino Acids and Derivatives	Branched-Chain Amino Acid Biosynthesis	**3-isopropylmalate dehydratase small subunit (EC 4.2.1.33)**
Amino Acids and Derivatives	Methionine Biosynthesis	**5-methyltetrahydrofolate--homocysteine methyltransferase (EC 2.1.1.13)**
Amino Acids and Derivatives	Threonine degradation	**Threonine dehydrogenase and related Zn-dependent dehydrogenases**
Carbohydrates	Chitin and N-acetylglucosamine utilization	**Chitinase (EC 3.2.1.14)**
Carbohydrates	Chitin and N-acetylglucosamine utilization	**N-Acetyl-D-glucosamine ABC transport system, permease protein 1**
Carbohydrates	Chitin and N-acetylglucosamine utilization	**PTS system, N-acetylglucosamine-specific IIB component (EC 2.7.1.69)**
Carbohydrates	Glycolysis and Gluconeogenesis	**Fructose-bisphosphate aldolase class II (EC 4.1.2.13)**
Carbohydrates	Pentose phosphate pathway	**Transketolase, C-terminal section (EC 2.2.1.1)**
Carbohydrates	Pyruvate metabolism I: anaplerotic reactions, PEP	**Malolactic enzyme (EC 1.-.-.-)**
Carbohydrates	Pyruvate metabolism II: acetyl-CoA, acetogenesis from pyruvate	**Pyruvate oxidase (EC 1.2.3.3)**
Carbohydrates	Beta-Glucoside Metabolism	**6-phospho-beta-glucosidase (EC 3.2.1.86)**
Carbohydrates	Beta-Glucoside Metabolism	**Outer surface protein of unknown function, cellobiose operon**
Carbohydrates	Beta-Glucoside Metabolism	**PTS system, cellobiose-specific IIA component (EC 2.7.1.69)**
Carbohydrates	Beta-Glucoside Metabolism	**PTS system, cellobiose-specific IIB component (EC 2.7.1.69)**
Carbohydrates	Beta-Glucoside Metabolism	**PTS system, cellobiose-specific IIC component (EC 2.7.1.69)**
Carbohydrates	Maltose and Maltodextrin Utilization	**Alpha-glucosidase (EC 3.2.1.20)**
Carbohydrates	Maltose and Maltodextrin Utilization	**Beta-phosphoglucomutase (EC 5.4.2.6)**
Carbohydrates	Maltose and Maltodextrin Utilization	**Maltose operon transcriptional repressor MalR, LacI family**
Carbohydrates	Maltose and Maltodextrin Utilization	**Maltose phosphorylase (EC 2.4.1.8)**
Carbohydrates	Maltose and Maltodextrin Utilization	**Maltose/maltodextrin ABC transporter, permease protein MalF**
Carbohydrates	Maltose and Maltodextrin Utilization	**Maltose/maltodextrin ABC transporter, permease protein MalG**
Carbohydrates	Maltose and Maltodextrin Utilization	**Maltose/maltodextrin transport ATP-binding protein MalK (EC 3.6.3.19)**
Carbohydrates	Maltose and Maltodextrin Utilization	**Neopullulanase (EC 3.2.1.135)**
Carbohydrates	Maltose and Maltodextrin Utilization	**Oligo-1,6-glucosidase (EC 3.2.1.10)**
Carbohydrates	Sucrose utilization	**Sucrose permease, major facilitator superfamily**
Carbohydrates	Trehalose Biosynthesis	**Trehalose phosphorylase (EC 2.4.1.64)**
Carbohydrates	Trehalose Uptake and Utilization	**Trehalose 6-phosphate phosphorylase (EC 2.4.1.216)**
Carbohydrates	Acetoin, butanediol metabolism	**Acetoin dehydrogenase E1 component alpha-subunit (EC 1.2.4.-)**
Carbohydrates	Acetoin, butanediol metabolism	**Dihydrolipoamide acetyltransferase component (E2) of acetoin dehydrogenase complex (EC 2.3.1.-)**
Carbohydrates	Acetoin, butanediol metabolism	**Dihydrolipoamide dehydrogenase of acetoin dehydrogenase (EC 1.8.1.4)**
Carbohydrates	D-gluconate and ketogluconates metabolism	**2-dehydro-3-deoxygluconate kinase (EC 2.7.1.45)**
Carbohydrates	D-gluconate and ketogluconates metabolism	**Glucose 1-dehydrogenase (EC 1.1.1.47)**
Carbohydrates	D-gluconate and ketogluconates metabolism	**L-idonate 5-dehydrogenase (EC 1.1.1.264)**
Carbohydrates	Deoxyribose and Deoxynucleoside Catabolism	**Deoxyribonucleoside regulator DeoR (transcriptional repressor)**
Carbohydrates	Deoxyribose and Deoxynucleoside Catabolism	**Phosphopentomutase (EC 5.4.2.7)**
Carbohydrates	L-ascorbate utilization (and related gene clusters)	**3-keto-L-gulonate 6-phosphate decarboxylase**
Carbohydrates	L-ascorbate utilization (and related gene clusters)	**Ascorbate-specific PTS system, EIIA component (EC 2.7.1.-)**
Carbohydrates	L-ascorbate utilization (and related gene clusters)	**Ascorbate-specific PTS system, EIIB component (EC 2.7.1.69)**
Carbohydrates	L-ascorbate utilization (and related gene clusters)	**L-xylulose 5-phosphate 3-epimerase (EC 5.1.3.-)**
Carbohydrates	L-ascorbate utilization (and related gene clusters)	**Probable L-ascorbate-6-phosphate lactonase UlaG (EC 3.1.1.-) (L-ascorbate utilization protein G)**
Carbohydrates	L-ascorbate utilization (and related gene clusters)	**Transcriptional antiterminator with PTS regulation domain, SPy0181 ortholog**
Carbohydrates	Xylose utilization	**Beta-xylosidase (EC 3.2.1.37)**
Carbohydrates	Xylose utilization	**Xyloside transporter XynT**
Carbohydrates	Formaldehyde assimilation: Ribulose monophosphate pathway	**6-phospho-3-hexuloisomerase**
Carbohydrates	Lactate utilization	**D-Lactate dehydrogenase (EC 1.1.2.5)**
Carbohydrates	Alpha-Amylase locus in Streptocococcus	**Maltose/maltodextrin ABC transporter, substrate binding periplasmic protein MalE**
Carbohydrates	Glycerol and Glycerol-3-phosphate Uptake and Utilization	**Glycerol-3-phosphate ABC transporter, periplasmic glycerol-3-phosphate-binding protein (TC 3.A.1.1.3)**
Carbohydrates	Glycerol and Glycerol-3-phosphate Uptake and Utilization	**Glycerol-3-phosphate transporter**
Carbohydrates	VC0266	**Hypothetical protein VC0266 (sugar utilization related?)**
Cell Wall and Capsule	Teichoic and lipoteichoic acids biosynthesis	**CDP-glycerol: N-acetyl-beta-D-mannosaminyl-1,4-N-acetyl-D-glucosaminyldiphosphoundecaprenyl glycerophos**
Cell Wall and Capsule	Teichoic and lipoteichoic acids biosynthesis	**Putative polyribitolphosphotransferase**
Cell Wall and Capsule	Peptidoglycan Biosynthesis	**D-alanine--D-alanine ligase (EC 6.3.2.4)**
Clustering-based subsystems	CBSS-393130.3.peg.794	**Free methionine-(R)-sulfoxide reductase, contains GAF domain**
Clustering-based subsystems	Bacterial Cell Division	**Cell division transporter, ATP-binding protein FtsE (TC 3.A.5.1.1)**
Clustering-based subsystems	Peptidyl-prolyl cis-trans isomerase containing cluster	**FIG056164: rhomboid family serine protease**
Cofactors, Vitamins, Prosthetic Gro	Biotin biosynthesis	**Biotin-protein ligase (EC 6.3.4.15)**
Cofactors, Vitamins, Prosthetic Gro	Biotin biosynthesis Experimental	**Competence protein F homolog, phosphoribosyltransferase domain**
Cofactors, Vitamins, Prosthetic Gro	5-FCL-like protein	**Thiaminase II (EC 3.5.99.2)**
Cofactors, Vitamins, Prosthetic Gro	NAD and NADP cofactor biosynthesis global	**Amidases related to nicotinamidase**
Cofactors, Vitamins, Prosthetic Gro	NAD and NADP cofactor biosynthesis global	**Niacin transporter NiaP**
Cofactors, Vitamins, Prosthetic Gro	NAD and NADP cofactor biosynthesis global	**Nicotinamide phosphoribosyltransferase (EC 2.4.2.12)**
Cofactors, Vitamins, Prosthetic Gro	NAD and NADP cofactor biosynthesis global	**Nudix-related transcriptional regulator NrtR**
Cofactors, Vitamins, Prosthetic Gro	NAD and NADP cofactor biosynthesis global	**Predicted N-ribosylNicotinamide CRP-like regulator**
Cofactors, Vitamins, Prosthetic Gro	Menaquinone and Phylloquinone Biosynthesis	**Ubiquinone/menaquinone biosynthesis methyltransferase UbiE (EC 2.1.1.-)**
Cofactors, Vitamins, Prosthetic Gro	Riboflavin, FMN and FAD metabolism in plants	**FIG000859: hypothetical protein YebC**
DNA Metabolism	CRISPRs	**CRISPR-associated protein Cas1**
DNA Metabolism	CRISPRs	**CRISPR-associated protein Cas2**
DNA Metabolism	CRISPRs	**CRISPR-associated protein, Csn1 family**
DNA Metabolism	2-phosphoglycolate salvage	**Phosphoglycolate phosphatase (EC 3.1.3.18)**
DNA Metabolism	DNA repair, bacterial	**DNA repair exonuclease family protein YhaO**
DNA Metabolism	DNA repair, bacterial	**DNA-cytosine methyltransferase (EC 2.1.1.37)**
DNA Metabolism	DNA repair, bacterial photolyase	**Deoxyribodipyrimidine photolyase (EC 4.1.99.3)**
DNA Metabolism	DNA replication strays	**DNA polymerase III polC-type (EC 2.7.7.7)**
DNA Metabolism	Gram Positive Competence	**Late competence protein ComGB, access of DNA to ComEA**
DNA Metabolism	Restriction-Modification System	**Type I restriction-modification system, DNA-methyltransferase subunit M (EC 2.1.1.72)**
DNA Metabolism	Restriction-Modification System	**Type I restriction-modification system, restriction subunit R (EC 3.1.21.3)**
Membrane Transport	ABC transporter alkylphosphonate (TC 3.A.1.9.1)	**Phosphonate ABC transporter ATP-binding protein (TC 3.A.1.9.1)**
Membrane Transport	ABC transporter alkylphosphonate (TC 3.A.1.9.1)	**Phosphonate ABC transporter permease protein phnE1 (TC 3.A.1.9.1)**
Membrane Transport	ABC transporter alkylphosphonate (TC 3.A.1.9.1)	**Phosphonate ABC transporter permease protein phnE2 (TC 3.A.1.9.1)**
Membrane Transport	ABC transporter alkylphosphonate (TC 3.A.1.9.1)	**Phosphonate ABC transporter phosphate-binding periplasmic component (TC 3.A.1.9.1)**
Membrane Transport	Proton-dependent Peptide Transporters	**Di-/tripeptide transporter**
Membrane Transport	ECF class transporters	**Duplicated ATPase component YkoD of energizing module of thiamin-regulated ECF transporter for HydroxyMe**
Membrane Transport	ECF class transporters	**Substrate-specific component PdxU of predicted pyridoxine ECF transporter**
Membrane Transport	ECF class transporters	**Substrate-specific component PdxU2 of predicted pyridoxin-related ECF transporter**
Membrane Transport	ECF class transporters	**Substrate-specific component ThiT of thiamin ECF transporter**
Membrane Transport	Ton and Tol transport systems	**TolA protein**
Metabolism of Aromatic Compound	Biphenyl Degradation	**Biphenyl-2,3-diol 1,2-dioxygenase (EC 1.13.11.39)**
Miscellaneous	Muconate lactonizing enzyme family	**L-alanine-DL-glutamate epimerase**
Nucleosides and Nucleotides	Housecleaning nucleoside triphosphate pyrophosphatases	**5'-nucleotidase YjjG (EC 3.1.3.5)**
Nucleosides and Nucleotides	Nudix proteins (nucleoside triphosphate hydrolases)	**Putative Nudix hydrolase YfcD (EC 3.6.-.-)**
Nucleosides and Nucleotides	Purine Utilization	**Cytosine/purine/uracil/thiamine/allantoin permease family protein**

