CORA Cork Open Research Archive
Cartlann Taighde Oscailte Chorcaí

Title	From field to fermentation: characterisation and application of non-dairy cultures in dairy foods
Authors	Cavanagh, Daniel
Publication date	2014
Original Citation	Cavanagh, D. 2014. From field to fermentation: characterisation and application of non-dairy cultures in dairy foods. PhD Thesis, University College Cork.
Type of publication	Doctoral thesis
Rights	© 2014, Daniel Cavanagh. - http://creativecommons.org/licenses/by-nc-nd/3.0/
Download date	2026-11-05 22:55:00
Item downloaded from	https://hdl.handle.net/10468/3458

UCC
University College Cork, Ireland
Coláiste na hOllscoile Corcaigh

University College Cork, Ireland
Coláiste na hOllscoile Corcaigh

From Field to Fermentation: Characterisation & Application of Non-Dairy Cultures in Dairy Foods

A Thesis Presented to the National University of Ireland

for the Degree of Doctor of Philosophy

by

Daniel Cavanagh B.Sc. B.Ed.

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

School of Microbiology, University College Cork, Co. Cork, Ireland

Date: 19th December 2014

Head of School : Prof. Gerald Fitzgerald

Research Supervisors:

Dr. Olivia McAuliffe, Dr. Kieran Kilcawley, Prof. Gerald Fitzgerald



The price of success is hard work, dedication to the job at hand, and the determination that whether we win or lose, we have applied the best of ourselves to the task at hand.

Vince Lombardi

Declaration

I hereby certify that this material, which I submit for assessment on the programme of study leading to the award of Doctor of Philosophy, is entirely my own work and has not been submitted for another degree, either at University College Cork or elsewhere.

_____

Daniel Cavanagh

Student number: 110224075

Date: _____

Table of contents

Declaration i

Table of contents ii

Abstract viii

Chapter 1

Literature Review

From Field to Fermentation: the origins of *Lactococcus lactis* and its domestication to the dairy environment 1

1.1. Abstract 2

1.2. Introduction 3

1.3. Diverse niches of *Lactococcus lactis* 6

1.4. Taxonomic classification of *Lactococcus lactis* 8

1.4.1. Subspecies designation 8

1.4.2. Genotype-phenotype mismatching 12

1.5. Characteristics of *L. lactis* genomes 17

1.5.1. Core and dispensable genome 18

1.6. Adaptation to the dairy environment 21

1.6.1. Genetic determinants of domestication 21

1.6.1.1. Horizontal gene transfer (HGT) 21

1.6.1.2. The role of mobile genetic elements 23

1.6.1.2.1. Plasmids 23

1.6.1.2.2. Transposable elements 26

1.6.1.3. Gene loss and genome reduction 28

1.6.2. Phenotypic traits associated with domestication 29

1.6.2.1. Protein utilisation 29

1.6.2.2. Lactose utilisation 31

1.6.2.3. The curious incidence of *lacE* and *prtP* 32

1.6.2.4. Amino acid auxotrophy in *Lactococcus lactis* 35

1.6.2.5. Carbohydrate and sugar metabolism 38

1.7. Applications of ‘wild-type’ strains in fermented foods 42

1.7.1 Documented food applications 42

1.7.2. Potential restrictions on the use of non-dairy LAB in foods 44
1.8. Conclusions 46
1.9. Future perspectives 47
1.10. Acknowledgements 49
1.11. References 50

Chapter 2

Evaluation of non-dairy *Lactococcus lactis* with potential dairy applications, reveals extensive phenotype-genotype disparity 83
2.1. Abstract 84
2.2. Introduction 85
2.3. Materials and methods 88
2.3.1. Strains and growth conditions 88
2.3.2. *Lactococcus* isolation from non-dairy sources 88
2.3.3. Subspecies phenotype identification 89
2.3.4. Identification of *Lactococcus* species and subspecies genotypes 89
2.3.5. Multi-locus sequence typing (MLST) 90
2.3.6. Average nucleotide identity (ANI) analysis 91
2.3.7. Biosample accession numbers 92
2.3.8. Assessment of key technological traits in non-dairy isolates 92
2.3.8.1. Lactose utilisation 92
2.3.8.2. Amino acid transferase activity (methionine and phenylalanine) 92
2.3.8.3. Antibiotic resistance and biogenic amine formation 93
2.4. Results 94
2.4.1. Species confirmation of non-dairy isolates 94
2.4.2. Subspecies phenotype identification 94
2.4.3. Subspecies genotype identification 95
2.4.4. Genetic diversity of *L. lactis* strains from different environments 96
2.4.4.1. Clustering of *lactis* and *cremoris* genotypes 96
2.4.4.2. ANI: Revision of *Lactococcus* subspecies as separate specie 97
2.4.5. Key technological traits 99
2.4.5.1. Starter activity: milk acidification 99

2.4.5.2. Enzyme activity analysis 100
2.4.5.3. Antibiotic Resistance 100
2.5. Discussion 102
2.6. Conclusions 107
2.7. Acknowledgements 108
2.8. References 109

Chapter 3

Assessment of wild non-dairy lactococcal strains for flavour diversification in a mini Gouda-type cheese model 129
3.1. Abstract 130
3.2. Introduction 131
3.3. Materials and methods 133
3.3.1. Bacterial strains and growth conditions 133
3.3.2. Technological characterisation of strains 134
3.3.3. Mini cheese production 135
3.3.4. Cheese physicochemical and microbiological analysis 136
3.3.5. Cheese extract preparation 136
3.3.6. Enzyme activity assays 137
3.3.7. Proteolysis 137
3.3.8. Sensory evaluation 137
3.3.9. Statistical analysis 138
3.4. Results and discussion 140
3.4.1.1. Technological characterisation of strains 140
3.4.1.2. Production of volatile compounds 141
3.4.2. Cheese physicochemical analysis 142
3.4.3. Stability of non-dairy cultures over ripening 143
3.4.4. Peptidase activity 145
3.4.5. Proteolysis 146
3.4.5.1. pH 4.6 soluble nitrogen 146
3.4.5.2. Free amino acid analysis 147
3.4.6. Sensory evaluation 147
3.5. Conclusions 151

3.6. Acknowledgements 152
3.7. References 153
3.8. Appendices 168
Appendix 3-1 169
Appendix 3-2 170
Appendix 3-3 171

Chapter 4

Phages of non-dairy lactococci: isolation and characterisation of ϕL47, a phage infecting the grass isolate *Lactococcus lactis* ssp. *cremoris* DPC6860 172
4.1. Abstract 173
4.2. Introduction 175
4.3. Materials and methods 178
4.3.1. Bacterial strains, bacteriophage and culture conditions 178
4.3.2. Sensitivity of non-dairy isolates to dairy phages 178
4.3.3. Phage isolation and propagation 179
4.3.4. Lytic spectrum and adsorption to lactococcal cells 179
4.3.5. Electron microscopy 180
4.3.6. Structural analysis of phage proteins 180
4.3.7. Restriction digest profiles 181
4.3.8. DNA sequencing, annotation and comparative analysis 181
4.4. Results and discussion 183
4.4.1. Isolation of ϕL47 from sewage 183
4.4.2. Limited host range of ϕL47 184
4.4.3. Extended tail fiber of ϕL47 184
4.4.4. Structural protein analysis 185
4.4.5. Restriction digestion profiling 186
4.5. Analysis of the ϕL47 genome 187
4.5.1. General characteristics 187
4.5.2. Tail morphology 188
4.5.3. Host cell lysis 189
4.5.4. DNA replication and nucleotide biosynthesis 189
4.5.5. tRNA encoding genes in ϕL47 genome 191

