frontiers in Microbiology

Genomic characterization of dairy associated Leuconostoc species and diversity of leuconostocs in undefined mixed mesophilic starter cultures

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

[1]Department of Chemistry, Biotechnology and Food Science, Norwegian University of Life Sciences, Norway, [2]TINE SA, Norway, [3]Department of Environmental Science, Aarhus University, Denmark, [4]Department of Food Science, University of Copenhagen, Denmark, [5]Department of Nutrition, Dietetics and Food Sciences, Utah State University, USA, [6]Department of Microbiology and Biotechnology, Max Rubner-Institut, Germany, [7]Scuola di Scienze del Farmaco e dei Prodotti della Salute, Universita di Camerino, Italy, [8]ACD Pharmaceuticals AS, Norway

Submitted to Journal:
Frontiers in Microbiology

Specialty Section:
Food Microbiology

ISSN:
1664-302X

Article type:
Original Research Article

Received on:
17 Nov 2016

Accepted on:
18 Jan 2017

Provisional PDF published on:
18 Jan 2017

Frontiers website link:
www.frontiersin.org

Citation:
Frantzen CA, Kot W, Pedersen TB, Ardo YM, Broadbent JR, Neve H, Hansen LH, Dal_bello F, Ostlie 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

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 Ardö[3], Jeff R. Broadbent[4], Horst Neve[5], Lars H. Hansen[2], Fabio Dal Bello[6], Hilde M. Østlie[1], Hans Petter Kleppen[1,7], Finn Vogensen[3], Helge Holo[1, 8, *]

[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]Department of Environmental Science, Aarhus University, Frederiksborgsvej 399, Roskilde, Denmark

[3]Department of Food Science, University of Copenhagen, Denmark

[4]Department of Nutrition, Dietetics and Food Sciences, Utah State University, United States of America

[5]Department of Microbiology and Biotechnology, Max Rubner-Institut, Kiel, Germany

[6]Scuola di Scienze del Farmaco e dei Prodotti della Salute, Universita di Camerino, Via S. Agostino 1, 62032 Camerino, Italy

[7]ACD Pharmaceuticals AS, Leknes, Norway

[8]TINE SA, Oslo, Norway

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

Abstract

Undefined mesophilic mixed (DL-type) starter cultures are composed of predominantly *Lactococcus lactis* subspecies and 1 to 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 fifty 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 diary-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.

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 to 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 (Ardo 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 thirteen 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; Bjorkroth et al., 2000; Cibik et al., 2000; Perez et al., 2002; Sanchez 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; Garvie et al., 1974; Garvie, 1979; 1983), 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., 2014b; a; 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.

Method and Materials

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 culture (A, B, C, D, and E) were suspended in MPCA to an optical density at 600nm (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 added 15 % (w/v) glycerol (Sigma-Aldrich) and stored at -70°C.

Genome sequencing, assembly and annotation

Genomic DNA from *Leuconostoc* isolates was extracted from 1mL of overnight culture using Qiagen DNeasy Blood & Tissue Kit (Qiagen, Hilden, Germany). The cells were lysed with 40mg/mL lysozyme (Qiagen, Hilden, Germany) and bead-beating in a FastPrep®-24 (MP Biomedicals, Santa Ana, California) using 0.5 g acid-washed beads ($<10^6$ 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 less than 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 1mL 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 ($<10^6$ 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 (2x300 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 Mb – 1.8 Mb) compared to *Ln. mesenteroides*, *Ln. dextranicum*, and *Ln. pseudomesenteroides* (1.8 Mb – 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.*

198 *cremoris* isolates did not have any genetic functionality absent in *Ln. mesenteroides* or *Ln.*
199 *dextranicum*. Moreover, several truncated genes and deletions were found in *Ln. cremoris* isolates,
200 likely the result of a degenerative evolutionary process through a long period of growth in the milk
201 environment.

202

203 **Comparative genomics of intra-species *Leuconostoc* lineages**

204 To explore differences in functional genetic potential between the lineages within the species and
205 subspecies, comparative analysis of intra-lineage pan-genomes was performed.

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

213 **(II) *Ln. mesenteroides* and *Ln. dextranicum* lineages.** Comparison of the genetic content showed a
214 large variance between and within the *Ln. mesenteroides* lineages. Interestingly, no major
215 difference between subspecies *Ln. mesenteroides* and *Ln. dextranicum* was found. *Ln. dextranicum*
216 20484 is grouped together with *Ln. mesenteroides* isolates LN32 and LN34, while *Ln. dextranicum*
217 LbE16 is grouped together with *Ln. mesenteroides* LbE15 and LN08. This subspecies segregation
218 of *Ln. dextranicum* and *Ln. mesenteroides* was based on the phenotypical ability to produce dextran
219 from sucrose. Dextransucrase, the enzyme involved in this process, is a glucosyltransferase that
220 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 caseine (*prtP*) was identified in *Ln. mesenteroides* lineages M1 and M3, 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 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

245 the category of methionine biosynthesis, β-glucoside metabolism, sucrose metabolism, as well as an
246 additional lactate dehydrogenase was identified in *Ln. pseudomesenteroides* P4 but not P1, P2 and
247 P3. Moreover, P4 isolates were missing the genes for reduction of diacetyl to acetoin and 2,3-
248 butandiol, and contained genes for a different capsular and extracellular polysaccharide biosynthesis
249 pathway, compared to P1, P2 and P3 isolates.