Nucleosides and Nucleotides	Purine Utilization	**Xanthine permease**
Nucleosides and Nucleotides	Purine conversions	**2',3'-cyclic-nucleotide 2'-phosphodiesterase (EC 3.1.4.16)**
Nucleosides and Nucleotides	Purine conversions	**GMP reductase (EC 1.7.1.7)**
Nucleosides and Nucleotides	De Novo Pyrimidine Synthesis	**Dihydroorotase (EC 3.5.2.3)**
Nucleosides and Nucleotides	pyrimidine conversions	**Uridine kinase (EC 2.7.1.48) [C1]**
Nucleosides and Nucleotides	Adenosyl nucleosidases	**Purine nucleoside phosphorylase (EC 2.4.2.1)**
Phages, Prophages, Transposable e	Phage capsid proteins	**Phage head maturation protease**
Phages, Prophages, Transposable e	Phage packaging machinery	**Phage terminase, small subunit**
Phages, Prophages, Transposable e	Phage replication	**DNA helicase, phage-associated**
Phages, Prophages, Transposable e	Phage replication	**DNA primase/helicase, phage-associated**
Phages, Prophages, Transposable e	Phage tail proteins	**Phage major tail protein**
Phages, Prophages, Transposable e	Phage tail proteins	**Phage tail length tape-measure protein**
Potassium metabolism	Potassium homeostasis	**Kup system potassium uptake protein**
Protein Metabolism	Ribosome biogenesis bacterial	**Ribosomal-protein-L7p-serine acetyltransferase**
Protein Metabolism	tRNAs	**tRNA-Ala-CGC**
Protein Metabolism	tRNAs	**tRNA-Pro-CGG**
Protein Metabolism	Omega peptidases (EC 3.4.19.-)	**Acylamino-acid-releasing enzyme (EC 3.4.19.1)**
Protein Metabolism	Peptidyl-prolyl cis-trans isomerase	**Putative parvulin type peptidyl-prolyl isomerase, similarity with PrsA foldase**
Protein Metabolism	Ribosomal protein S5p acylation	**Ribosomal-protein-S5p-alanine acetyltransferase**
RNA Metabolism	RNA processing orphans	**2H phosphoesterase superfamily protein BC2899**
RNA Metabolism	Rrf2 family transcriptional regulators	**Rrf2 family transcriptional regulator**
RNA Metabolism	Rrf2 family transcriptional regulators	**Rrf2 family transcriptional regulator, group III**
Respiration	Quinone oxidoreductase family	**Quinone oxidoreductase (EC 1.6.5.5)**
Stress Response	Cold shock, CspA family of proteins	**Cold shock protein CspC**
Stress Response	Oxidative stress	**Ferroxidase (EC 1.16.3.1)**
Stress Response	Oxidative stress	**Iron-binding ferritin-like antioxidant protein**
Stress Response	Oxidative stress	**Non-specific DNA-binding protein Dps**
Virulence, Disease and Defense	Beta-lactamase	**Beta-lactamase class A**
Virulence, Disease and Defense	Cobalt-zinc-cadmium resistance	**Probable cadmium-transporting ATPase (EC 3.6.3.3)**
Virulence, Disease and Defense	Copper homeostasis	**Copper chaperone**
Virulence, Disease and Defense	Streptococcus pneumoniae Vancomycin Tolerance Locus	**ABC transporter membrane-spanning permease, Pep export, Vex3**

Lactis vs mesenteroides

Category	Subsystem	Role
Amino Acids and Derivatives	Chorismate: Intermediate for synthesis of Tryptophan, PAPA antibiotics, PABA, 3-hydroxyanthranilate and more.	**Aminodeoxychorismate lyase (EC 4.1.3.38)**
Amino Acids and Derivatives	Chorismate: Intermediate for synthesis of Tryptophan, PAPA antibiotics, PABA, 3-hydroxyanthranilate and more.	**Isochorismatase (EC 3.3.2.1)**
Amino Acids and Derivatives	Chorismate: Intermediate for synthesis of Tryptophan, PAPA antibiotics, PABA, 3-hydroxyanthranilate and more.	**Para-aminobenzoate synthase, aminase component (EC 2.6.1.85)**
Amino Acids and Derivatives	Threonine degradation	**Threonine dehydrogenase and related Zn-dependent dehydrogenases**
Carbohydrates	Glycolysis and Gluconeogenesis	**Fructose-bisphosphate aldolase class II (EC 4.1.2.13)**
Carbohydrates	Pentose phosphate pathway	**Transketolase, C-terminal section (EC 2.2.1.1)**
Carbohydrates	Pyruvate Alanine Serine Interconversions	**D-serine/D-alanine/glycine transporter**
Carbohydrates	Pyruvate metabolism I: anaplerotic reactions, PEP	**Malolactic enzyme (EC 1.-.-.-)**
Carbohydrates	Pyruvate metabolism I: anaplerotic reactions, PEP	**Malolactic regulator**
Carbohydrates	Pyruvate metabolism II: acetyl-CoA, acetogenesis from pyruvate	**Acylphosphate phosphohydrolase (EC 3.6.1.7), putative**
Carbohydrates	Pyruvate metabolism II: acetyl-CoA, acetogenesis from pyruvate	**Pyruvate oxidase (EC 1.2.3.3)**
Carbohydrates	Maltose and Maltodextrin Utilization	**Alpha-glucosidase (EC 3.2.1.20)**
Carbohydrates	Maltose and Maltodextrin Utilization	**Beta-phosphoglucomutase (EC 5.4.2.6)**
Carbohydrates	Maltose and Maltodextrin Utilization	**Maltose operon transcriptional repressor MalR, LacI family**
Carbohydrates	Maltose and Maltodextrin Utilization	**Maltose phosphorylase (EC 2.4.1.8)**
Carbohydrates	Maltose and Maltodextrin Utilization	**Maltose/maltodextrin ABC transporter, permease protein MalF**
Carbohydrates	Maltose and Maltodextrin Utilization	**Maltose/maltodextrin ABC transporter, permease protein MalG**
Carbohydrates	Maltose and Maltodextrin Utilization	**Neopullulanase (EC 3.2.1.135)**
Carbohydrates	Maltose and Maltodextrin Utilization	**Oligo-1,6-glucosidase (EC 3.2.1.10)**
Carbohydrates	Trehalose Biosynthesis	**Trehalose phosphorylase (EC 2.4.1.64)**
Carbohydrates	Trehalose Uptake and Utilization	**PTS system, trehalose-specific IIA component (EC 2.7.1.69)**
Carbohydrates	Acetoin, butanediol metabolism	**Acetoin dehydrogenase E1 component alpha-subunit (EC 1.2.4.-)**
Carbohydrates	Acetoin, butanediol metabolism	**Dihydrolipoamide acetyltransferase component (E2) of acetoin dehydrogenase complex (EC 2.3.1.-)**
Carbohydrates	Acetoin, butanediol metabolism	**Dihydrolipoamide dehydrogenase of acetoin dehydrogenase (EC 1.8.1.4)**
Carbohydrates	Fructose utilization	**1-phosphofructokinase (EC 2.7.1.56)**
Carbohydrates	Fructose utilization	**PTS system, fructose-specific IIA component (EC 2.7.1.69)**
Carbohydrates	Fructose utilization	**PTS system, fructose-specific IIB component (EC 2.7.1.69)**
Carbohydrates	Fructose utilization	**PTS system, fructose-specific IIC component (EC 2.7.1.69)**
Carbohydrates	Fructose utilization	**Transcriptional repressor of the fructose operon, DeoR family**
Carbohydrates	Xylose utilization	**Beta-xylosidase (EC 3.2.1.37)**
Carbohydrates	Xylose utilization	**Xyloside transporter XynT**
Carbohydrates	Formaldehyde assimilation: Ribulose monophosphate pathway	**6-phospho-3-hexuloisomerase**
Carbohydrates	Lactate utilization	**D-Lactate dehydrogenase (EC 1.1.2.5)**
Carbohydrates	Alpha-Amylase locus in Streptococcoccus	**Maltose/maltodextrin ABC transporter, substrate binding periplasmic protein MalE**
Carbohydrates	Alpha-Amylase locus in Streptococcoccus	**putative esterase**
Carbohydrates	VC0266	**Hypothetical protein VC0266 (sugar utilization related?)**
Cell Wall and Capsule	Exopolysaccharide Biosynthesis	**Tyrosine-protein kinase EpsD (EC 2.7.10.2)**
Cell Wall and Capsule	Exopolysaccharide Biosynthesis	**Tyrosine-protein kinase transmembrane modulator EpsC**
Cell Wall and Capsule	Murein Hydrolases	**Membrane-bound lytic murein transglycosylase D precursor (EC 3.2.1.-)**
Cell Wall and Capsule	Peptidoglycan Biosynthesis	**D-alanine--D-alanine ligase (EC 6.3.2.4)**
Clustering-based subsystems	CBSS-393130.3.peg.794	**Free methionine-(R)-sulfoxide reductase, contains GAF domain**
Clustering-based subsystems	CBSS-262719.3.peg.410	**Replicative DNA helicase (EC 3.6.1.-) [SA14-24]**