4.5.6. Comparative genome analysis 193
4.6. Conclusions 196
4.7. Acknowledgements 197
4.8. References 198

Chapter 5

Location, location, location; examination of niche specialisation of *Lactococcus lactis* through comparative genome analysis 222
5.1. Abstract 223
5.2. Introduction 224
5.3. Materials and methods 226
5.3.1. Bacterial strains and growth conditions 226
5.3.2. Plasmid profile analysis 226
5.3.3. Draft genome sequencing and annotation 226
5.3.4. Mining of non-conserved ORFs 227
5.3.5. Minimal media to assess carbohydrate fermentation 228
5.3.6. Carbohydrate fermentation 228
5.3.7. Growth inhibition 229
5.3.8. Lysine decarboxylase activity 229
5.4. Results and discussion 231
5.4.1. Genome comparison of DPC strains from grass, corn and the bovine rumen 231
5.4.2. Technologically important traits 233
5.4.2.1. Milk-specific gene sets 234
5.4.2.2. Bacteriophage resistance 235
5.4.2.2.1. Diverse anti-phage systems of non-dairy isolates 236
5.4.2.3. Flavour formation 238
5.4.3. Niche specific traits 239
5.4.3.1. Diverse carbohydrate fermentation capacity of non-dairy isolates 240
5.4.3.1.1. Specific carbohydrate utilisation genes in grass and corn isolates 240

5.4.3.1.2. Insights into carbohydrate utilisation by rumen-associated *L. lactis* 241
5.4.3.1.2.1. Common carbohydrate-degrading enzymes associated with rumen and grass niches 242
5.4.3.1.2.2. Specialisation to the rumen environment 243
5.4.3.2. Surviving the cold: genetic determinants associated with growth at low temperatures 245
5.4.3.3. Surviving low pH: Biogenic amine production 247
5.4.3.3.1. Cadaverine production 247
5.4.3.3.2. Putrescine production 248
5.4.3.4. Resistance and immunity to antimicrobial compounds 249
5.4.4. Heavy metal ion transport 250
5.4.4.1. Increased insensitivity to cobalt 250
5.4.4.2. Cobalt sensitivity of DPC6856 252
5.4.4.3. Cadmium 252
5.4.5. Extra-cellular secretions 253
5.4.5.1. WXG100 secretion system (Wss) 253
5.4.5.2. Exopolysaccharide (EPS) production 254
5.4.5.3. *yydFGHIJ*-like gene cluster of DPC6856 255
5.4.6. Unique presence of *cobQ* in *L. lactis* 256
5.5. Conclusions 258
5.6. Acknowledgements 259
5.7. References 260
5.8. Appendices 284
Appendix 5-1 285
Appendix 5-2 296
Appendix 5-3 303

Chapter 6

General Discussion 308
6.1. General discussion 309

6.2. References 318

Acknowledgements **322**

Abstract

This thesis investigates the phenotypic and genotypic diversity of non-dairy *L. lactis* strains and their application to dairy fermentations. A bank of non-dairy lactococci were isolated from grass, vegetables and the bovine rumen. Subsequent analysis of these *L. lactis* strains revealed seven strains to possess *cremoris* genotypes which did not correlate with their observed phenotypes. Multi-locus sequence typing (MLST) and average nucleotide identity (ANI) highlighted the genetic diversity of *lactis* and *cremoris* subspecies. The application of these non-dairy lactococci to cheese production was also assessed. In milk, non-dairy strains formed diverse volatile profiles and selected strains were used as adjuncts in a mini Gouda-type cheese system. Sensory analysis showed non-dairy strains to be strongly associated with the development of off-flavours and bitterness. However, microfluidisation appeared to reduce bitterness. A novel bacteriophage, ϕL47, was isolated using the grass isolate *L. lactis* ssp. *cremoris* DPC6860 as a host. The phage, a member of the *Siphoviridae*, possessed a long tail fiber, previously unseen in dairy lactococcal phages. Genome sequencing revealed ϕL47 to be the largest sequenced lactococcal phage to date and owing to the high % similarity with ϕ949, a second member of the 949 group. Finally, to identify and characterise specific genes which may be important in niche adaptation and for applications to dairy fermentations, comparative genome sequence analysis was performed on *L. lactis* from corn (DPC6853), the bovine rumen (DPC6853) and grass (DPC6860). This study highlights the contribution of niche specialisation to the intra-species diversity of *L. lactis* and the adaptation of this organism to different environments. In summary this thesis describes the genetic diversity of *L. lactis* strains from outside the dairy environment and their potential application in dairy fermentations.

Chapter 1

Literature Review

From Field to Fermentation: the origins of *Lactococcus lactis* and its domestication to the dairy environment

All of this chapter has been published in Food Microbiology (2014). 47, 45-61.
DOI: 10.1016/j.fm.2014.11.001.

1.1. Abstract

Lactococcus lactis is an organism of substantial economic importance, used extensively in the production of fermented foods and widely held to have evolved from plant strains. The domestication of this organism to the milk environment is associated with genome reduction and gene decay, and the acquisition of specific genes involved in protein and lactose utilisation by horizontal gene transfer. In recent years, numerous studies have focused on uncovering the physiology and molecular biology of lactococcal strains from the wider environment for exploitation in the dairy industry. This in turn has facilitated comparative genome analysis of lactococci from different environments and provided insight into the natural phenotypic and genetic diversity of *L. lactis*. This diversity may be exploited in dairy fermentations to develop products with improved quality and sensory attributes. In this review, we discuss the classification of *L. lactis* and the problems that arise with phenotype/ genotype designation. We also discuss the adaptation of non-dairy lactococci to milk, the traits associated with this adaptation and the potential application of non-dairy lactococci to dairy fermentations.

1.2. Introduction

The lactic acid bacteria (LAB) comprise a group of Gram-positive organisms which are primarily anaerobic, non-sporulating bacteria which produce lactic acid as the principal end product of sugar fermentation (Kandler and Weiss, 1986). Numerous species of bacteria are capable of producing lactic acid but the term LAB is restricted to specific genera including *Lactococcus* amongst others (Stiles and Holzapfel, 1997). LAB have been exploited for thousands of years in the production of fermented foods whereby they contribute to flavour, quality, texture and safety of the products (Settanni and Corsetti, 2008). While these organisms produce mainly lactic acid as a by-product of sugar fermentation (homofermentation), there are many examples of heterofermenting LAB, producing an array of different fermentation products including ethanol, carbon dioxide, acetic acid and formic acid (Kleerebezem and Hugenholtz, 2003). Among the most widely selected genera for industrial application are *Oenococcus* (wine), *Lactobacillus* (meat, vegetables, dairy, cereals) and *Lactococcus* (dairy) (Bourdichon *et al.*, 2012). Two species of *Lactococcus* are listed in the 'Inventory of Microbial Food Cultures' with documented use in food fermentations, *Lactococcus raffinolactis* and *Lactococcus lactis* (Bourdichon *et al.*, 2012).