250

251 **Genetic potential of *Leuconostoc.* (I) Amino acid biosynthesis.**

252 The amino acid requirements of leuconostocs have been described as highly variable between
253 strains. Glutamic acid and valine are required by most leuconostocs, methionine usually stimulates
254 growth, while no *Leuconostoc* are reported to require alanine (Garvie, 1967). Comparative analysis
255 of genes involved in amino acid biosynthesis showed that *Ln. cremoris* and *Ln. mesenteroides*
256 subsp. carried the genetic potential to produce a wide range of amino acids while *Ln. lactis* and *Ln.*
257 *pseudomesenteroides* did not (Table 2). This included genes encoding biosynthesis of histidine,
258 tryptophan, methionine and lysine. Studies on the amino acid requirement of leuconostocs show
259 that most the *Ln. mesenteroides* subspecies do require isoleucine and leucine to grow. The *ilv* and
260 *leu* operons involved in biosynthesis of the branched-chain amino acids isoleucine, leucine and
261 valine were present in all *Ln. mesenteroides* isolates, however both operons were truncated when
262 compared to functional *ilv* and *leu* operons from lactococci. The *leuA* gene in the *leuABCD* operon
263 is truncated in leuconostocs (391 aa) compared to lactococci (513 aa) likely resulting in an inactive
264 product and a nonfunctional pathway. This has been documented in the dairy strain *Lactococcus*
265 *lactis* IL1403 where a similar truncation of the *leuA* gene led to an inactivation of the leucine/valine
266 pathway (Godon et al., 1993). Likewise, the *ilv* operon of sequenced leuconostocs is missing the
267 *ilvD* gene, and has truncated *ilvA* and *ilvH* genes when compared to the lactococcal *ilv* operon. The
268 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 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-Heyerhof 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 1,200 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 diary associated *Ln. mesenteroides* or *Ln. cremoris* (<36 % identity). Genes coding for maltose-phosphorylase (*malP*) and sucrose-6-phosphate hydrolase (*scrB*) was 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*) was 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.

316

317 **(III) Citrate metabolism.** All the dairy strains in this study contained the genes necessary for
318 uptake and metabolism of citrate. These genes are found in an operon comprised of *citC* (citrate
319 lyase ligase), *citDEF* (citrate lyase), *citG* (holo-ACP synthase), *citO* (transcriptional regulator) and
320 *citS* (Na+ dependent citrate transporter). A citrate/malate transporter annotated *cimH* was present in
321 *Ln. mesenteroides* subspecies isolates, but was not present in any of the *Ln. lactis* or *Ln.*
322 *pseudomesenteroides* isolates. In the *Ln. cremoris* and *Ln. pseudomesenteroides* genomes, the *cit*
323 operon is flanked by two IS116/IS110/IS902 family transposases, suggesting it may have been
324 acquired by horizontal gene transfer. In these bacteria, the operon appears to be located on the
325 chromosome, a finding supported by the genome assembly, which organizes the *cit* operon on a
326 contig containing a number of essential genes, and by read coverage analysis that shows a
327 continuous gapless coverage through the contig, with no elevation in read coverage across the *cit*
328 operon. The *citCDEFGOS* operons of *Ln. mesenteroides* and *Ln. lactis*, however, appears to be
329 located on a plasmid, since in all cases they assembled on a contig, which includes a site of
330 replication and not essential genes. The *cit* operon is highly conserved in the *Ln. cremoris* and *Ln.*
331 *pseudomesenteroides* genomes with >97 % DNA sequence identity between all the isolates. The
332 likely to be plasmid-encoded cit operon found in *Ln. mesenteroides* and *Ln. lactis* genomes is also
333 highly conserved between the isolates (>99% identity), however it is significantly different from the
334 chromosomally encoded *cit* operon present in *Ln. cremoris* and *Ln. pseudomesenteroides* (50 % to
335 65 % DNA sequence identity for each gene). None of the strains of non-dairy origin included in this
336 study contained the citrate genes, indicating that the ability to metabolize citrate plays an important
337 role in the successful adaption to the milk environment.

338 **(IV) Proteolytic activity.** Leuconostocs grow in association with the lactococci in dairy
339 fermentations, and commonly grow poorly in milk without the presence of lactococci. The general

explanation for this poor growth is their lack of proteinase activity, making them dependent on small peptides from lactococcal proteinase activity. Screening all the isolates for genes involved in peptide and proteolytic activity revealed a number of differences between the lineages (Table 4). The genes encoding for the OppABCDF system were found in all *Leuconostoc* genomes. However, in *Ln. cremoris* genomes, the *oppA* gene was missing, and the *oppB* gene was severely truncated. A gene encoding for a PII-type serine proteinase (PrtP) best known for its action on caseins was found in all *Ln. pseudomesenteroides* genomes, *Ln. lactis* dairy 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, 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 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, homologues 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 lacks 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.