Clustering-based subsystems	Bacterial Cell Division	**Cell division transporter, ATP-binding protein FtsE (TC 3.A.5.1.1)**
Clustering-based subsystems	CBSS-1313.3.peg.391	**Iojap protein**
Cofactors, Vitamins, Prosthetic Groups, Pigments	5-FCL-like protein	**Substrate-specific component ThiW of predicted thiazole ECF transporter**
Cofactors, Vitamins, Prosthetic Groups, Pigments	Folate Biosynthesis	**2-amino-4-hydroxy-6-hydroxymethyldihydropteridine pyrophosphokinase (EC 2.7.6.3)**
Cofactors, Vitamins, Prosthetic Groups, Pigments	Folate Biosynthesis	**Dihydroneopterin aldolase (EC 4.1.2.25)**
Cofactors, Vitamins, Prosthetic Groups, Pigments	Folate Biosynthesis	**Dihydropteroate synthase (EC 2.5.1.15)**
Cofactors, Vitamins, Prosthetic Groups, Pigments	Folate Biosynthesis	**GTP cyclohydrolase I (EC 3.5.4.16) type 1**
Cofactors, Vitamins, Prosthetic Groups, Pigments	NAD and NADP cofactor biosynthesis global	**Niacin transporter NiaP**
Cofactors, Vitamins, Prosthetic Groups, Pigments	NAD and NADP cofactor biosynthesis global	**Nicotinamide phosphoribosyltransferase (EC 2.4.2.12)**
Cofactors, Vitamins, Prosthetic Groups, Pigments	NAD and NADP cofactor biosynthesis global	**Nudix-related transcriptional regulator NrtR**
Cofactors, Vitamins, Prosthetic Groups, Pigments	Pyridoxin (Vitamin B6) Biosynthesis	**Pyridoxine biosynthesis glutamine amidotransferase, glutaminase subunit (EC 2.4.2.-)**
Cofactors, Vitamins, Prosthetic Groups, Pigments	Pyridoxin (Vitamin B6) Biosynthesis	**Pyridoxine biosynthesis glutamine amidotransferase, synthase subunit (EC 2.4.2.-)**
Cofactors, Vitamins, Prosthetic Groups, Pigments	Menaquinone and Phylloquinone Biosynthesis	**Ubiquinone/menaquinone biosynthesis methyltransferase UbiE (EC 2.1.1.-)**
Cofactors, Vitamins, Prosthetic Groups, Pigments	Riboflavin, FMN and FAD metabolism	**5-amino-6-(5-phosphoribosylamino)uracil reductase (EC 1.1.1.193)**
Cofactors, Vitamins, Prosthetic Groups, Pigments	Riboflavin, FMN and FAD metabolism	**6,7-dimethyl-8-ribityllumazine synthase (EC 2.5.1.78)**
Cofactors, Vitamins, Prosthetic Groups, Pigments	Riboflavin, FMN and FAD metabolism	**Diaminohydroxyphosphoribosylaminopyrimidine deaminase (EC 3.5.4.26)**
Cofactors, Vitamins, Prosthetic Groups, Pigments	Riboflavin, FMN and FAD metabolism	**GTP cyclohydrolase II (EC 3.5.4.25)**
Cofactors, Vitamins, Prosthetic Groups, Pigments	Riboflavin, FMN and FAD metabolism	**Riboflavin synthase eubacterial/eukaryotic (EC 2.5.1.9)**
Cofactors, Vitamins, Prosthetic Groups, Pigments	Riboflavin, FMN and FAD metabolism in plants	**FIG000859: hypothetical protein YebC**
DNA Metabolism	2-phosphoglycolate salvage	**Phosphoglycolate phosphatase (EC 3.1.3.18)**
DNA Metabolism	DNA repair, bacterial	**DNA repair exonuclease family protein YhaO**
DNA Metabolism	DNA repair, bacterial	**DNA-cytosine methyltransferase (EC 2.1.1.37)**
DNA Metabolism	DNA repair, bacterial	**Exonuclease SbcC**
DNA Metabolism	Gram Positive Competence	**Late competence protein ComGB, access of DNA to ComEA**
DNA Metabolism	Restriction-Modification System	**Type I restriction-modification system, restriction subunit R (EC 3.1.21.3)**
Fatty Acids, Lipids, and Isoprenoids	Fatty Acid Biosynthesis FASII	**Enoyl-[acyl-carrier-protein] reductase [NADPH] (EC 1.3.1.10)**
Iron acquisition and metabolism	Encapsulating protein for DyP-type peroxidase and ferritin-like protein oligomers	**Predicted dye-decolorizing peroxidase (DyP), encapsulated subgroup**
Membrane Transport	ABC transporter alkylphosphonate (TC 3.A.1.9.1)	**Phosphonate ABC transporter ATP-binding protein (TC 3.A.1.9.1)**
Membrane Transport	ABC transporter alkylphosphonate (TC 3.A.1.9.1)	**Phosphonate ABC transporter permease protein phnE1 (TC 3.A.1.9.1)**
Membrane Transport	ABC transporter alkylphosphonate (TC 3.A.1.9.1)	**Phosphonate ABC transporter permease protein phnE2 (TC 3.A.1.9.1)**
Membrane Transport	ABC transporter alkylphosphonate (TC 3.A.1.9.1)	**Phosphonate ABC transporter phosphate-binding periplasmic component (TC 3.A.1.9.1)**
Membrane Transport	Proton-dependent Peptide Transporters	**Di-/tripeptide transporter**
Membrane Transport	ECF class transporters	**Duplicated ATPase component YkoD of energizing module of thiamin-regulated ECF transporter for HydroxyMethylPyrimidine**
Membrane Transport	ECF class transporters	**Substrate-specific component PdxU2 of predicted pyridoxin-related ECF transporter**
Membrane Transport	Ton and Tol transport systems	**TolA protein**
Nucleosides and Nucleotides	Nudix proteins (nucleoside triphosphate hydrolases)	**Putative Nudix hydrolase YfcD (EC 3.6.-.-)**
Nucleosides and Nucleotides	Purine Utilization	**Cytosine/purine/uracil/thiamine/allantoin permease family protein**
Nucleosides and Nucleotides	Purine Utilization	**Guanine deaminase (EC 3.5.4.3)**
Nucleosides and Nucleotides	Purine conversions	**2',3'-cyclic-nucleotide 2'-phosphodiesterase (EC 3.1.4.16)**
Nucleosides and Nucleotides	De Novo Pyrimidine Synthesis	**Dihydroorotase (EC 3.5.2.3)**
Nucleosides and Nucleotides	pyrimidine conversions	**Uridine kinase (EC 2.7.1.48) [C1]**
Phages, Prophages, Transposable elements, Plasmids	Phage capsid proteins	**Phage capsid and scaffold**
Phages, Prophages, Transposable elements, Plasmids	Phage capsid proteins	**Phage major capsid protein**
Phages, Prophages, Transposable elements, Plasmids	Phage packaging machinery	**Phage terminase, small subunit**
Phages, Prophages, Transposable elements, Plasmids	Phage replication	**DNA replication protein, phage-associated**
Phages, Prophages, Transposable elements, Plasmids	Phage tail proteins	**Phage tail length tape-measure protein**
Potassium metabolism	Potassium homeostasis	**Kup system potassium uptake protein**
Potassium metabolism	Potassium homeostasis	**Potassium-transporting ATPase A chain (EC 3.6.3.12) (TC 3.A.3.7.1)**
Potassium metabolism	Potassium homeostasis	**Potassium-transporting ATPase B chain (EC 3.6.3.12) (TC 3.A.3.7.1)**
Potassium metabolism	Potassium homeostasis	**Potassium-transporting ATPase C chain (EC 3.6.3.12) (TC 3.A.3.7.1)**
Protein Metabolism	tRNAs	**tRNA-Ser-GGA**
Protein Metabolism	Peptidyl-prolyl cis-trans isomerase	**Putative parvulin type peptidyl-prolyl isomerase, similarity with PrsA foldase**
RNA Metabolism	RNA processing orphans	**2H phosphoesterase superfamily protein BC2899**
RNA Metabolism	Rrf2 family transcriptional regulators	**Rrf2 family transcriptional regulator**
RNA Metabolism	Rrf2 family transcriptional regulators	**Rrf2 family transcriptional regulator, group III**
Regulation and Cell signaling	LysR-family proteins in Salmonella enterica Typhimurium	**LysR family transcriptional regulator STM2281**
Respiration	Respiratory dehydrogenases 1	**Methanol dehydrogenase large subunit protein (EC 1.1.99.8)**
Respiration	Quinone oxidoreductase family	**Quinone oxidoreductase (EC 1.6.5.5)**
Stress Response	Oxidative stress	**Ferroxidase (EC 1.16.3.1)**
Stress Response	Oxidative stress	**Iron-binding ferritin-like antioxidant protein**
Stress Response	Oxidative stress	**Non-specific DNA-binding protein Dps**
Virulence, Disease and Defense	Copper homeostasis	**Copper chaperone**
Virulence, Disease and Defense	Multidrug Resistance Efflux Pumps	**Multi antimicrobial extrusion protein (Na(+)/drug antiporter), MATE family of MDR efflux pumps**