L. lactis is the main constituent of dairy starter culture systems used worldwide for the production of numerous fermented dairy products including cheese of both artisanal and commercial origin, and fermented milks such as buttermilk and sour cream. Indeed, through the consumption of fermented dairy products, it is estimated that humans ingest up to 10^{18} lactococcal cells per annum (Mills *et al.*, 2010). Based on its history of use in food fermentations, *L. lactis* has GRAS, or Generally Regarded As Safe, status (FDA, 2010). The predominant role of *L. lactis* in dairy starter cultures

is to produce lactic acid at a sufficient rate and contribute to the breakdown of milk proteins during fermentation (Wouters *et al.*, 2002), thus significantly contributing to the final product in terms of organoleptic properties and microbial quality (Ross *et al.*, 2000). A small number of *L. lactis* strains are routinely used in fermented food production in an effort to develop more consistent products. These strains are chosen primarily for their acidification activity and resistance to bacteriophage infection (Marshall, 1991). However, consumer demands for products of more diverse flavour is driving manufacturers of fermented dairy products to expand their culture systems. These new cultures must be able to create desirable products in terms of flavour and texture while also enduring the environmental stresses associated with the manufacture of fermented dairy foods (Mills *et al.*, 2010). The creation of genetically modified organisms (GMOs) has the potential to resolve this problem, but the introduction of such organisms into the food chain is met with much opposition from both governmental agencies and consumers alike (Pedersen *et al.*, 2005). An alternative non-GM approach is through the examination of the natural biodiversity which exists in the LAB group, including *L. lactis* from outside the dairy environment. Such an approach may provide a means to identify novel starter cultures with the desired industrial traits for production in terms of flavour development and bacteriophage resistance (Ayad *et al.*, 2000; Mills *et al.*, 2010).

A high degree of sequence similarity exists between *Streptococcaceae*, yet they can be found in a broad range of different environmental niches. Members of the *Lactococcus* genus can be isolated from raw-milk, raw-milk cheeses and non-milk environments and are collectively referred to as 'wild-type' (Ayad *et al.*, 2000). Indeed, strains of *L. lactis* have been isolated from a range of sources including drain water and human vaginal samples (Gao *et al.*, 2011, Kato *et al.*, 2012). Although not

a common resident of the gastro intestinal tract (GIT), *L. lactis* is capable of surviving gut passage (Kimoto *et al.*, 1999; Meyrand *et al.*, 2013). This in turn has opened up the potential of these strains for probiotic use, and the delivery of therapeutic drugs *in-vivo* (Steidler and Rottiers, 2006; Wells and Mercenier, 2008) and the targeted delivery of vaccines by this organism has been examined (Asensi *et al.*, 2013). In recent years the number of sequenced lactococcal strains from different environmental niches has grown considerably (Table 1). This has shed further light on the diversity within the *L. lactis* species and identified genes present in these strains which may be harnessed to impart added value to dairy fermentations. The aim of this review is to discuss the potential origins and natural diversity of *L. lactis*, and to highlight the mechanisms by which this industrially important organism has become adapted to the dairy environment. The potential features present in strains of *L. lactis* from outside the dairy environment which could be beneficial in dairy fermentations are also highlighted.

1.3. Diverse niches of *Lactococcus lactis*

LAB are often referred to as fastidious organisms, found in nutrient rich habitats; however, in such environments these bacteria can be exposed to extremes of pH, differing nutrient availability and challenges from other microbiota for resources (van de Guchte *et al.*, 2002). *L. lactis* possesses a wide ecological distribution from sourdough bread (Passerini *et al.*, 2013a), to sugar cane plants (Serna Cock and Rodríguez de Stouvenel, 2006) to the GIT of brown trout (Pérez *et al.*, 2011). Recently, the genome sequence of *L. lactis* ssp. *lactis* IO-1, isolated from drain water, has been elucidated which possesses the capacity to utilise xylose and generates increased levels of L-lactic acid (Kato *et al.*, 2012). Although *L. lactis* may naturally inhabit many different environments (see Table 1 for origins of sequenced strains), it is most widely known for its association with the milk environment and in the production of dairy products. Based on multi-locus sequence typing (MLST) and the formation of clonal complexes, Passerini *et al.* (2010) proposed that *L. lactis* strains be classed as 'domesticated' or 'environmental' in reference to their origin. Domesticated strains are thus defined as strains used as dairy starters, in milk production and found in fermented products, while environmental strains are defined as isolates from plants, animals and raw-milk.

The consensus is that industrial dairy strains used in production today are believed to have descended from plants (Kelly *et al.*, 2010) and have adapted over time to thrive in milk. It is plausible that this species initially colonised milk from contact with grass or other plants which would have been used as fodder or bedding for cattle. *L. lactis* is one of the first bacteria to occupy plant material, where it becomes the less dominant species as the pH of the environment lowers (Kelly *et al.*, 1998b; Kelly and Ward, 2002). It is thought that these bacteria initially colonised

seeds, prior to germination, which enables them to quickly establish themselves as the dominant microbial community (Kelly *et al.*, 1998b). This in turn may be aided by the production of anti-microbial compounds, as numerous strains isolated from plant matter have been found to produce nisin and other uncharacterised bacteriocins (Kelly *et al.*, 1998b). The material used in bedding preparation strongly influences the bacteria found on a cow's teat and in materials such as straw, higher levels of mesophilic bacteria have been identified (Monsallier *et al.*, 2012). Currently, *L. lactis* is rarely encountered in the farm environment (Vacheyrou *et al.*, 2011) but this may not always have been the case, and may be attributed to the modern use of pesticides which could adversely affect microbial growth (Clair *et al.*, 2012). Analysis of the microbial content of raw-milk from French dairy farms, showed a reduction in the levels of lactococci over time (Mallet *et al.*, 2012). This was mainly attributed to the treatment of the cows' teats, the surface of which is considered an important source of bacteria for the production of raw-milk cheeses (Vacheyrou *et al.*, 2011).

1.4. Taxonomic classification of *Lactococcus lactis*

Bacteria of the genus *Lactococcus* are recognised as belonging to one of nine species: *L. garvieae, L. lactis, L. plantarum, L. formosensis, L. piscium, L. fujiensis, L. chungangensis, L. taiwanensis* and *L. raffinolactis* (Schleifer *et al.*, 1985; Cho *et al*., 2008; Cai *et al*., 2011; Chen *et al.*, 2013; Chen *et al.*, 2014). *L. garvieae* is a pathogenic organism of fish such as trout and yellow fin, and has also been associated with bovine mastitis and human clinical specimens (Teixeira *et al.*, 1996; Vinh *et al.*, 2006; Wang *et al.*, 2007; Yiu *et al.*, 2007). *L. raffinolactis* has been identified as a fish commensal but also an opportunistic pathogen (Michel *et al.*, 2007), while *L. piscium* has been isolated from packaged beef products (Sakala *et al.*, 2002). New species of *Lactococcus, L. taiwanensis* sp. *nov*. and *L. formosensis* have recently been isolated from cummingcordia, a plant used in the production of a fermented Taiwanese foodstuff, and fermented broccoli stems, respectively (Chen *et al.*, 2013; Chen *et al.*, 2014).