Conflict of Interest

The authors declare no conflict of interest.

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.

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 Science & Technology* 93(6)**,** 657-673. doi: 10.1007/s13594-013-0128-3.

Ardo, Y., and Varming, C. (2010). Bacterial influence on characteristic flavour of cheeses made with mesophilic DL-starter. *Australian Journal of Dairy Technology* 65(3)**,** 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. *Applied and Environmental Microbiology* 67(1)**,** 420-425. doi: 10.1128/aem.67.1.420-425.2001.

Bandell, M., Lhotte, M.E., Marty-Teysset, C., Veyrat, A., Prevost, H., Dartois, V., et al. (1998). Mechanism of the citrate transporters in carbohydrate and citrate cometabolism in *Lactococcus* and *Leuconostoc* species. *Applied and Environmental Microbiology* 64(5)**,** 1594-1600.

Barrangou, R., Yoon, S.S., Breidt Jr, F., Jr., Fleming, H.P., and Klaenhammer, T.R. (2002). Characterization of six *Leuconostoc fallax* bacteriophages isolated from an industrial sauerkraut fermentation. *Applied and Environmental Microbiology* 68(11)**,** 5452-5458.

Bjorkroth, 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. *Applied and Environmental Microbiology* 66(9)**,** 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. *Journal of Dairy Research* 71(2)**,** 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(1)**,** 421. doi: 10.1186/1471-2105-10-421.

Campedelli, I., Florez, A.B., Salvetti, E., Delgado, S., Orru, L., Cattivelli, L., et al. (2015). Draft Genome Sequence of Three Antibiotic-Resistant *Leuconostoc mesenteroides* Strains of Dairy Origin. *Genome Announcements* 3(5). 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 Science & Technology* 91(4)**,** 457-470. doi: 10.1007/s13594-011-0022-9.

Cavin, J.F., Dartois, V., Labarre, C., and Divies, C. (1999). Cloning of branched chain amino acid biosynthesis genes and assays of alpha-acetolactate synthase activities in *Leuconostoc mesenteroides* subsp. *cremoris*. *Research in Microbiology* 150(3)**,** 189-198.

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. *Systematic and Applied Microbiology* 23. doi: 10.1016/s0723-2020(00)80014-4.

Cogan, T.M., and Jordan, K.N. (1994). Metabolism of *Leuconostoc* Bacteria. *Journal of Dairy Science* 77(9)**,** 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(18)**,** 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. *Applied and Environmental Microbiology* 79(24)**,** 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 multi-locus sequence typing (MLST) protocol for *Leuconostoc lactis* isolates from traditional dairy products in China and Mongolia. *BMC Microbiology* 14(1)**,** 1-9. doi: 10.1186/1471-2180-14-150.

Drider, D., Bekal, S., and Prevost, H. (2004). Genetic organization and expression of citrate permease in lactic acid bacteria. *Genetics and Molecular Research* 3(2)**,** 273-281.

Edgar, R.C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26(19)**,** 2460-2461. doi: 10.1093/bioinformatics/btq461.

Edgar, R.C. (2013). UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nat Methods* 10(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 Research* 30(7)**,** 1575-1584.

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(11)**,** 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. *International Journal of Systematic and Evolutionary Microbiology* 39(3)**,** 279-283. doi: doi:10.1099/00207713-39-3-279.

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).

Garvie, E.I. (1960). The genus *Leuconostoc* and its nomenclature. *Journal of Dairy Research* 27(02)**,** 283-292. doi: 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(3)**,** 439-447. doi: doi:10.1099/00221287-48-3-439.

Garvie, E.I. (1969). Lactic dehydrogenases of strains of the genus *Leuconostoc*. *Microbiology* 58(1)**,** 85-94. doi: 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. *International Journal of Systematic and Evolutionary Microbiology* 29(2)**,** 149-151. doi: doi:10.1099/00207713-29-2-149.

Garvie, E.I. (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(1)**,** 118-119. doi: doi:10.1099/00207713-33-1-118.

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.

International Journal of Systematic and Evolutionary Microbiology 24(2)**,** 248-251. doi: 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. *Journal of Bacteriology* 175(14)**,** 4383-4390.

Grissa, I., Vergnaud, G., and Pourcel, C. (2007). CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Research* 35(Web Server issue)**,** W52-57. 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. *International Journal of Systematic and Evolutionary Microbiology* 62(7)**,** 1548-1551. doi: 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*. *Journal of Biological Chemistry* 275(44)**,** 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*. *Systematic and Applied Microbiology* 22(4)**,** 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. *International Dairy Journal* 14. doi: 10.1016/j.idairyj.2003.10.005.