mesenteroides vs lactis

Category	Subsystem	Role
Amino Acids and Derivatives	Alanine biosynthesis	**Cysteine desulfurase (EC 2.8.1.7), SufS subfamily**
Amino Acids and Derivatives	Arginine Biosynthesis -- gjo	**Glutamate N-acetyltransferase (EC 2.3.1.35)**
Amino Acids and Derivatives	Arginine Biosynthesis -- gjo	**N-acetylglutamate synthase (EC 2.3.1.1)**
Amino Acids and Derivatives	Chorismate: Intermediate for synthesis of Tryptophan, PAPA antibiotics, PABA, 3-hydroxyanthranilat	**Anthranilate synthase, amidotransferase component (EC 4.1.3.27)**
Amino Acids and Derivatives	Chorismate: Intermediate for synthesis of Tryptophan, PAPA antibiotics, PABA, 3-hydroxyanthranilat	**Anthranilate synthase, aminase component (EC 4.1.3.27)**
Amino Acids and Derivatives	Chorismate: Intermediate for synthesis of Tryptophan, PAPA antibiotics, PABA, 3-hydroxyanthranilat	**Indole-3-glycerol phosphate synthase (EC 4.1.1.48)**
Amino Acids and Derivatives	Chorismate: Intermediate for synthesis of Tryptophan, PAPA antibiotics, PABA, 3-hydroxyanthranilat	**Phosphoribosylformimino-5-aminoimidazole carboxamide ribotide isomerase (EC 5.3.1.16)**
Amino Acids and Derivatives	Branched-Chain Amino Acid Biosynthesis	**2-isopropylmalate synthase (EC 2.3.3.13)**
Amino Acids and Derivatives	Branched-Chain Amino Acid Biosynthesis	**3-isopropylmalate dehydratase large subunit (EC 4.2.1.33)**
Amino Acids and Derivatives	Branched-Chain Amino Acid Biosynthesis	**3-isopropylmalate dehydrogenase (EC 1.1.1.85)**
Amino Acids and Derivatives	Branched-Chain Amino Acid Biosynthesis	**Dihydroxy-acid dehydratase (EC 4.2.1.9)**
Amino Acids and Derivatives	Branched-Chain Amino Acid Biosynthesis	**Ketol-acid reductoisomerase (EC 1.1.1.86)**
Amino Acids and Derivatives	Branched-Chain Amino Acid Biosynthesis	**Threonine dehydratase (EC 4.3.1.19)**
Amino Acids and Derivatives	Histidine Biosynthesis	**ATP phosphoribosyltransferase (EC 2.4.2.17)**
Amino Acids and Derivatives	Histidine Biosynthesis	**ATP phosphoribosyltransferase regulatory subunit (EC 2.4.2.17)**
Amino Acids and Derivatives	Histidine Biosynthesis	**Histidinol dehydrogenase (EC 1.1.1.23)**
Amino Acids and Derivatives	Histidine Biosynthesis	**Imidazole glycerol phosphate synthase amidotransferase subunit (EC 2.4.2.-)**
Amino Acids and Derivatives	Histidine Biosynthesis	**Imidazole glycerol phosphate synthase cyclase subunit (EC 4.1.3.-)**
Amino Acids and Derivatives	Histidine Biosynthesis	**Imidazoleglycerol-phosphate dehydratase (EC 4.2.1.19)**
Amino Acids and Derivatives	Histidine Biosynthesis	**Phosphoribosyl-AMP cyclohydrolase (EC 3.5.4.19)**
Amino Acids and Derivatives	Histidine Biosynthesis	**Phosphoribosyl-ATP pyrophosphatase (EC 3.6.1.31)**
Amino Acids and Derivatives	Lysine Biosynthesis DAP Pathway	**N-acetyl-L,L-diaminopimelate aminotransferase (EC 2.6.1.-)**
Amino Acids and Derivatives	Methionine Biosynthesis	**5-methyltetrahydropteroyltriglutamate--homocysteine methyltransferase (EC 2.1.1.14)**
Amino Acids and Derivatives	Methionine Biosynthesis	**Cystathionine beta-synthase (EC 4.2.1.22)**
Amino Acids and Derivatives	Methionine Biosynthesis	**Cystathionine gamma-lyase (EC 4.4.1.1)**
Amino Acids and Derivatives	Methionine Biosynthesis	**O-acetylhomoserine sulfhydrylase (EC 2.5.1.49)**
Amino Acids and Derivatives	Methionine Biosynthesis	**O-succinylhomoserine sulfhydrylase (EC 2.5.1.48)**
Carbohydrates	Chitin and N-acetylglucosamine utilization	**N-Acetyl-D-glucosamine ABC transport system, permease protein 2**
Carbohydrates	Pyruvate metabolism I: anaplerotic reactions, PEP	**Malate permease**
Carbohydrates	Acetoin, butanediol metabolism	**Acetoin (diacetyl) reductase (EC 1.1.1.304)**
Carbohydrates	Acetoin, butanediol metabolism	**Acetolactate synthase large subunit (EC 2.2.1.6)**
Carbohydrates	Acetoin, butanediol metabolism	**Acetolactate synthase small subunit (EC 2.2.1.6)**
Carbohydrates	D-ribose utilization	**Ribose operon repressor**
Carbohydrates	Glycerol and Glycerol-3-phosphate Uptake and Utilization	**Glycerophosphoryl diester phosphodiesterase, periplasmic (EC 3.1.4.46)**
Cell Wall and Capsule	Rhamnose containing glycans	**capsular polysaccharide biosynthesis protein**
Cell Wall and Capsule	Teichoic and lipoteichoic acids biosynthesis	**Teichoic acid glycosylation protein**
Cell Wall and Capsule	Peptidoglycan Biosynthesis	**D-alanine--D-alanine ligase A (EC 6.3.2.4)**
Clustering-based subsystems	Rad50-Mre11 DNA repair cluster	**DNA double-strand break repair Rad50 ATPase**
Clustering-based subsystems	Bacterial Cell Division	**Cell division protein FtsX**
Clustering-based subsystems	CBSS-176299.4.peg.1996A	**Endoribonuclease L-PSP**
Clustering-based subsystems	Divergent RNA modification related clusters	**Predicted nucleoside phosphatase**
Clustering-based subsystems	Glycyl-tRNA synthetase containing cluster	**Zinc uptake regulation protein ZUR**
Cofactors, Vitamins, Prosthetic Groups, Pigments	Biotin biosynthesis	**Adenosylmethionine-8-amino-7-oxononanoate aminotransferase (EC 2.6.1.62)**
Cofactors, Vitamins, Prosthetic Groups, Pigments	Biotin biosynthesis	**Dethiobiotin synthetase (EC 6.3.3.3)**
Cofactors, Vitamins, Prosthetic Groups, Pigments	5-FCL-like protein	**5,10-methylenetetrahydrofolate reductase (EC 1.5.1.20)**
Cofactors, Vitamins, Prosthetic Groups, Pigments	5-FCL-like protein	**Phosphomethylpyrimidine kinase (EC 2.7.4.7)**
Cofactors, Vitamins, Prosthetic Groups, Pigments	Menaquinone and Phylloquinone Biosynthesis	**2-heptaprenyl-1,4-naphthoquinone methyltransferase (EC 2.1.1.163)**
Cofactors, Vitamins, Prosthetic Groups, Pigments	Heme and Siroheme Biosynthesis	**Ferrochelatase, protoheme ferro-lyase (EC 4.99.1.1)**
DNA Metabolism	DNA repair, UvrABC system	**Excinuclease ABC subunit A paralog of unknown function**
DNA Metabolism	DNA repair, bacterial	**Exonuclease SbcD**
DNA Metabolism	Restriction-Modification System	**Type I restriction-modification system, specificity subunit S (EC 3.1.21.3)**
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/Co/Ni transporter MgtE**
Membrane Transport	Transport of Nickel and Cobalt	**HoxN/HupN/NixA family nickel/cobalt transporter**
Membrane Transport	ECF class transporters	**Substrate-specific component FolT of folate ECF transporter**
Membrane Transport	ECF class transporters	**Substrate-specific component STY3230 of queuosine-regulated ECF transporter**
Metabolism of Aromatic Compounds	Benzoate degradation	**benzoate MFS transporter BenK**
Miscellaneous	Conserved gene cluster possibly involved in RNA metabolism	**Serine acetyltransferase (EC 2.3.1.30)**
Miscellaneous	Iron-sulfur cluster assembly	**Iron-sulfur cluster assembly ATPase protein SufC**
Miscellaneous	Iron-sulfur cluster assembly	**Iron-sulfur cluster assembly protein SufB**
Miscellaneous	Iron-sulfur cluster assembly	**Iron-sulfur cluster assembly protein SufD**
Miscellaneous	Iron-sulfur cluster assembly	**PaaD-like protein (DUF59) involved in Fe-S cluster assembly**
Miscellaneous	Iron-sulfur cluster assembly	**Putative iron-sulfur cluster assembly scaffold protein for SUF system, SufE2**
Miscellaneous	Phosphoglycerate mutase protein family	**Phosphoglycerate mutase family 2**
Miscellaneous	Phosphoglycerate mutase protein family	**Phosphoglycerate mutase family 5**
Nucleosides and Nucleotides	Purine conversions	**Adenosine deaminase (EC 3.5.4.4)**
Nucleosides and Nucleotides	Purine conversions	**Nucleotide pyrophosphatase (EC 3.6.1.9)**
Nucleosides and Nucleotides	Purine conversions	**preQ1-regulated inosine-uridine nucleoside hydrolase (EC 3.2.2.1)**
Nucleosides and Nucleotides	Ribonucleotide reduction	**Ribonucleotide reductase of class III (anaerobic), activating protein (EC 1.97.1.4)**
Nucleosides and Nucleotides	Ribonucleotide reduction	**Ribonucleotide reductase of class III (anaerobic), large subunit (EC 1.17.4.2)**
Phages, Prophages, Transposable elements, Plasmids	Phage tail fiber proteins	**Phage tail fiber protein**
Potassium metabolism	Hyperosmotic potassium uptake	**Potassium uptake protein TrkH**
Protein Metabolism	Metallocarboxypeptidases (EC 3.4.17.-)	**D-alanyl-D-alanine carboxypeptidase (EC 3.4.16.4)**
Protein Metabolism	Protein degradation	**Aminopeptidase C (EC 3.4.22.40)**
Protein Metabolism	Protein deglycation	**Ribulosamine/erythrulosamine 3-kinase potentially involved in protein deglycation**
RNA Metabolism	RNA processing orphans	**2'-5' RNA ligase**
Regulation and Cell signaling	LysR-family proteins in Escherichia coli	**Chromosome initiation inhibitor**
Regulation and Cell signaling	LysR-family proteins in Escherichia coli	**Cys regulon transcriptional activator CysB**
Respiration	F0F1-type ATP synthase	**ATP synthase F0 sector subunit a (EC 3.6.3.14)**
Respiration	F0F1-type ATP synthase	**ATP synthase F0 sector subunit b (EC 3.6.3.14)**
Respiration	F0F1-type ATP synthase	**ATP synthase F0 sector subunit c (EC 3.6.3.14)**
Respiration	Biogenesis of c-type cytochromes	**Cytochrome c-type biogenesis protein DsbD, protein-disulfide reductase (EC 1.8.1.8)**
Secondary Metabolism	Auxin biosynthesis	**Anthranilate phosphoribosyltransferase (EC 2.4.2.18)**
Secondary Metabolism	Auxin biosynthesis	**Tryptophan synthase alpha chain (EC 4.2.1.20)**
Secondary Metabolism	Auxin biosynthesis	**Tryptophan synthase beta chain (EC 4.2.1.20)**
Stress Response	Cold shock, CspA family of proteins	**Cold shock protein CspG**
Stress Response	Glutathione: Biosynthesis and gamma-glutamyl cycle	**Glutamate--cysteine ligase (EC 6.3.2.2)**
Sulfur Metabolism	Galactosylceramide and Sulfatide metabolism	**Beta-galactosidase large subunit (EC 3.2.1.23)**
Sulfur Metabolism	Galactosylceramide and Sulfatide metabolism	**Beta-galactosidase small subunit (EC 3.2.1.23)**
Sulfur Metabolism	Thioredoxin-disulfide reductase	**Alkyl hydroperoxide reductase protein C (EC 1.6.4.-)**
Sulfur Metabolism	Thioredoxin-disulfide reductase	**Alkyl hydroperoxide reductase protein F (EC 1.6.4.-)**
Virulence, Disease and Defense	Cobalt-zinc-cadmium resistance	**Cobalt-zinc-cadmium resistance protein**
Virulence, Disease and Defense	Cobalt-zinc-cadmium resistance	**Cobalt-zinc-cadmium resistance protein CzcD**
Virulence, Disease and Defense	Copper homeostasis	**Multicopper oxidase**

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. Frantzen[1], Hans Petter Kleppen[1,2], Helge Holo[1, 3, *]

[1]Laboratory 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

[2]ACD Pharmaceuticals AS, Leknes, Norway

[3]TINE 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 leuconostocs. 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 erronous 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

130 determine phylogenetic distances between genomes and to construct a WPGMA phylogenetic
131 supertree using the sequence alignment metric functions in the Decipher v2.0 (34) and MASS
132 v7.3-47 (35) packages in R. A distance cut-off for the number of clusters was determined using
133 the knee of the curve approach (36), binning the isolates into genomic lineages.

134

135 **Relative quantification of the microbial community in starter cultures**

136 Compositional analysis in starter cultures was performed in triplicates on total DNA extracted
137 from the starter cultures using 1 mL of starter culture diluted to an OD_{600} of 1. The samples were
138 treated with 20 mg/mL lysozyme (Sigma-Aldrich) and 3 U/L mutanolysin (Sigma-Aldrich),
139 mechanically lysed using FastPrep (MP Biomedicals) with 0.5 g acid-washed glass beads (<106
140 µm) (Sigma-Aldrich) and purified using the Qiagen DNeasy Blood & Tissue Kit (Qiagen). A
141 suitable amplicon target was identified by screening the softcore-genes for nucleotide sequence
142 variation using the sequence alignment metrics functions in the DECIPHER package v1.16.1
143 (34). Genes without flanking consensus regions within a <500bp variable region adequate for
144 differentiation, or which did not provide sufficient discrimination between lineages, were
145 discarded. The loci *purR* and *epsD*, and the v2-v3 region of 16S rDNA were amplified by PCR
146 using the KAPA HiFi PCR Kit (KAPA Biosystems, Wilmington, Massachusetts, USA) with the
147 following primers: purR-324F (5'-YACTCCATCAAATCTTCGTAAAAT-3'), purR-811R (5'-
148 TGTCATTAAATATATTTCCCAATTGAACA-3'), epsD-138F (5'-
149 KCTTATYGCGGCTGCATT-3'), epsD-604R (5'-GATARTARAGTTCTAAATCTGCTCGT-
150 3'), 16S-44F (5'- GCGTGCCTAATACATGCAAGTYGA-3'), 16S-536R (5'-
151 CTGCTGGCACGTAKTTAGCCGTCC-3'). Forward (5'-
152 TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG-) and reverse

(5’GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG-) 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-

811), and the *epsD* gene (pos. 138-604). 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.

371 *cremoris* by 16S rDNA was demonstrated. This highlights the advantage of species-specific
372 amplicon targets compared to that of 16S rDNA. However, the sequence variation within the
373 *purR* amplicon was insufficient to differentiate between many of the genetic lineages. Thus, the
374 variance within the amplicons found among our core genes, is not high enough to expose the
375 complexity of DL starter cultures.

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

390 The discovery of *epsD* as a suitable target for strain differentiation was surprising, as the
391 gene was only present in 9 of the 30 reference strains. The *eps* operon has been found both on
392 plasmids (61, 62). and to be chromosomally located (46). The *epsD* gene was highly represented
393 among the starter culture strains, missing in only two of our 97 starter culture isolates. Apart

394 from the missing *eps* operon, we were unable to distinguish the two isolates CF124 and CF223
395 from their nearest pan- and core-genomic neighbors. In the laboratory, strains harboring *eps*
396 plasmids have been cured of their eps positive phenotype by serial transfers (61), and no
397 evidence exists that suggests a chromosomal locality confers higher stability over multiple
398 transfers (46). The high degree of sequence variation in the *eps* operon, and more specifically the
399 sequence variation in the *epsD* amplicon represents evolutionary diversification indicating a
400 history of selection pressure. Typically, lactococcal strains with different phage sensitivities also
401 contain different EPS, and strains that do not produce EPS have been demonstrated to exhibit
402 phage sensitivities different from strains which do produce EPS (61). Moreover, the production
403 of EPS has been shown to confer resistance to phages (62, 63). Regardless of what might be the
404 cause of the high degree of sequence variation in the *epsD* gene, its applied use in discrimination
405 and quantification of lactococcal diversity provides culture-independent, robust, and
406 reproducible data. Moreover, it provides the means to monitor temporal shifts in lactococcal
407 diversity, as well as comparing the genetic diversity of *Lactococcus lactis* between starter
408 cultures and starter culture batches.