1.4.1. Subspecies designation

Prior to 1985 members of the genus *Lactococcus* were referred to as dairy streptococci, however, in order to differentiate these organisms from the pathogenic bacteria streptococci *sensu stricto, Streptococcus lactis* and *Streptococcus cremoris* were reclassified as *L. lactis* ssp. *lactis* and *L. lactis* ssp. *cremoris*, respectively. Of all *Lactococcus* species, *L. lactis* is the most widely studied due to the substantial industrial importance of this organism in the dairy industry. *L. lactis* is divided into four subspecies: *lactis, cremoris, hordniae* (leaf hopper) and most recently, *tructae* (trout intestine) (Pérez *et al.*, 2011).

Molecular fingerprinting of *L. lactis* strains has revealed the presence of two major genotypes termed *L. lactis* ssp. *lactis* and *L. lactis* ssp. *cremoris* (Jarvis and Jarvis, 1981; Rademaker *et al.*, 2007). Comparative genome analysis between the two subspecies show that they possess 85% overall identity (Wegmann *et al.*, 2007). In the past, limited high throughput technology was available for bacterial characterisation and numerous assays were developed for subspecies genotype identification (see Table 2) including DNA-DNA hybridisation and PCR fingerprint analysis. The method used for assignment of subspecies genotypes has varied amongst reports (Biscola *et al.*, 2013; Parapouli *et al.*, 2013; Alemayehu *et al.*, 2014) which included 16S rRNA analysis amongst others. The 16S rRNA sequence differs by as little as 0.7% between subspecies depending on strain. The use of this gene alone for subspecies identification has previously led to inaccurate assignment of genotypes which was corrected by MLST analysis (Xu *et al.*, 2014). In comparison to MLST, multi-locus sequence analysis (MLSA) is more applicable for use where clonal evolution occurs and where recombination events rarely take place. It is important to note that these two terms are not interchangeable as each analysis possesses its own features and uses. Glutamate decarboxylase catalyzes the conversion of glutamate to γ-aminobutyric acid (GABA) and is involved in acid resistance at low pH (Sanders *et al.*, 1998). It has been reported that subspecies *lactis* strains are capable of producing GABA while *cremoris* strains are negative for this trait (Nomura *et al.*, 1999), due to an inactive *gadB* gene caused by a frameshift mutation (Nomura *et al.*, 2000). The presence or absence of the *AseI* restriction site appeared to correspond to GAD^+ or GAD^- phenotype respectively. This has subsequently led to the development of a PCR-based technique for the differentiation of subspecies (Nomura *et al.*, 2002) as described in Table 2. To date, MLST has been

successfully applied for characterisation of *L. lactis* (Rademaker *et al.*, 2007; Taïbi *et al.*, 2010) and Xu *et al.* (2014) proposed that future subspecies genotyping should be determined by MLST in order to avoid inaccuracies based on 16S rRNA analysis.

DNA-DNA hybridisation has, in the past, been used as a foremost technique for genomically circumscribing prokaryotic species. With advances in sequencing technology and an abundance of data available on public databases, this technique has consequently been considered obsolete (Stackebrandt *et al.*, 2002). The use of average nucleotide identity (ANI) to replace DNA-DNA hybridisation has shown to be a valuable tool in prokaryotic species circumscription (Richter and Rosselló-Móra, 2009). Complete chromosomes of sequenced *L. lactis* strains were downloaded from Genbank (http://www.ncbi.nlm.nih.gov/genome/?term=Lactococcus+lactis) and examined using the JSpecies software tool (Richter and Rosselló-Móra, 2009) which uses either the BLAST (Altschul *et al.*, 1997) or MuMmer algorithm (Kurtz *et al.*, 2004). This software tool also allowed for the calculation of tetranucleotide frequency correlation coefficients (TETRA), which has previously been applied in the phylogenetic analysis of metagenomic samples (Teeling *et al.*, 2004). When compared to the type-strain, IL1403, previously identified *lactis* subspecies possessed an ANIm value of between 98.01% and 99.79% (Table 3) while similar results were observed for subspecies *cremoris* strains and the prototype-strain MG1363 (98.18%-99.47%). These results indicate that *lactis* strains belong to the same species as strain IL1403 and likewise for *cremoris* strains and strain MG1363. Contrastingly, when strain IL1403 was examined against *cremoris* subspecies ANIm values ranged from 88.28%-88.6% and from 88.25% to 88.62%, when MG1363 was compared to *lactis* subspecies. Richter and Rosselló-Móra (2009) suggested that two strains be treated as separate species, if an ANIm (average nucleotide identity determined using the MuMmer algorithm)

value of <95% is observed. These low ANIm values observed between type-strains hints that these two subspecies be classed as different species according to this analysis. It is important to note that this analysis is based upon comparative sequence analysis only and does not take into account the phenotypic similarity between the strains. In accordance with Tomás *et al.* (2013) the authors agree that bacterial species be classed according to numerous criteria; taking into account genotype, phenotype, niche, and utilising the expanding tools in genetics and genome analysis.

Within the *lactis* subspecies, there exists a biovariant *diacetylactis* which ferments citrate, yielding diacetyl and acetoin. Diacetyl present in small amounts imparts a buttery and creamy aroma to dairy products and is an essential compound in Camembert, Emmental and Cheddar cheese varieties (Curioni and Bosset, 2002). In the past, the presence of the *cit* operon for citrate fermentation was used to distinguish between *lactis* subspecies; however, this method is not a dependable method because of the unstable nature of this plasmid-encoded phenotype (Kempler and McKay, 1981). Citrate breakdown requires the presence of a permease to take up citrate and a chromosomally encoded lyase to initiate citrate catabolism (Drider *et al.*, 2004). *L. lactis* IL1403 is a plasmid-cured strain of *L. lactis* ssp. *lactis* biovar. *diacetylactis* CNRZ157 and possesses chromosomal genes encoding citrate breakdown, which are absent in *L. lactis* ssp. *cremoris* SK11 and strain MG1363 (Price *et al.*, 2012). Analysis of the population network of *L. lactis* subspecies *lactis* grouped *diacetylactis* variants into a single clonal complex composed of strains considered ‘domesticated’ to the dairy environment (Passerini *et al.*, 2010). A low degree of genetic diversity was evident amongst these strains in contrast to environmental strains from skin, raw-milk and plant matter. Of note, ‘wild-type’ isolates were found to produce as much diacetyl as domesticated strains that possess

a plasmid-encoded citrate permease (*citP*) and the chromosomally encoded gene cluster *citM-citl-citCDEFXG* (Passerini *et al.*, 2013b). One of these strains lacked the plasmid encoded citrate permease, while the other lacked both the *citP* gene and the *citM-citI-citCDEFXG* cluster. The rerouting of pyruvate to aroma compound formation pathways, was hypothesised as the mechanism by which strains lacking any citrate utilization genes were able to generate acetoin (Passerini *et al.*, 2013b).