Hugenholtz, J. (1993). Citrate metabolism in lactic acid bacteria. *FEMS Microbiology Reviews* 12(1)**,** 165-178. doi: http://dx.doi.org/10.1016/0168-6445(93)90062-E.

Johansen, E., and Kibenich, A. (1992). Characterization of *Leuconostoc* Isolates from Commercial Mixed Strain Mesophilic Starter Cultures. *Journal of Dairy Science* 75(5)**,** 1186-1191. doi: 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. *Journal of Bacteriology* 194(3)**,** 730-731. doi: 10.1128/jb.06498-11.

Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Research* 28(1)**,** 27-30.

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(12)**,** 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. *Applied and Environmental Microbiology* 78(18)**,** 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. *Applied Microbiology and Biotechnology* 34(5)**,** 628-631. doi: 10.1007/bf00167912.

Lodics, T.A., and Steenson, L.R. (1990). Characterization of Bacteriophages and Bacteria Indigenous to a Mixed-Strain Cheese Starter. *Journal of Dairy Science* 73(10)**,** 2685-2696.

673 Magni, C., de Felipe, F.L., Sesma, F., López, P., and de Mendoza, D. (1994). Citrate transport in
674 *Lactococcus lactis* biovar *diacetylactis*: Expression of the plasmid-borne citrate permease
675 P. *FEMS Microbiology Letters* 118(1-2)**,** 75-82. doi: 10.1111/j.1574-
676 6968.1994.tb06806.x.
677 Marty-Teysset, C., Posthuma, C., Lolkema, J.S., Schmitt, P., Divies, C., and Konings, W.N.
678 (1996). Proton motive force generation by citrolactic fermentation in *Leuconostoc*
679 *mesenteroides*. *Journal of Bacteriology* 178(8)**,** 2178-2185.
680 Meslier, V., Loux, V., and Renault, P. (2012). Genome sequence of *Leuconostoc*
681 *pseudomesenteroides* strain 4882, isolated from a dairy starter culture. *Journal of*
682 *Bacteriology* 194(23)**,** 6637. doi: 10.1128/jb.01696-12.
683 Monnet, C., Phalip, V., Schmitt, P., and Diviès, C. (1994). Comparison of α-acetolactate
684 synthase and α-acetolactate decarboxylase in *Lactococcus* spp. and *Leuconostoc* spp.
685 *Biotechnology Letters* 16(3)**,** 257-262. doi: 10.1007/bf00134622.
686 Nieto-Arribas, P., Seseña, S., Poveda, J.M., Palop, L., and Cabezas, L. (2010). Genotypic and
687 technological characterization of *Leuconostoc* isolates to be used as adjunct starters in
688 Manchego cheese manufacture. *Food Microbiology* 27. doi: 10.1016/j.fm.2009.08.006.
689 Nurk, S., Bankevich, A., Antipov, D., Gurevich, A.A., Korobeynikov, A., Lapidus, A., et al.
690 (2013). Assembling single-cell genomes and mini-metagenomes from chimeric MDA
691 products. *Journal of Computational Biology* 20(10)**,** 714-737. doi:
692 10.1089/cmb.2013.0084.
693 Obst, M., Meding, E.R., Vogel, R.F., and Hammes, W.P. (1995). Two genes encoding the beta-
694 galactosidase of *Lactobacillus sakei*. *Microbiology* 141 (Pt 12)(12)**,** 3059-3066. doi:
695 10.1099/13500872-141-12-3059.
696 Olsen, K.N., Brockmann, E., and Molin, S. (2007). Quantification of *Leuconostoc* populations in
697 mixed dairy starter cultures using fluorescence in situ hybridization. *Journal of Applied*
698 *Microbiology* 103(4)**,** 855-863. doi: 10.1111/j.1365-2672.2007.03298.x.
699 Overbeek, R., Olson, R., Pusch, G.D., Olsen, G.J., Davis, J.J., Disz, T., et al. (2014). The SEED
700 and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST).
701 *Nucleic Acids Research* 42(Database issue)**,** D206-214. doi: 10.1093/nar/gkt1226.
702 Pedersen, T.B., Kot, W.P., Hansen, L.H., Sorensen, S.J., Broadbent, J.R., Vogensen, F.K., et al.
703 (2014a). Genome Sequence of *Leuconostoc mesenteroides* subsp. *cremoris* Strain T26,
704 Isolated from Mesophilic Undefined Cheese Starter. *Genome Announc* 2(3). doi:
705 10.1128/genomeA.00485-14.
706 Pedersen, T.B., Kot, W.P., Hansen, L.H., Sorensen, S.J., Broadbent, J.R., Vogensen, F.K., et al.
707 (2014b). Genome Sequences of Two *Leuconostoc pseudomesenteroides* Strains Isolated
708 from Danish Dairy Starter Cultures. *Genome Announc* 2(3). doi:
709 10.1128/genomeA.00484-14.
710 Pedersen, T.B., Vogensen, F.K., and Ardo, Y. (2016). Effect of heterofermentative lactic acid
711 bacteria of DL-starters in initial ripening of semi-hard cheese. *International Dairy*
712 *Journal* 57**,** 72-79. doi: 10.1016/j.idairyj.2016.02.041.
713 Perez, G., Cardell, E., and Zarate, V. (2002). Random amplified polymorphic DNA analysis for
714 differentiation of *Leuconostoc mesenteroides* subspecies isolated from Tenerife cheese.
715 *Letters in Applied Microbiology* 34(2)**,** 82-85. doi: 10.1046/j.1472-765x.2002.01050.x.
716 Porcellato, D., Ostlie, H.M., Brede, M.E., Martinovic, A., and Skeie, S.B. (2013). Dynamics of
717 starter, adjunct non-starter lactic acid bacteria and propionic acid bacteria in low-fat and