409 The great rate of advancement in next-generation sequencing technologies over the past
410 decade has been accompanied by a rapid development of bioinformatics applications. The
411 reduced cost of sequencing has promoted whole-genome sequencing of bacterial isolates, and the
412 vast improvements to the downstream analysis of genomic data has taken comparative analysis
413 to a completely new level. Pan-genomic analysis of several hundred genomes enables
414 characterization and differentiation of bacteria, and facilitates the development of rapid and
415 robust methods such as targeted-amplicon sequencing of discriminatory loci. Dairy starter
416 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

1. **Vedamuthu ER.** 1994. The Dairy *Leuconostoc*: Use in Dairy Products. J Dairy Sci **77:**2725-2737.
2. **Smid EJ, Erkus O, Spus M, Wolkers-Rooijackers JC, Alexeeva S, Kleerebezem M.** 2014. Functional implications of the microbial community structure of undefined mesophilic starter cultures. Microb Cell Fact **13:**S2.
3. **Erkus O, de Jager VC, Spus M, van Alen-Boerrigter IJ, van Rijswijck IM, Hazelwood L, Janssen PW, van Hijum SA, Kleerebezem M, Smid EJ.** 2013. Multifactorial diversity sustains microbial community stability. ISME J **7**.
4. **Kleppen HP, Bang T, Nes IF, Holo H.** 2011. Bacteriophages in milk fermentations: Diversity fluctuations of normal and failed fermentations. Int Dairy J **21:**592-600.
5. **Rousseau GM, Moineau S.** 2009. Evolution of *Lactococcus lactis* phages within a cheese factory. Appl Environ Microbiol **75:**5336-5344.
6. **Boucher I, Moineau S.** 2001. Phages of *Lactococcus lactis*: An ecological and economical equilibrium. Recent Research and Developments in Virology **3:**243-256.
7. **Jany J-L, Barbier G.** 2008. Culture-independent methods for identifying microbial communities in cheese. Food Microbiology **25:**839-848.
8. **Verreault D, Gendron L, Rousseau GM, Veillette M, Massé D, Lindsley WG, Moineau S, Duchaine C.** 2011. Detection of Airborne Lactococcal Bacteriophages in Cheese Manufacturing Plants. Appl Environ Microbiol **77:**491-497.
9. **Neve H, Berger A, Heller KJ.** 1995. A method for detecting and enumerating airbourne virulent bacteriophage of dairy starter cultures. Kieler Milchwirtschaftliche Forschungsberichte **47:**193-207.
10. **Terzaghi BE, Sandine WE.** 1975. Improved medium for lactic streptococci and their bacteriophages. Appl Microbiol **29:**807-813.
11. **Hugenholtz J, Splint R, Konings WN, Veldkamp H.** 1987. Selection of protease-positive and protease-negative variants of Streptococcus cremoris. Appl Environ Microbiol **53:**309-314.
12. **Nurk S, Bankevich A, Antipov D, Gurevich AA, Korobeynikov A, Lapidus A, Prjibelski AD, Pyshkin A, Sirotkin A, Sirotkin Y, Stepanauskas R, Clingenpeel SR, Woyke T, McLean JS, Lasken R, Tesler G, Alekseyev MA, Pevzner PA.** 2013. Assembling single-cell genomes and mini-metagenomes from chimeric MDA products. J Comput Biol **20:**714-737.
13. **Seemann T.** 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics **30:**2068-2069.
14. **Bolotin A, Wincker P, Mauger S, Jaillon O, Malarme K, Weissenbach J, Ehrlich SD, Sorokin A.** 2001. The Complete Genome Sequence of the Lactic Acid Bacterium *Lactococcus lactis* ssp. *lactis* IL1403. Genome Res **11:**731-753.
15. **Wegmann U, O'Connell-Motherway M, Zomer A, Buist G, Shearman C, Canchaya C, Ventura M, Goesmann A, Gasson MJ, Kuipers OP, van Sinderen D, Kok J.** 2007. Complete Genome Sequence of the Prototype Lactic Acid Bacterium *Lactococcus lactis* subsp. *cremoris* MG1363. J Bacteriol **189:**3256-3270.
16. **Kato H, Shiwa Y, Oshima K, Machii M, Araya-Kojima T, Zendo T, Shimizu-Kadota M, Hattori M, Sonomoto K, Yoshikawa H.** 2012. Complete genome sequence

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.

17. **Makarova K, Slesarev A, Wolf Y, Sorokin A, Mirkin B, Koonin E, Pavlov A, Pavlova N, Karamychev V, Polouchine N, Shakhova V, Grigoriev I, Lou Y, Rohksar D, Lucas S, Huang K, Goodstein DM, Hawkins T, Plengvidhya V, Welker D, Hughes J, Goh Y, Benson A, Baldwin K, Lee JH, Diaz-Muniz I, Dosti B, Smeianov V, Wechter W, Barabote R, Lorca G, Altermann E, Barrangou R, Ganesan B, Xie Y, Rawsthorne H, Tamir D, Parker C, Breidt F, Broadbent J, Hutkins R, O'Sullivan D, Steele J, Unlu G, Saier M, Klaenhammer T, Richardson P, Kozyavkin S, Weimer B, Mills D.** 2006. Comparative genomics of the lactic acid bacteria. Proc Natl Acad Sci U S A **103:**15611-15616.

18. **Siezen RJ, Bayjanov J, Renckens B, Wels M, van Hijum SA, Molenaar D, van Hylckama Vlieg JE.** 2010. Complete genome sequence of *Lactococcus lactis* subsp. *lactis* KF147, a plant-associated lactic acid bacterium. J Bacteriol **192:**2649-2650.

19. **Linares DM, Kok J, Poolman B.** 2010. Genome sequences of *Lactococcus lactis* MG1363 (revised) and NZ9000 and comparative physiological studies. J Bacteriol **192:**5806-5812.

20. **Gao Y, Lu Y, Teng KL, Chen ML, Zheng HJ, Zhu YQ, Zhong J.** 2011. Complete Genome Sequence of *Lactococcus lactis* subsp. *lactis* CV56, a Probiotic Strain Isolated from the Vaginas of Healthy Women. J Bacteriol **193:**2886-2887.

21. **Bolotin A, Quinquis B, Ehrlich SD, Sorokin A.** 2012. Complete genome sequence of *Lactococcus lactis* subsp. *cremoris* A76. J Bacteriol **194:**1241-1242.

22. **Ainsworth S, Zomer A, de Jager V, Bottacini F, van Hijum S, Mahony J, van Sinderen D.** 2013. Complete Genome of *Lactococcus lactis* subsp. *cremoris* UC509.9, Host for a Model Lactococcal P335 Bacteriophage. Genome Announc **1**.

23. **Kelly WJ, Altermann E, Lambie SC, Leahy SC.** 2013. Interaction between the genomes of *Lactococcus lactis* and phages of the P335 species. Front Microbiol **4:**257.

24. **Yang X, Wang Y, Huo G.** 2013. Complete Genome Sequence of *Lactococcus lactis* subsp. *lactis* KLDS4.0325. Genome Announc **1:**e00962-00913.

25. **Oliveira LC, Saraiva TD, Soares SC, Ramos RT, Sa PH, Carneiro AR, Miranda F, Freire M, Renan W, Junior AF, Santos AR, Pinto AC, Souza BM, Castro CP, Diniz CA, Rocha CS, Mariano DC, de Aguiar EL, Folador EL, Barbosa EG, Aburjaile FF, Goncalves LA, Guimaraes LC, Azevedo M, Agresti PC, Silva RF, Tiwari S, Almeida SS, Hassan SS, Pereira VB, Abreu VA, Pereira UP, Dorella FA, Carvalho AF, Pereira FL, Leal CA, Figueiredo HC, Silva A, Miyoshi A, Azevedo V.** 2014. Genome Sequence of *Lactococcus lactis* subsp. *lactis* NCDO 2118, a GABA-Producing Strain. Genome Announc **2**.

26. **McCulloch JA, de Oliveira VM, de Almeida Pina AV, Pérez-Chaparro PJ, de Almeida LM, de Vasconcelos JM, de Oliveira LF, da Silva DEA, Rogez HLG, Cretenet M, Mamizuka EM, Nunes MRT.** 2014. Complete Genome Sequence of *Lactococcus lactis* Strain AI06, an Endophyte of the Amazonian Açaí Palm. Genome Announc **2**.

27. **Kelleher P, Bottacini F, Mahony J, Kilcawley KN, van Sinderen D.** 2017. Comparative and functional genomics of the *Lactococcus lactis* taxon; insights into evolution and niche adaptation. BMC Genomics **18:**267.

28. **Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL.** 2009. BLAST+: architecture and applications. BMC Bioinformatics **10:**421.
29. **Contreras-Moreira B, Vinuesa P.** 2013. GET_HOMOLOGUES, a versatile software package for scalable and robust microbial pangenome analysis. Appl Environ Microbiol **79:**7696-7701.
30. **Zhao Y, Jia X, Yang J, Ling Y, Zhang Z, Yu J, Wu J, Xiao J.** 2014. PanGP: a tool for quickly analyzing bacterial pan-genome profile. Bioinformatics **30:**1297-1299.
31. **Enright AJ, Van Dongen S, Ouzounis CA.** 2002. An efficient algorithm for large-scale detection of protein families. Nucleic Acids Res **30:**1575-1584.
32. **Warnes GR, Bolker B, Bonebakker L, Gentleman R, Huber W, Liaw A, Lumley T, Maechler M, Magnusson A, Moeller S, Schwartz M, Venables B.** 2009. gplots: Various R programming tools for plotting data, http://cran.r-project.org/web/packages/gplots/index.html.
33. **Galili T.** 2015. dendextend: an R package for visualizing, adjusting and comparing trees of hierarchical clustering. Bioinformatics **31:**3718-3720.
34. **Wright E.** 2016. Using DECIPHER v2.0 to Analyze Big Biological Sequence Data in R. The R Journal **8:**352-359.
35. **Venables WN, Ripley BD.** 2002. Modern Applied Statistics with R, Fourth ed. Springer, New York.
36. **Salvador S, Chan P.** 2004. Determining the number of clusters/segments in hierarchical clustering/segmentation algorithms, abstr Proceedings of the 16th IEEE International Conference on Tools with Artificial Intelligence, IEEE Computer Society, 1033653.
37. **Zhang J, Kobert K, Flouri T, Stamatakis A.** 2014. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. Bioinformatics **30:**614-620.
38. **Rognes T, Flouri T, Nichols B, Quince C, Mahe F.** 2016. VSEARCH: a versatile open source tool for metagenomics. PeerJ **4:**e2584.
39. **Edgar RC.** 2013. UPARSE: highly accurate OTU sequences from microbial amplicon reads. Nat Methods **10:**996-998.
40. **Nguyen N-P, Warnow T, Pop M, White B.** 2016. A perspective on 16S rRNA operational taxonomic unit clustering using sequence similarity. **2:**16004.
41. **Konstantinidis KT, Tiedje JM.** 2005. Genomic insights that advance the species definition for prokaryotes. Proceedings of the National Academy of Sciences of the United States of America **102:**2567-2572.
42. **Frantzen C, Kleppen HP, Holo H.** 2016. 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. Int Dairy J **53:**45-50.
43. **Cavanagh D, Casey A, Altermann E, Cotter PD, Fitzgerald GF, McAuliffe O.** 2015. Evaluation of *Lactococcus lactis* Isolates from Nondairy Sources with Potential Dairy Applications Reveals Extensive Phenotype-Genotype Disparity and Implications for a Revised Species. Applied and Environmental Microbiology **81:**3961-3972.
44. **Xu H, Sun Z, Liu W, Yu J, Song Y, Lv Q, Zhang J, Shao Y, Menghe B, Zhang H.** 2014. Multilocus sequence typing of *Lactococcus lactis* from naturally fermented milk foods in ethnic minority areas of China. J Dairy Sci **97:**2633-2645.
45. **Kilstrup M, Martinussen J.** 1998. A transcriptional activator, homologous to the *Bacillus subtilis* PurR repressor, is required for expression of purine biosynthetic genes in *Lactococcus lactis*. J Bacteriol **180:**3907-3916.