1.4.2. Genotype-phenotype mismatching

In the past, species and subspecies were identified by their metabolic traits, and thus *cremoris* or *lactis* designation was established by a variety of phenotypic characteristics (Fig. 1) (Garvie and Farrow, 1982). Phenotypic identification as *cremoris* is determined by the possession of three metabolic traits: 1) the inability to grow at 40 ^{o}C, 2) the inability to grow in the presence of 4% (w/v) NaCl and 3) the inability to deaminate arginine (Orla-Jensen, 1919). The impaired growth rate of *cremoris* phenotypes in the presence of 4% NaCl has been associated with the failure to accumulate betaine caused by genetic alterations, including the functional loss of the *busA* operon (Obis *et al.*, 2001). Arginine catabolism in *Streptococcaceae* takes place via the arginine deiminase pathway containing three enzymes: arginine deiminase, catabolic ornithine carbamoyltransferase and carbamate kinase (Cunin *et al.*, 1986). According to classical phenotypic identification, *L. lactis* MG1363 was initially designated as subspecies *lactis*, as it possesses an active arginine deiminase (ADI) pathway for arginine breakdown (Godon *et al.*, 1992; Van Hylckama Vlieg *et al.*, 2006). However, following various genetic analyses, this strain was later reclassified as subspecies *cremoris*, possessing two of the three phenotypic traits associated with this subspecies and a *cremoris* genotype (Wegmann *et al.*, 2007).

L. lactis ssp. *cremoris* SK11 possesses the three phenotypic traits associated with the *cremoris* phenotype due to an inactive ADI pathway caused by a single base pair deletion in the *arcA* gene. This strain is identified as being a "true" *cremoris* subspecies as it possesses a *cremoris* genotype and tests negative for all three phenotypic traits used to define the subspecies. A similar phenotype to subspecies *cremoris* is also observed in *L. raffinolactis* (Boucher *et al.*, 2003). As mentioned previously, glutamate decarboxylase activity, identified in *lactis* but not *cremoris* subspecies, is also employed as a means of phenotypic subspecies identification (Nomura *et al.*, 1999; Nomura *et al.*, 2000; Nomura *et al.*, 2006). Thus the phenotype of a subspecies can be easily identified. However, the results of phenotypic assays can vary according to conditions in which the organism was cultured and the phase of growth of the organism. This can lead to incorrect identification of a particular phenotype and highlights the need for more reliable means of phenotypic identification (Salama *et al.*, 1993; Urbach *et al.*, 1997; Garde *et al.*, 1999; Tanigawa *et al.*, 2010). Genetic classification alone may not adequately represent the phenotypic diversity of the strain and in the past, this has further complicated the taxonomic classification of *L. lactis* (Rademaker *et al.*, 2007; Kelly *et al.*, 2010). Therefore it appears that to accurately represent the nature of a strain, genotypic and phenotypic analysis should be done in tandem. In order to eliminate incongruences in phenotypic identification a standard approach should be taken to prepare strains for phenotypic analysis. Such approaches could include culturing strains in a given species-specific defined medium prior to analysis and performing phenotypic assays under specific stages of growth.

Some misunderstanding and confusion can arise in subspecies classification, as the phenotype and the genotype often do not correlate (Tailliez *et al.*, 1998). For

example, *L. lactis* strains identified as possessing a *lactis* phenotype (i.e. the ability to breakdown arginine, grow at 40 ^{o}C and in 4% NaCl), have previously clustered genotypically as subspecies *cremoris* (Weerkamp *et al.*, 1996; Fernández *et al.*, 2011) and conversely for *cremoris* phenotypes and *lactis* genotypes (Wouters *et al.*, 2002; Kelly *et al.*, 2010). In an effort to develop assays to account for genotype-phenotype mismatching, an amplified fragment length polymorphism (AFLP) method has been developed which allows for the identification of *cremoris* and *lactis* genotypes rooted in type-strain genomic lineages (Kütahya *et al.*, 2011). This method also allowed for the identification of *cremoris* phenotypic markers based on the formation of specific fragments using selected primer/ enzyme combinations. Sequence analysis of this fragment revealed it to encode a hypothetical membrane protein absent in strain IL1403 but present in strains MG1363 and SK11 (Kütahya *et al.*, 2011). The characterisation of *L. lactis* ssp. *cremoris* using the combined approaches of MLST and comparative genome hybridisation (CGH) has proven useful in strain classification and in identifying strain diversity (Taïbi *et al.*, 2010). Analysis performed in that study, allowed for the detection of specific genes and their respective variants, which included genes encoding industrially important functions such as nisin resistance and peptidase activity. This could be exploited in the dairy industry to identify strains which possess specific phenotypes which would enable the formation of tailored starter mixes whereby strains are selected based on compatibility. Differences were detected in the presence or absence of genes for osmoprotection between strains which also function in thermoprotection. The absence or mutation of these genes may lead to increased autolysis caused by higher temperatures and salt concentrations during cheese manufacturing, which has been shown to contribute to cheese flavour formation (Taïbi *et al.*, 2010; Yarlagadda *et al.*,

2014). In comparison to high salt cheeses, reduced cell autolysis by certain strains may be observed in low salt cheeses. Overall, this may contribute to less flavour development owing to the poor release of intra-cellular enzymes into the cheese matrix and reduce consumer acceptability of such cheeses.

Kelly and Ward (2002) examined over 500 *L. lactis* isolates from different environments and classified *L. Lactis* strains into one of five groups; (1) *lactis* genotype, *lactis* phenotype, (2) *lactis* genotype, *diacetylactis* phenotype, (3) *lactis* genotype, *cremoris* phenotype, (4) *cremoris* genotype, *lactis* phenotype and (5) *cremoris* genotype, *cremoris* phenotype. Previous analysis of *Lactococcus* strains from different environments and geographical regions led to the identification of two major lineages within the species which do not correspond to groups defined by phenotype (Rademaker *et al.*, 2007). According to Rademaker *et al.*, (2007), the first lineage comprises subspecies *cremoris* type-strain genotypes of both *cremoris* and *lactis* phenotypes, while the second is comprised of *lactis* type-strain genotypes and contained *lactis* phenotypes only. Notably, a lesser third lineage was also formed, containing two non-dairy *L. lactis* ssp. *lactis* isolates possessing both a *lactis* genotype and phenotype. In a comprehensive study conducted by Fernández *et al.* (2011), it was proposed that *lactis* and *cremoris* genotypes that have a *lactis* phenotype could form a new subspecies. It is possible that an increase in the isolation and genetic characterisation of *L. lactis* strains from the wider environment could lead to a breakdown in traditional methods for subspecies classification. Indeed new classification approaches may be required such as that previously set out by Kelly and Ward (2002) which refers to specific genotype and phenotype combinations.