full-fat Dutch-type cheese. *International Dairy Journal* 33(2)**,** 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. *International Dairy Journal* 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.* IEEE Computer Society).

Sanchez, J.I., Martinez, B., and 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(3)**,** 377-387. doi: 10.1016/j.ijfoodmicro.2005.04.025.

Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30(14)**,** 2068-2069. doi: 10.1093/bioinformatics/btu153.

Thunell, R.K. (1995). Taxonomy of the Leuconostocs. *Journal of Dairy Science* 78(11)**,** 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. *Applied and Environmental Microbiology* 61(8)**,** 3172-3176.

Vedamuthu, E.R. (1994). The dairy *Leuconostoc*: Use in dairy products. *Journal of Dairy Science* 77(9)**,** 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. *International Journal of Food Microbiology* 136. 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. *Journal of Applied Microbiology* 82. 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).

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(10)**,** 2591-2600. doi: 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(6270)**,** 63-65.

Wright, E. (2015). "DECIPHER: Database Enabled Code for Ideal Probe Hybridization Employing R. ". 1.16.1 ed.).

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*. *Journal für Verbraucherschutz und Lebensmittelsicherheit* 8(3)**,** 205-214. doi: 10.1007/s00003-013-0826-z.

764 Zhang, J., Kobert, K., Flouri, T., and Stamatakis, A. (2014). PEAR: a fast and accurate Illumina
765 Paired-End reAd mergeR. *Bioinformatics* 30(5)**,** 614-620. doi:
766 10.1093/bioinformatics/btt593.
767 Zhang, W., Liu, W., Song, Y., Xu, H., Menghe, B., Zhang, H., et al. (2015). Multilocus sequence
768 typing of a dairy-associated *Leuconostoc mesenteroides* population reveals clonal
769 structure with intragenic homologous recombination. *Journal of Dairy Science* 98(4)**,**
770 2284-2293. doi: http://dx.doi.org/10.3168/jds.2014-9227.
771 Zhao, Y., Jia, X., Yang, J., Ling, Y., Zhang, Z., Yu, J., et al. (2014). PanGP: a tool for quickly
772 analyzing bacterial pan-genome profile. *Bioinformatics* 30(9)**,** 1297-1299. doi:
773 10.1093/bioinformatics/btu017.
774 Østlie, H.M., Kraggerud, H., Longva, A.B., and Abrahamsen, R.K. (2016). Characterisation of
775 the microflora during ripening of a Norwegian semi-hard cheese with adjunct culture of
776 propionic acid bacteria. *International Dairy Journal* 54**,** 43-49. doi:
777 http://dx.doi.org/10.1016/j.idairyj.2015.10.005.

778

Tables

Table 1: Average genome size and coding sequences of *Leuconostoc* isolates binned into pan-genome lineages. Information on each individual isolate is included in supplementary table S1.

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)

784 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	-	-	-	-

785 +, presence of predicted pathway functionality. -, absence of predicted pathway functionality.
786

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=12)	C2 (n=6)	C3 (n=3)	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	#	#	#	+	+	+	+	+	+	+	+	+
citCDEFGS	+	+	+	+	+(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* & *levE*, fructose PTS; *frk*, fructokinase; *citCDEFGS*, citrate metabolism operon; *fba*, fructose bisphosphate aldolase

800 Table 4: Genetic potential for proteolytic activity.

	Ln. cremoris			***Ln. mesenteroides***				***Ln. pseudomesenteroides***				***Ln. lactis***
Gene(s)	C1 (n=12)	C2 (n=6)	C3 (n=3)	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	#	#	#	+	+	+	+	+	+	+	+	+

801 +, gene presence. -, gene absence. #, gene(s) present but truncated. Number in parenthesis indicates percentage of isolates where gene
802 was present. Genes are abbreviated as follows: *prtP*, type-II serine proteinase; *pepA*, glutamyl aminopeptidase; *pepC*, aminopeptidase
803 C; *pepF*, oligoendopeptidase; *pepN*, aminopeptidase N; *pepO*, neutral endopeptidase; *pepS*, aminopeptidase; *pepT*, peptidase T; *pepV*,
804 beta-ala-xaa dipeptidase; *pepX*, xaa-pro depeptidyl-peptidase; *oppABCDF*, peptide ABC transporter operon.