46. **Dabour N, LaPointe G.** 2005. Identification and Molecular Characterization of the Chromosomal Exopolysaccharide Biosynthesis Gene Cluster from *Lactococcus lactis* subsp. *cremoris* SMQ-461. Appl Environ Microbiol **71:**7414-7425.
47. **Forde A, Fitzgerald GF.** 2003. Molecular organization of exopolysaccharide (EPS) encoding genes on the lactococcal bacteriophage adsorption blocking plasmid, pCI658. Plasmid **49:**130-142.
48. **Godon J-J, Delorme C, Ehrlich SD, Renault P.** 1992. Divergence of Genomic Sequences between *Lactococcus lactis* subsp. *lactis* and *Lactococcus lactis* subsp. *cremoris*. Applied and Environmental Microbiology **58:**4045-4047.
49. **Nomura M, Kobayashi M, Narita T, Kimoto-Nira H, Okamoto T.** 2006. Phenotypic and molecular characterization of *Lactococcus lactis* from milk and plants. J Appl Microbiol **101:**396-405.
50. **Kelly WJ, Ward LJH, Leahy SC.** 2010. Chromosomal diversity in *Lactococcus lactis* and the origin of dairy starter cultures. Genome Biol Evol **2**.
51. **Kütahya OE, Starrenburg MJC, Rademaker JLW, Klaassen CHW, van Hylckama Vlieg JET, Smid EJ, Kleerebezem M.** 2011. High-Resolution Amplified Fragment Length Polymorphism Typing of *Lactococcus lactis* Strains Enables Identification of Genetic Markers for Subspecies-Related Phenotypes. Applied and Environmental Microbiology **77:**5192-5198.
52. **Passerini D, Beltramo C, Coddeville M, Quentin Y, Ritzenthaler P, Daveran-Mingot M-L, Le Bourgeois P.** 2010. Genes but Not Genomes Reveal Bacterial Domestication of *Lactococcus Lactis*. PLOS ONE **5:**e15306.
53. **Frantzen CA, Kot W, Pedersen TB, Ardö YM, Broadbent JR, Neve H, Hansen LH, Dal Bello F, Østlie HM, Kleppen HP, Vogensen FK, Holo H.** 2017. Genomic Characterization of Dairy Associated *Leuconostoc* Species and Diversity of Leuconostocs in Undefined Mixed Mesophilic Starter Cultures. Front Microbiol **8**.
54. **Smit G, Smit BA, Engels WJ.** 2005. Flavour formation by lactic acid bacteria and biochemical flavour profiling of cheese products. FEMS Microbiol Rev **29:**591-610.
55. **Stackebrandt E, GoebelL BM.** 1994. Taxonomic Note: A Place for DNA-DNA Reassociation and 16S rRNA Sequence Analysis in the Present Species Definition in Bacteriology. Int J Syst Evol Microbiol **44:**846-849.
56. **Edgar RC.** 2017. Updating the 97% identity threshold for 16S ribosomal RNA OTUs. bioRxiv doi:10.1101/192211.
57. **Edgar RC.** 2016. UNOISE2: improved error-correction for Illumina 16S and ITS amplicon sequencing. bioRxiv doi:10.1101/081257.
58. **Mahé F, Rognes T, Quince C, de Vargas C, Dunthorn M.** 2014. Swarm: robust and fast clustering method for amplicon-based studies. PeerJ **2:**e593.
59. **Ndoye B, Rasolofo EA, Gisele LaPointe G, Roy D.** 2011. A review of the molecular approaches to investigate the diversity and activity of cheese microbiota. Dairy Sci Technol **91**.
60. **Palys T, Nakamura LK, Cohan FM.** 1997. Discovery and classification of ecological diversity in the bacterial world: the role of DNA sequence data. Int J Syst Bacteriol **47:**1145-1156.
61. **Deveau H, Van Calsteren M-R, Moineau S.** 2002. Effect of Exopolysaccharides on Phage-Host Interactions in *Lactococcus lactis*. Appl Environ Microbiol **68:**4364-4369.

62. **Forde A, Fitzgerald GF.** 1999. Analysis of exopolysaccharide (EPS) production mediated by the bacteriophage adsorption blocking plasmid, pCI658, isolated from *Lactococcus lactis* ssp. *cremoris* HO2. Int Dairy J **9:**465-472.
63. **Moineau S, Borkaev M, Holler BJ, Walker SA, Kondo JK, Vedamuthu ER, Vandenbergh PA.** 1996. Isolation and Characterization of Lactococcal Bacteriophages from Cultured Buttermilk Plants in the United States. J Dairy Sci **79:**2104-2111.

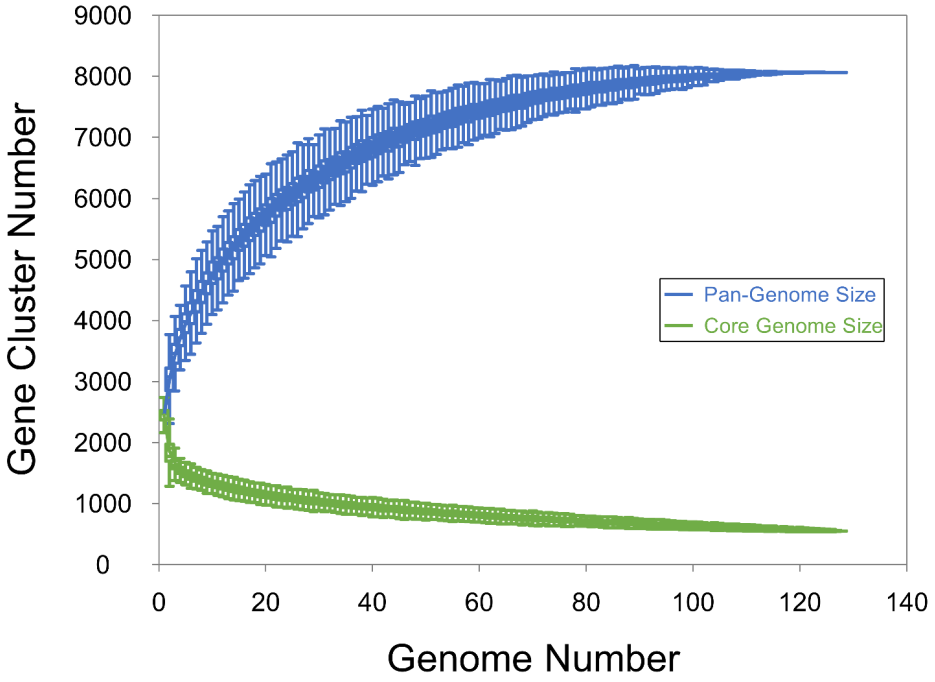


Figure 1: Pan- and core-genome sizes for 127 *Lactococcus lactis* isolates using PanGP and the traverse all approach was estimated at 8064, and 551, respectively.

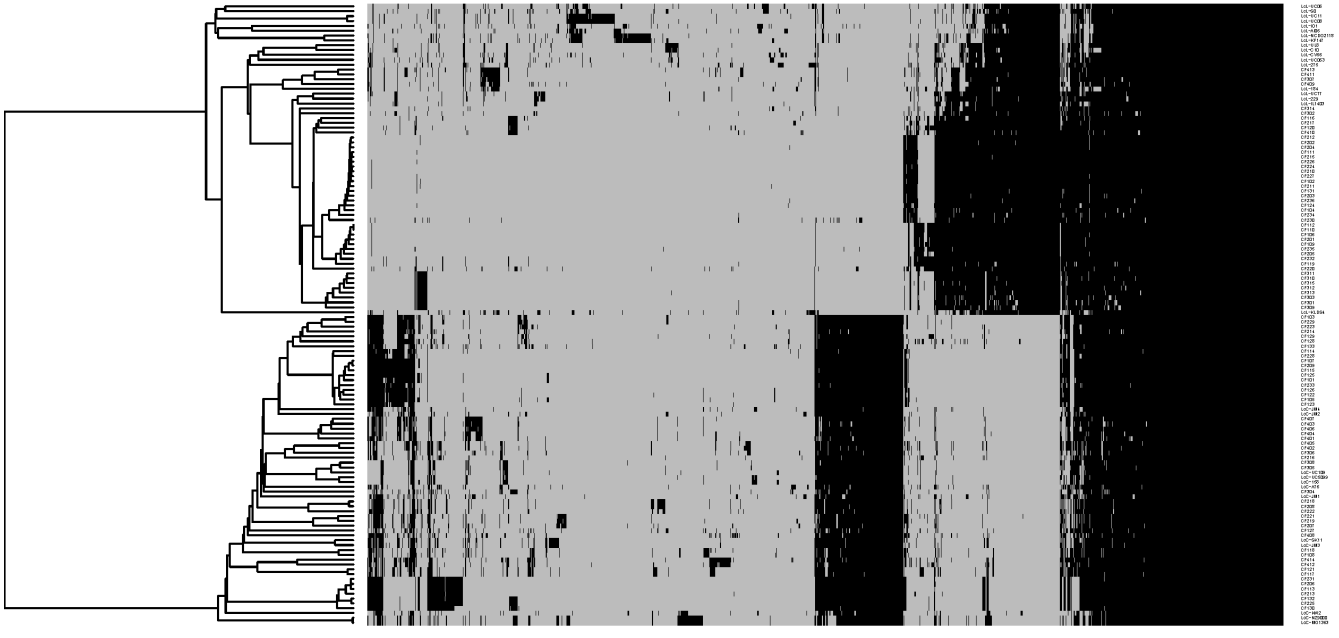


Figure 2: Heatmap representation of the pan-genome content of 127 *Lactococcus lactis* isolates. The black regions indicate orthologous groups that are present, while the grey regions indicate orthologous groups that are absent. Using hierarchal single-linkage clustering, with a distance cut-off determined 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.

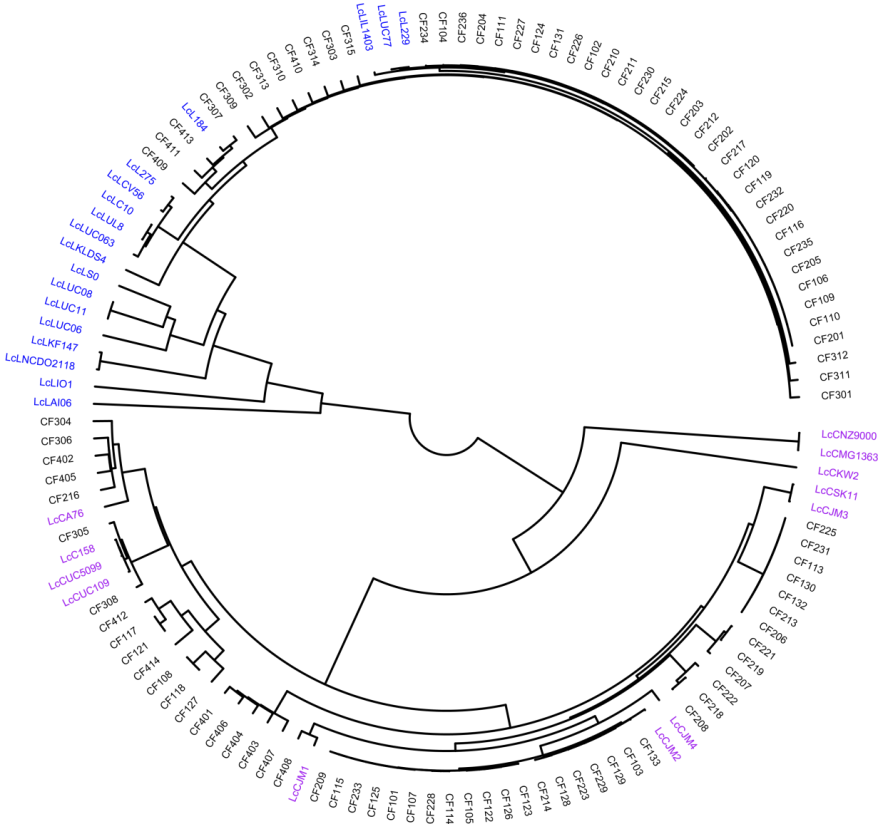


Figure 3: Differentiation of *Lactococcus lactis* using 551 core-genes.