1.5. Characteristics of *L. lactis* genomes

Organisms of the *Lactococcus* species possess relatively small genomes of between ~2,000 to ~2,800 protein-encoding genes (Makarova *et al.*, 2006). The NCBI database lists twenty three *L. lactis* genomes while the Genomes OnLine Database (GOLD www.genomesonline.org) provides details of eleven complete and two incomplete sequencing projects. Industrial dairy lactococci are commonly found to possess large 'plasmidomes' or plasmid complements which can contribute as much as 200 kb to the total genome size (Kelly *et al.*, 2010). These plasmids encode for a range of functions which may infer a competitive advantage in milk such as bacteriophage resistance, lactose utilisation and casein breakdown (Siezen *et al.*, 2005; Górecki *et al.*, 2011; Wegmann *et al.*, 2012; Ainsworth *et al.*, 2013). To date, over 80 sequenced plasmids are available on the NCBI database which includes isolates from dairy, human and plant-based environments. The plasmids of lactococci and the important role that they play in adaptation to milk will be discussed in more detail in section 1.5.1.

Kelly *et al.* (2010) examined the chromosomal diversity of industrial dairy and wild-type *L. lactis* strains using pulsed-field gel electrophoresis (PFGE) with both *Sma*I, to include plasmid DNA, and I-*Ceu*I, to include chromosomal DNA only. The origin of the strain was the most significant determinant of chromosome length with 'true' *cremoris* strains (having both a *cremoris* phenotype and genotype) possessing the smallest chromosome length. This is attributed to the process of reductive evolution as an outcome of adaptation to milk as all of these strains were of industrial dairy origin. A reduction in genome size has also been observed in other industrial dairy bacteria such as *Lactobacillus bulgaricus* and *Streptococcus thermophilus* (Hols *et al.*, 2005; Van de Guchte *et al.*, 2006). Conversely, strains of *L. lactis* ssp. *cremoris*

(genotype) with a *lactis* phenotype have been found to possess a larger genome size compared to 'wild-type' strains (Kelly *et al.*, 2010). The authors propose that strains that possess such a phenotype/ genotype combination are perhaps more closely related to plant strains (Kelly *et al.*, 2010).

The genetic and genome (chromosome size, plasmid complement, plasmid size(s), genome size) diversity of *L. lactis* from different regions and environmental niches, was examined by Passerini *et al.* (2010), employing a six gene MLST scheme and PFGE. MLST analysis separated strains from 'domesticated' habitats into two clonal complexes with 'environmental' strains forming unique sequence types (STs). These STs differed from each other by three to six loci, indicating the high degree of genetic diversity amongst these strains. Phylogenetic analysis suggested that the environmental group emerged first, with the domesticated group emerging more recently, which is confirmed in other reports (Kelly *et al.*, 2010). PFGE analysis did not correlate with MLST analysis but showed that although domesticated strains possessed the same STs, they differed overall by a total of 230 kb. Thus, the authors concluded that a high degree of gene loss and acquisition in the dispensable, or non-core genome, is responsible for the phenotypic differences of *L. lactis* strains and the formation of different clonal complexes.

1.5.1. Core and dispensable genome

In order to describe the genomic diversity within a species, the term 'pangenome' has been coined which comprises the total genes present within all members of a particular species (Smokvina *et al.*, 2013). Amongst bacteria, there can be a wide variation in the gene complement of strains that may be closely related which can be largely due to the environment from which the strains originated. For

example, comparison between *L. lactis* ssp. *lactis* A12, isolated from sourdough, and *L. lactis* ssp. *lactis* IL1403 revealed that 23% of the A12 genome corresponded to strain-specific genes which were mainly responsible for niche specialisation (Passerini *et al.*, 2013a). The term 'core genome' represents the genes that are shared between all members of the species. These genes are typically involved in important cellular processes such as central and nucleotide metabolism, and can extend to genes involved in sugar utilization and cell-surface proteins (Smokvina *et al.*, 2013). Analysis of nineteen *L. garvieae* genomes, five *L. lactis* genomes and one *L. raffinolactis* genome, showed that 70% of the core genes of *L. garvieae* were shared amongst the other *Lactococcus* species, suggesting a core genome of the genus (Ferrario *et al.*, 2013).

Examination of the core and dispensable genome of *L. lactis* has not been performed in great detail such as that for other LAB (Douillard *et al.*, 2013; Smokvina *et al.*, 2013). With the advanced number of sequencing projects undertaken in recent years, different reports are identifying the shared amount of genetic information within *L. lactis*. Between subspecies there exists a high degree of genome similarity as shown by the comparative analysis of *L. lactis* ssp. *cremoris* A76 with six *L. lactis* genomes which share a core genome of 1.84 mb (Bolotin *et al.*, 2012). *L. lactis* strains possessing a *lactis* genotype have been reported to share 1, 915 common genes while an 85% identity in coding sequences has been found between subspecies *cremoris* and *lactis* strains (Kelly *et al.*, 2010; Tan-a-ram *et al.*, 2011). The 'dispensable genome' refers to genes specific to a particular strain or lineage. These genes are most commonly associated with phages, transposons/ mobile elements, and plasmids but can encode for important functions which give the strain a competitive advantage such as phage resistance, EPS production and sugar utilization in the dairy environment

(Smokvina *et al.*, 2013). Passerini *et al.* (2010) hypothesised that the large degree of variation between phenotypes was caused by the active loss and gain of genes within the dispensable genome, and suggested that a large pangenome existed within the subspecies.

1.6. Adaptation to the dairy environment

Milk provides all of the nutrients necessary for the cultivation of numerous microorganisms. It contains an abundance of carbohydrates, proteins, fats, vitamins and minerals and has a neutral pH (Marshall, 1991). While some of these nutrients can be taken up-into the cell and immediately metabolised by certain bacteria, others require processing before they can be utilised such as milk fat and protein. Repeated culturing in milk has led to bacterial species which may have initially inhabited a non-dairy niche becoming specialised for growth in the milk environment. It is hypothesised that the degree of specialisation has progressed so far as to make these strains unable to survive outside of the milk habitat and they may no longer be able to occupy non-dairy niches (Kelly *et al.*, 2010).