Figures

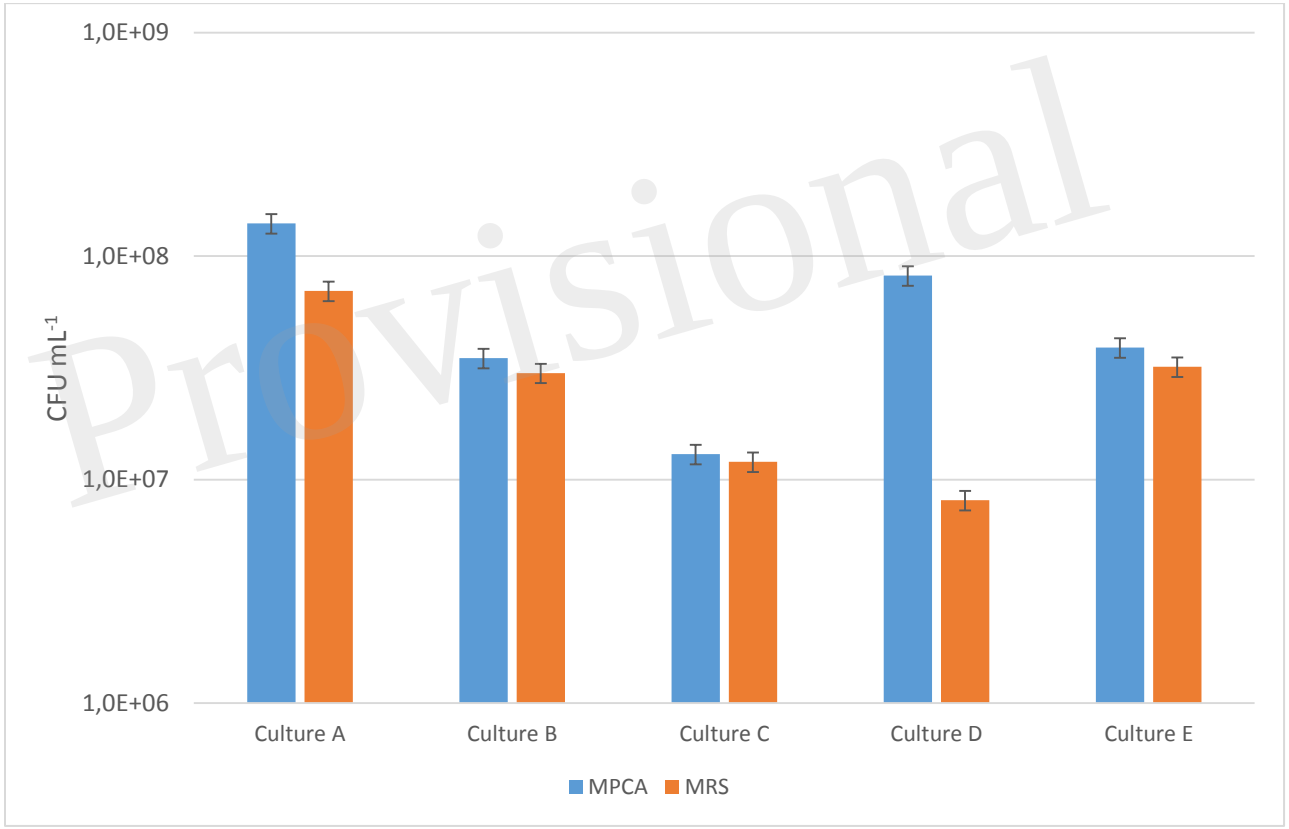


Figure 1: Bacterial counts for five starter cultures A, B, C, D, and E on MRS and MPCA supplemented with vancomycin to select for *Leuconostoc*. The counts are the mean of three separate extractions made from the same culture batch and the error bar indicates the standard deviation. The blue bars represent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