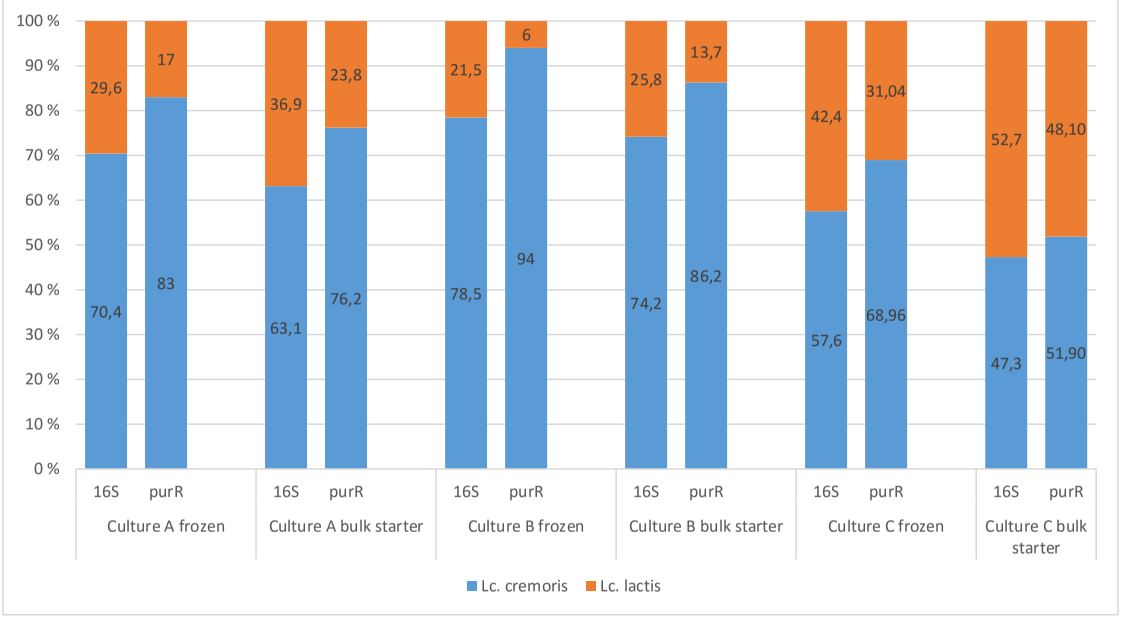


Figure 4: Comparison of the quantification of *Lactococcus lactis* subsp. in starter cultures using the 16S rDNA and *purR* loci. Compared to *purR,* 16S underreports the relative amount of subsp. *cremoris* compared to subsp. *lactis*.

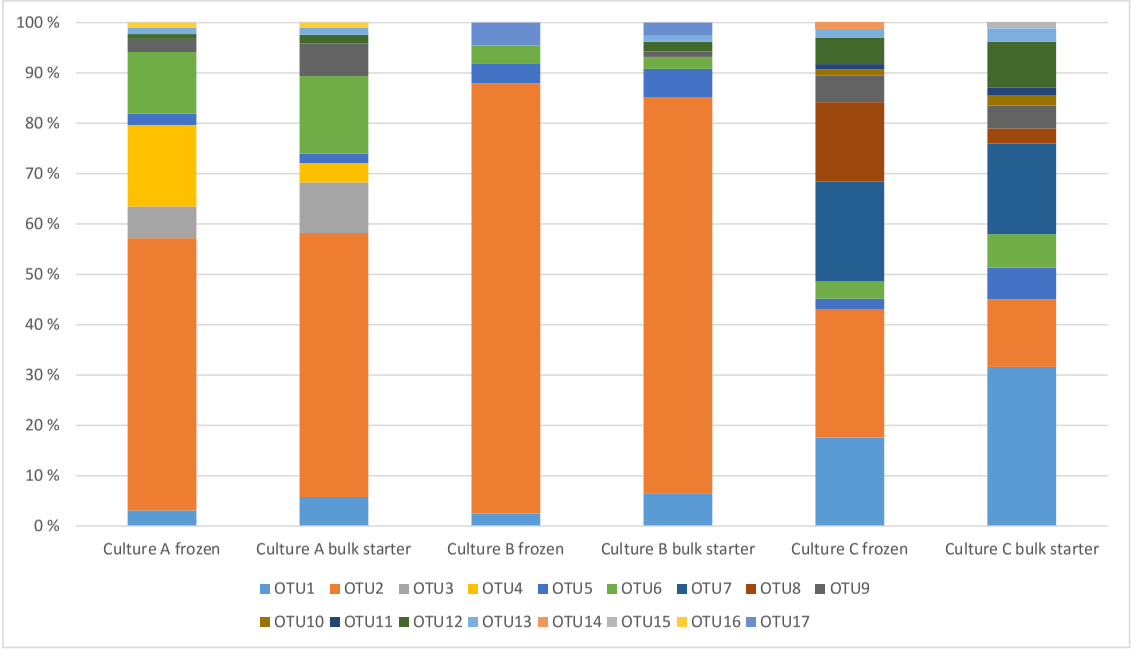


Figure 5: *Lactococcus lactis* diversity in three commercial starter cultures A, B, C using targeted-amplicon sequencing of *purR*. OTU1, 3, 10, 12 and 14 were identified as *L*. *lactis* subspecies *lactis*, while OTU2, 4, 5, 6, 7, 8, 9, 11, 13, 15, 16, 17 were identified as *L*. *lactis* subspecies *cremoris*.

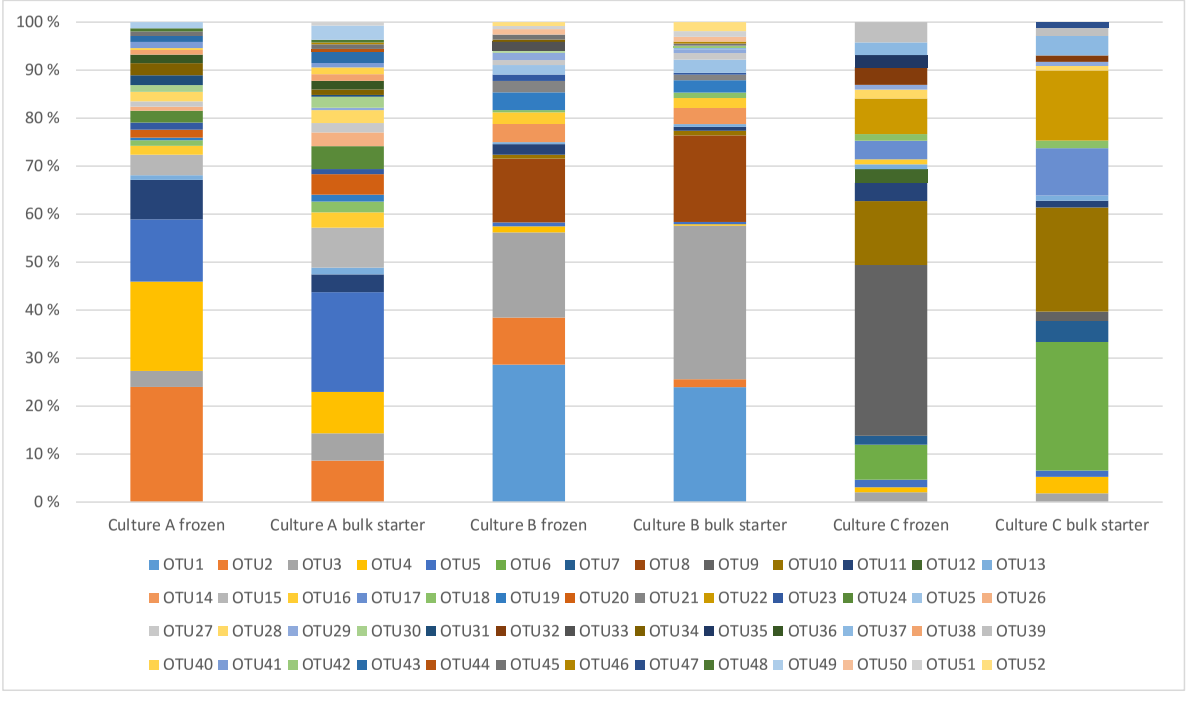


Figure 6: *Lactococcus lactis* diversity and relative abundance in starter cultures using targeted-amplicon sequencing of *epsD*. The OTUs identified as L. *lactis* subspecies *lactis* were OTU1, 3, 4, 7, 8, 15, 20, 22, 24, 27 and 50. The OTUs identified as *L. lactis* subspecies *cremoris* were OTU2, 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.

	Culture A		**Culture B**		**Culture C**	
	Frozen	Bulk	Frozen	Bulk	Frozen	Bulk
L. lactis* subsp. *cremoris	58.8	47.6	77.9	73.8	47.0	33.4
L. lactis* subsp. *lactis	24.7	27.8	21.4	25.7	34.6	37.2
***Leuconostoc* spp.**	16.6	24.6	0.8	0.5	18.4	29.4

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.

	Culture A		**Culture B**		**Culture C**	
#OTU ID	**Frozen**	**Bulk**	**Frozen**	**Bulk**	**Frozen**	**Bulk**
OTU1[a]	3	5,5	2,3	6,2	16,9	30,5
OTU2[b]	52	50,4	81,4	75	24,6	13
OTU3[a]	6,1	9,7	0	0	0	0
OTU4[b]	15,6	3,7	0	0	0	0
OTU5[b]	2,2	1,8	3,8	5,4	2	6,1
OTU6[b]	11,8	14,8	3,4	2,3	3,3	6,4
OTU7[b]	0	0	0	0	19,1	17,4
OTU8[b]	0	0	0,9	0	15,2	2,9
OTU9[b]	2,5	6,2	0	1	5,1	4,4
OTU10[a]	0	0	0	0	1,3	1,9
OTU11[b]	0	0	0	0	1	1,6
OTU12[a]	0,9	1,7	0	1,8	5	8,8
OTU13[b]	1,2	1,3	0	1,2	1,8	2,5
OTU14[a]	0	0	0	0	1,2	0
OTU15[b]	0	0	0	0	0	1,1
OTU16[b]	1	1	0	0	0	0
OTU17[b]	0	0	4,3	2,5	0	0

[a] OTUs identified as *L*. *lactis* subspecies *lactis*.

[b] 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.