1.6.1. Genetic determinants of domestication

1.6.1.1. Horizontal gene transfer (HGT)

The genetic changes which take place in the genome of a cell in response to specific environmental conditions are a significant source of diversity within a bacterial species (Van Hylckama Vlieg *et al.*, 2006). Horizontal gene transfer (HGT) is regarded as being a key mediator in the evolution of bacterial genomes (Cai *et al.*, 2009). In contrast to vertical transfer, which relates to the transfer of genetic information from one generation to the next, HGT is the movement of DNA between organisms of the same generation (Baltrus, 2013). Modern industrial dairy strains are believed to have emerged from a bottleneck event (a reduction in a population size leading to decreased genetic variation therein) or the colonisation of milk by a shared ancestor of non-dairy origin (Van Hylckama Vlieg *et al.*, 2006; Kelly *et al.*, 2010; Passerini *et al.*, 2010). This would suggest that the overall diversity in industrial dairy

strains is comparatively lower than that of non-dairy strains. The different environmental niches inhabited by *L. lactis* brings this organism into close contact with many diverse organisms, possessing many different genes for potential transfer events and *vice versa* (Licht *et al.*, 1999). For example, BLASTp analysis of the *pepX* homolog, *pepX*2, of *L. lactis* IL1403 showed that this gene was most similar to a gene of *Listeria monocytogenes* (Liu *et al.*, 2010). Genetic analysis of the plasmid complement of *L. lactis* has revealed numerous regions similar to those found in lactobacilli, enterococci and streptococci, all of which may have been acquired by HGT (Siezen *et al.*, 2005; Górecki *et al.*, 2011). In plasmid pLP712, the D-lactate dehydrogenase gene, *dld*, was similar to that of *E. coli* and was suggested to be acquired by HGT (Wegmann *et al.*, 2012).

HGT is also evident amongst dairy organisms of other species and acquired genes can encode for important metabolic functions to allow these organisms to thrive in the milk environment. For example, *met*C is involved in the biosynthesis of methionine, a relatively rare amino acid in milk. Genome sequence analysis of three *St. thermophilus* strains has revealed the presence of a *cys*M2*met*C*cys*E2 gene cluster with a 95% identity to another industrial dairy organism, *Lactobacillus bulgaricus* (Bolotin *et al.*, 2004a). With advances in molecular sequencing technology, it has become possible to uncover genes which facilitate an organism's adaptation to challenging environments and the important role which HGT may play. The acquisition of new traits can occur via three different mechanisms: 1) transformation, 2) transduction and 3) conjugation. Transduction may play only a limited role in the adaptation of bacteria for two reasons; due to the small size of the DNA fragment capable of being transferred during transduction and the narrow host range of bacteriophage which could prevent the transfer of DNA over a broader population of

bacteria. In contrast to conjugation, transformation is a slower process as it requires the existence of free DNA and the achievement of a competent cellular state, and in general a transformation event is more difficult to detect (Verraes *et al.*, 2013). Therefore transduction and transformation will not be discussed further.

1.6.1.2. The role of mobile genetic elements

1.6.1.2.1. Plasmids

Of the three mechanisms which promote the transfer of genetic material between bacterial cells, conjugation allows for the transfer of the largest genomic segments (Bolotin *et al.*, 2004b). Conjugation includes the movement of both transposons and conjugative and non-conjugative plasmids (Verraes *et al.*, 2013). Conjugative plasmids are self-transmissible, however mobilisable plasmids require the presence of additional conjugative elements encoded by the recipient chromosome or helper plasmids (Lorenzo-Díaz *et al.*, 2014). Mobilizable plasmids often contain an origin of transfer (*oriT*) and genes encoding for mobilization and may also encode genes important for survival in harsher environments. For example, plasmid analysis of the raw-milk cheese isolate, *L. lactis* DPC3758, identified mobilization elements in pAF14 which also carried genes for heavy metal resistance (Fallico *et al.*, 2012). It is important to note that many plasmids cannot be transferred by conjugation, however, conjugal plasmids are common in industrial dairy lactococci (Coakley *et al.*, 1997). In terms of starter culture development, conjugation equates to a natural process of DNA transfer and can be exploited in the dairy industry to create desirable phenotypes while maintaining food grade status (Fallico *et al.*, 2012).

Plasmid-encoded genes can account for up to 4.8% of the total genome of LAB (Makarova, *et al.* 2006). Recent genomic analysis of *L. lactis* ssp. *cremoris* UC509.9

reveals a complement of eight plasmids ranging from 4,623 bp (pCIS1) to 80,592 bp (pCIS8) in size (Ainsworth *et al.*, 2013). A large plasmid complement has also been observed in non-dairy lactococci such as *L. lactis* CV56 which possesses five plasmids (pCV56A 44.0 kb, pCV56B 35.9 kb, pCV56C 31.4 kb, pCV56D 5.5 kb, pCV56E 2.2 kb) (Gao *et al.*, 2011). Overall, plasmid profile analysis of *Lactococcus* strains from different niches has shown that industrial dairy strains possessed a higher average plasmid complement (seven) than non-dairy strains (two), with a greater abundance of plasmids smaller than 10 kb (Kelly *et al.*, 2010). Some plasmids can also become integrated into the host chromosome, increasing the stability of the phenotype. Analysis of the Ce6 region of strain SK11 revealed a 15 kb insertion which was likely acquired via integration of a small plasmid (Kelly *et al.*, 2010). Within this region, two putative type III restriction/ modification systems were identified as well as genes associated with plasmid replication. Overall, this insertion may play a role in the enhanced bacteriophage resistance of this strain.

Plasmids found in industrial dairy *Lactococcus* strains encode a range of important technological traits such as proteinase activity, bacteriophage resistance, lactose utilisation and bacteriocin production (Gasson and Shearman, 2003; Mills *et al.*, 2006). Over the last years, genomic sequencing of lactococcal plasmids of 'wild-type' strains, has revealed genes with potential application to the dairy industry such as bacteriocin production, flavour enhancement and probiotic characteristics (Kelly *et al.*, 1998b; Fallico *et al.*, 2012; Le *et al.*, 2013). In *L. lactis* KF147 putative EPS producing genes were also identified which may play an important role in the survival of this strain in its native habitat (Siezen *et al.*, 2008). EPS biosynthesis has been reported as crucial for biofilm formation and plant colonization in the plant-associated bacteria, *Gluconacetobacter diazotrophicus* PAL5 (Meneses *et al.*, 2011). Certain

industrial dairy strains also possess the capacity to produce EPS which, we speculate, may be a remnant from when these strains occupied a plant niche. The use of EPS producing *L. lactis* has shown to be beneficial when applied to cheddar cheese making, improving the textural and cooking properties of cheese in reduced fat systems (Costa *et al.*, 2010). The existence of EPS producing strains in modern times may have been largely influenced by the continued selection of strains with this phenotype by dairy producers. Within the milk environment the retention of the EPS producing phenotype may have aided in defending the cell from phage attack as EPS production has been associated with blocking phage adsorption (Forde and Fitzgerald, 1999; 2003). Ainsworth *et al.* (2014b) reported that there is a lack of sequence data for non-dairy lactococcal plasmids. This may be overcome in the near future, however, by the increased isolation and genome sequencing of lactococci from non-dairy niches which may shed further light on specific metabolic traits (such as EPS production) which may be important for survival in a particular ecosystem.