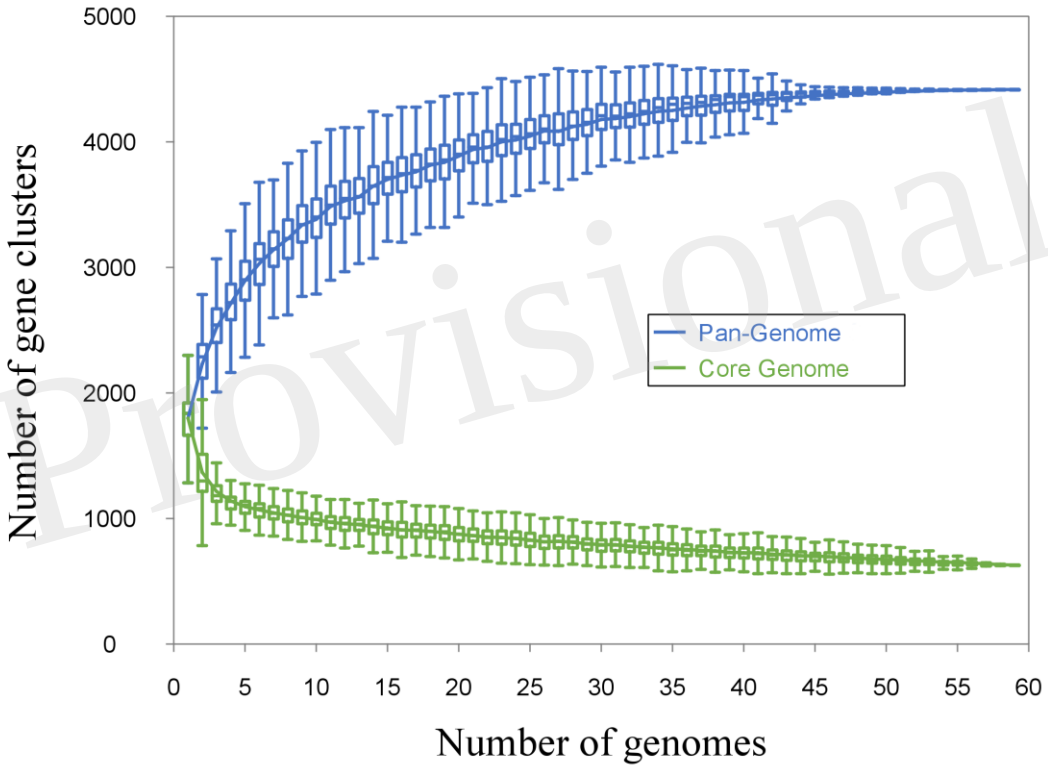
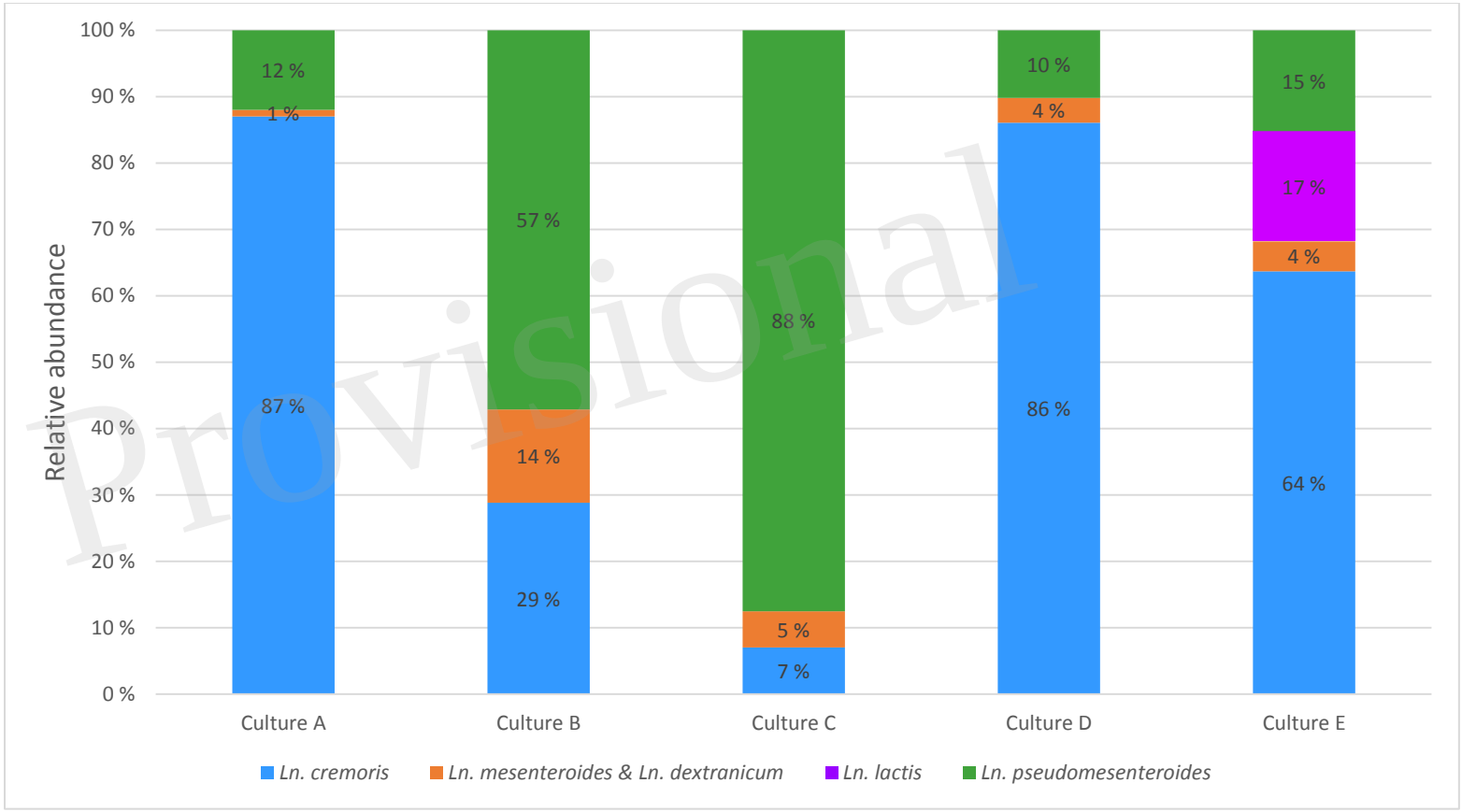