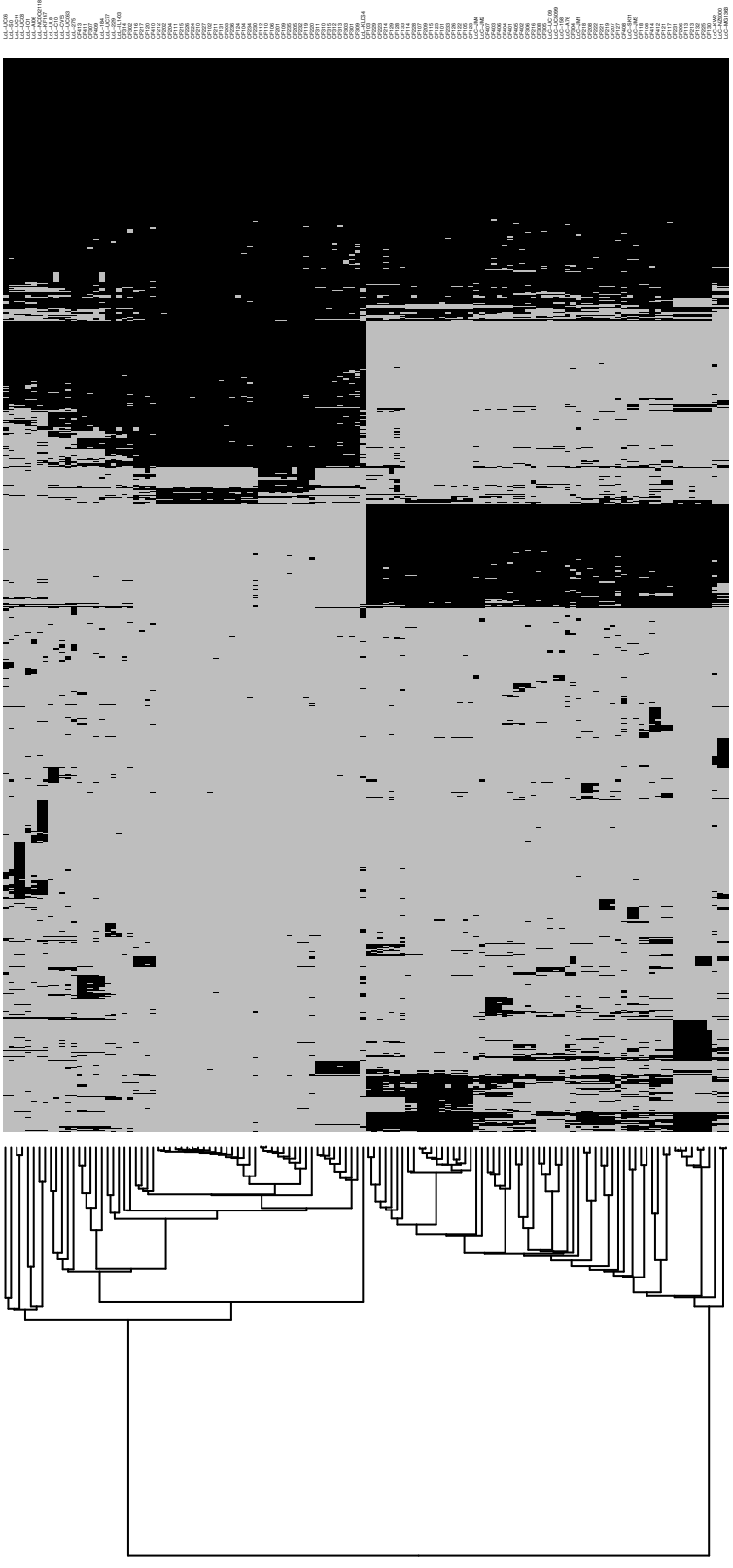
	Culture A		**Culture B**		**Culture C**	
#OTU ID	**Frozen**	**Bulk**	**Frozen**	**Bulk**	**Frozen**	**Bulk**
OTU01[a]	0.0	0.0	28.6	23.9	0.0	0.0
OTU02[b]	24.0	8.6	9.8	1.6	0.0	0.0
OTU03[a]	3.3	5.6	17.7	32.0	2.1	1.8
OTU04[a]	18.6	8.6	1.3	0.3	1.0	3.5
OTU05[b]	13.0	20.7	0.8	0.5	1.6	1.3
OTU06[c]	0.0	0.0	0.0	0.0	7.3	26.8
OTU07[a]	0.0	0.0	0.0	0.0	1.9	4.4
OTU08[a]	0.0	0.0	13.3	18.0	0.0	0.0
OTU09[c]	0.0	0.0	0.0	0.0	35.6	2.0
OTU10[c]	0.0	0.0	0.8	1.0	13.3	21.7
OTU11[b]	8.2	3.7	2.2	0.8	3.8	1.4
OTU12[c]	0.0	0.0	0.0	0.0	2.8	0.0
OTU13[b]	1.0	1.4	0.4	0.6	1.0	1.1
OTU14[b]	0.0	0.0	3.8	3.4	0.0	0.0
OTU15[a]	4.3	8.3	0.0	0.0	0.0	0.0
OTU16[b]	1.9	3.2	2.4	2.1	1.0	0.0
OTU17[c]	0.0	0.0	0.0	0.0	3.9	9.8
OTU18[b]	1.1	2.2	0.5	1.2	1.4	1.6
OTU19[b]	0.5	1.4	3.7	2.6	0.0	0.0
OTU20[a]	1.7	4.3	0.0	0.0	0.0	0.0
OTU21[c]	0.0	0.0	2.4	1.2	0.0	0.0
OTU22[a]	0.0	0.0	0.0	0.0	7.4	14.6
OTU23[b]	1.5	1.1	1.3	0.4	0.0	0.0
OTU24[a]	2.5	4.7	0.0	0.0	0.0	0.0
OTU25[b]	0.0	0.5	2.0	2.8	0.0	0.0
OTU26[c]	0.9	2.8	0.0	0.0	0.0	0.0
OTU27[a]	1.1	1.9	1.0	1.4	0.0	0.0
OTU28[c]	2.0	2.7	0.0	0.0	1.8	0.9
OTU29[b]	0.0	0.5	1.5	0.8	1.0	0.9
OTU30[c]	1.4	2.3	0.3	0.0	0.0	0.0
OTU31[b]	2.1	0.4	0.0	0.0	0.0	0.0
OTU32[c]	0.0	0.0	0.0	0.0	3.7	1.3
OTU33[c]	0.0	0.0	1.9	0.0	0.0	0.0
OTU34[b]	2.5	1.1	0.5	0.0	0.0	0.0
OTU35[b]	0.0	0.0	0.0	0.0	2.5	0.0
OTU36[b]	1.8	1.8	0.0	0.0	0.0	0.0
OTU37[c]	0.0	0.0	0.0	0.4	2.6	4.1

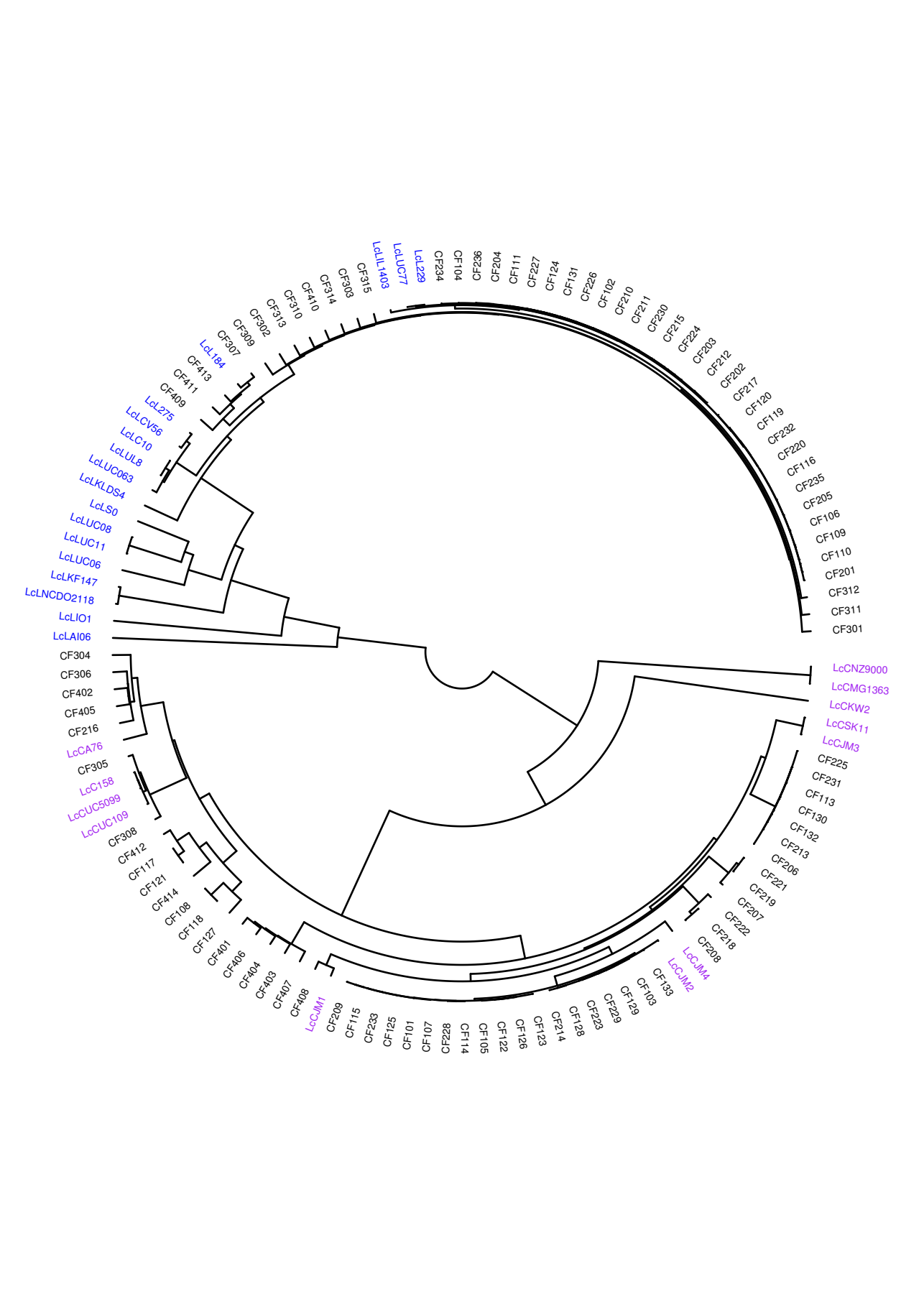
OTU38[b]	1.0	1.4	0.0	0.0	0.0	0.0
OTU39[c]	0.0	0.0	0.0	0.0	4.2	1.7
OTU40[c]	0.4	1.4	0.0	0.0	0.0	0.0
OTU41[b]	1.3	0.9	0.0	0.0	0.0	0.0
OTU42[c]	0.0	0.0	0.0	0.3	0.0	0.0
OTU43[c]	1.2	2.3	0.0	0.0	0.0	0.0
OTU44[c]	0.0	0.6	0.0	0.0	0.0	0.0
OTU45[b]	0.9	1.0	1.0	0.4	0.0	0.0
OTU46[b]	0.0	0.4	0.0	0.4	0.0	0.0
OTU47[c]	0.0	0.0	0.0	0.0	0.0	1.2
OTU48[c]	0.7	0.6	0.0	0.0	0.0	0.0
OTU49[c]	1.3	3.0	0.0	0.0	0.0	0.0
OTU50[a]	0.0	0.0	1.1	1.0	0.0	0.0
OTU51[b]	0.0	0.7	0.7	1.2	0.0	0.0
OTU52[c]	0.0	0.0	0.8	1.9	0.0	0.0

[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.




LcLIL1403
LcLUC77
LcL229
CF234
CF104
CF236
CF204
CF111
CF227
CF124
CF131
CF226
CF102
CF210
CF211
CF230
CF215
CF224
CF203
CF212
CF202
CF217
CF120
CF119
CF232
CF220
CF116
CF235
CF205
CF106
CF109
CF110
CF201
CF312
CF311
CF301
LcCNZ9000
LcCMG1363
LcCKW2
LcCSK11
LcCJM3
CF225
CF231
CF113
CF130
CF132
CF213
CF206
CF221
CF219
CF207
CF222
CF218
CF208
LcCJM4
LcCJM2
CF133
CF103
CF129
CF229
CF223
CF128
CF214
CF123
CF126
CF122
CF105
CF114
CF228
CF107
CF101
CF125
CF233
CF115
CF209
LcCJM1
CF408
CF407
CF403
CF404
CF406
CF401
CF127
CF118
CF106
CF414
CF121
CF117
CF412
CF308
LcCUC109
LcCUC5099
LcC158
CF305
LcCA76
CF216
CF405
CF402
CF306
CF304
LcLAI06
LcLIO1
LcLNCDO2118
LcLKF147
LcLUC06
LcLUC11
LcLUC08
LcLS0
LcLKLDS4
LcLUC063
LcLUL8
LcLC10
LcLCV56
LcL275
CF409
CF411
CF413
LcL184
CF307
CF309
CF302
CF313
CF310
CF410
CF314
CF303
CF315