Figure 2. Pan- and core-genome estimation. The estimation is made by including genomes one by one, matching the genetic content from each genome, with the growing pan- and the decreasing core-genome. Homologous genes are clustered together in orthologous groups. If the genomes included in the estimation 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.

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.



829

830 Figure 4: Composition of leuconostocs in five starter cultures A, B, C, D and E using targeted-amplicon sequencing of the *eno* gene.

831

832

833

Figure 02.TIFF

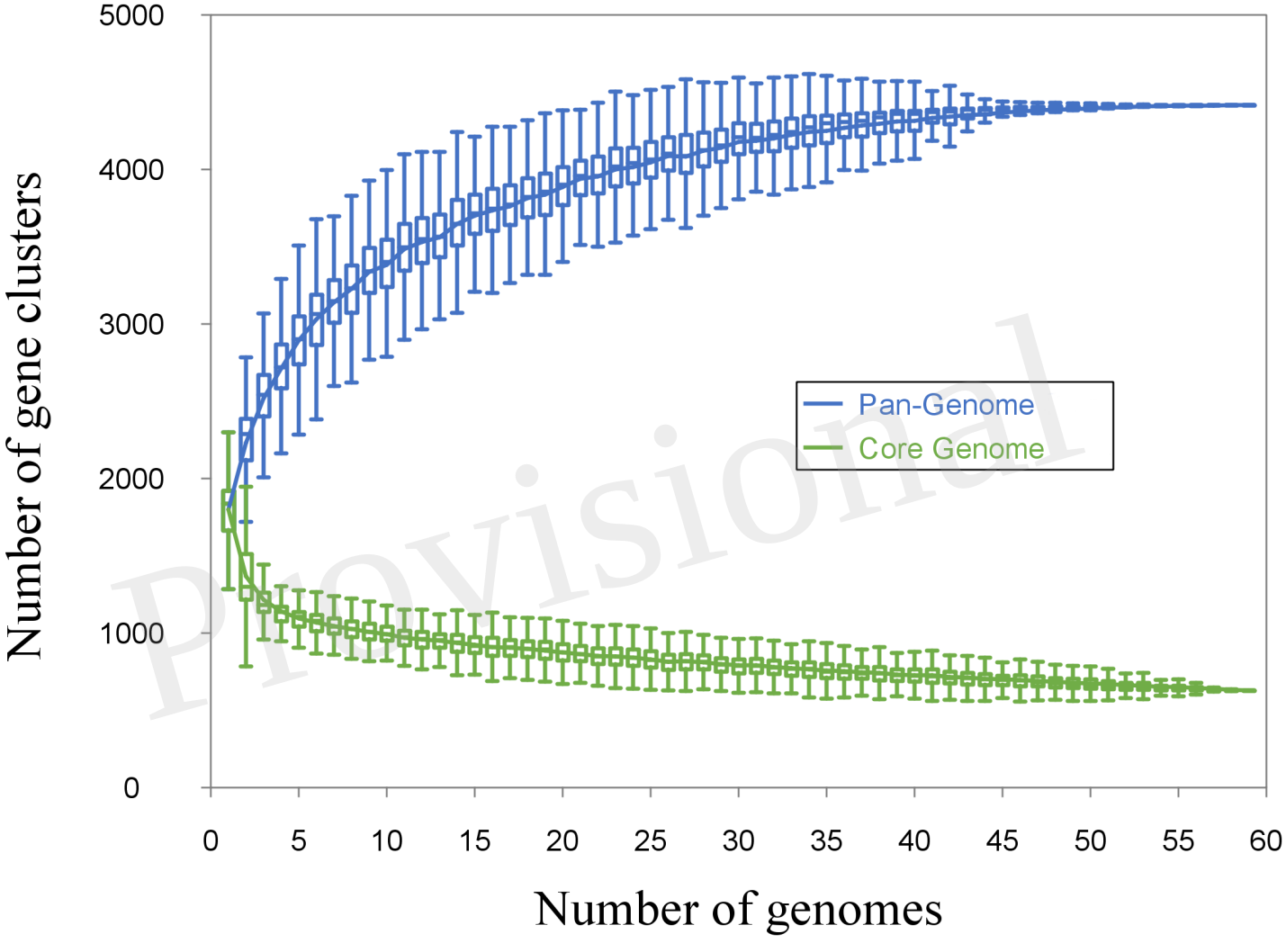


Figure 03.TIFF