The acquisition of plasmids or other mobile genetic elements via conjugation is dependent on a number of factors. These include environment-dependeent cell-to-cell contact between the bacterial strains, and plasmid compatibility (Verraes *et al.*, 2013). It is plausible that the acquisition of extra chromosomal elements by plant lactococcal strains from other bacteria following ingestion by cattle may have taken place (Salama *et al.*, 1995; Nomura *et al.*, 2006). The mammalian digestive tract presents an environment which favours the transfer of genetic material between lactococci and other bacteria. Following ingestion of plant material inhabited by lactococci, these organisms could potentially come into close contact with organisms of the resident gut microflora. Bolotin *et al.* (2004b) reported on a probable transfer event of the *ycdB* gene between lactococci and *Salmonella* via conjugation while other

studies have reported conjugation between lactococci and the microflora of the murine gut (Gruzza *et al.*, 1993; Gruzza *et al.*, 1994). Within the milk environment, *L. lactis* may have obtained plasmids from other species encoding important functions in milk. Genetic profiling of *Lactococcus* strains from dromedary's milk suggests that the ability to ferment citrate by some industrial dairy lactococcal starters is due to the acquisition of a Cit^+ plasmid from *Leuconostoc* species (Drici *et al.*, 2010). This suggests that the close association between organisms in the milk environment is a more important factor in determining the transfer of genetic material than the phylogenetic similarity between the strains.

1.6.1.2.2. Transposable elements

Transposable elements are commonly found in plasmids and in chromosomes of *Lactococcus*. From an evolutionary standpoint transposons and their corresponding insertion sequence (IS) elements are of significant interest due to their effect on gene regulation, expression and location (Mills *et al.*, 2006). It has been suggested that transposons have been crucial in the adaptation of lactococci to milk as many encode functions which many be beneficial to the strain (Romero and Klaenhammer, 1993). For instance, the genes involved in the production of lacticin 481 are plasmid encoded and flanked by two IS elements forming a composite transposon, which may give an advantage to the cell by eliminating other bacteria within close proximity in the milk environment (Dufour *et al.*, 2000). Similarly, restriction/ modification systems and abortive infection proteins may be found as potential transposable elements on conjugal plasmids, enabling the bacteria to resist phage attack (Romero and Klaenhammer, 1990).

Homologous genes of the oligopeptidase F, *pepF1*, have been identified on the chromosome of strains IL1403 and NCDO763, and on the lactose-proteinase plasmid of NCDO763 (Nardi *et al.*, 1997). Sequence comparison between the plasmid encoded gene of NCDO763 and chromosome encoded gene of strain IL1403, suggests that these genes are not divergently related but acquired by a recent transfer event. The presence of IS elements flanking the plasmid encoded gene, indicates the transfer to the plasmid possibly via IS-directed mobilisation. The plasmid may have then been transferred to different strains via conjugation as reported for lactose-proteinase plasmids (Mills *et al.*, 2006). This hypothesis was supported by the molecular characterisation of the plasmid of strain NCDO712 (Wegmann *et al.*, 2012). In non-dairy lactococci, the important role of transposons is also evident. *L. lactis* KF147 possesses a chromosomally-orientated conjugative transposon encoding alpha-galactoside metabolism, which has also been identified in other plant lactococcal strains (Kelly *et al.*, 1998a).

A relatively high proportion of DNA corresponding to IS elements has been identified in lactococcal genomes and has been found to be as much as 100 kb in *L. lactis* ssp. *cremoris* UC509.9 (Ainsworth *et al.*, 2013). IS elements are regarded as being important mediators of gene activation and deactivation and thus are important contributors to adaptation and evolution (Schneider and Lenski, 2004). Continuous propagation of an *ldh* negative mutant of strain NIZO9010 which displayed a reduced growth capacity in milk, led to a restoration of lactate dehydrogenase activity. In these bacteria, the insertion sequence IS*981* was located up-stream from the inactive gene, *ldhB*, resulting in transcription activation (Bongers *et al.*, 2003). De Visser *et al.* (2004) examined *L. lactis* ssp. *lactis* IL1403 under exposure to different growth and starvation conditions, and the resulting mutations arising from ISs. It was observed

that after numerous generations, the descendant cells possessed an increased ability to survive and reproduce compared to an ancestral strain, attributed in most cases to deletions arising from recombinations among IS*981*. The insertion of IS elements may also lead to the loss of a particular phenotype, as observed for the phage resistance genes on plasmid pGBK17 and the loss of an abortive infection mechanism by transposition of IS*1076* (Huang *et al.*, 1991; Polzin and McKay, 1991).

1.6.1.3. Gene loss and genome reduction

Examination of industrial dairy *Streptococcaceae* genomes shows that adaptation to milk has resulted in gene loss, along with a relatively high percentage of pseudogenes caused by nonsense mutations, deletions, truncations and/ or frameshifts. The recent elucidation of the industrial dairy *L. lactis* ssp. *cremoris* UC509.9 genome has revealed a high number of pseudogenes and transposons within the smallest, sequenced lactococcal genome (Ainsworth *et al.*, 2013). This relatively high number of pseudogenes has also been observed in other LAB species associated with the dairy environment, most notably *St. thermophilus* where it is estimated that as much as 10% of the genome is comprised of pseudogenes (Bolotin *et al.*, 2004a; Van de Guchte *et al.*, 2006). It has been theorized that strains currently in use in the dairy industry, have undergone such an extent of reductive evolution that they are no longer fit to survive in the wider environment (Kelly *et al.*, 2010). Similar to *L. lactis*, *Lactobacillus casei* inhabits many different environmental niches offering diverse substrates for growth. In this species, it has been proposed that environmental strains could be regarded as 'ecological generalists' while strains adapted to grow in milk, where less substrate diversity exists, could be regarded as 'niche specialists' and the same could be applied to *L. lactis* (Cai *et al.*, 2009). Two particular processes, which are crucial to the

survival of *L. lactis* in the environment and which have undergone significant alterations upon domestication are the biosynthesis of amino acids and the ability to ferment plant derived sugars.

1.6.2. Phenotypic traits associated with domestication

1.6.2.1. Protein utilisation

A key phenotypic trait of industrial dairy starters is their ability to degrade milk proteins and it is a key factor in the successful application of lactococcal species to dairy fermentations. The milk environment remains a relatively new niche for non-pathogenic organisms in comparison to the length of time that LAB have been in existence (Campbell-Platt, 1994). Strains from industrial dairy starters and other fermented products may be descended from a single founder event, which allowed this lineage to thrive in milk fermentations (Passerini *et al.*, 2010). This single founder event may have involved the acquisition of a plasmid which harboured genes inferring a competitive advantage to this organism over others in milk, such as the rapid release of amino acids from casein and/ or lactose utilization. *L. lactis* strains have been isolated from cow skin that possess a gene enabling them to utilise lactose and, upon introduction to milk, these strains may have had a significant advantage over other bacteria (Passerini *et al.*, 2010). However, Passerini *et al.* (2010) proposed that the acquisition of genes enabling the utilization of milk sugars and proteins are not the only key processes facilitating the adaptation to milk, but that the overall process of domestication involves numerous events involving plasmid gain and loss (Figure 2). In addition, the acquisition of plasmid-encoded phage resistance mechanisms may have led to a specific strain dominating the microbiota during cheese making and in

turn led to this strains being continually isolated and becoming domesticated to the milk environment.

Microorganisms isolated from nutrient-rich environments possess nutritional requirements which render them incapable of growth in media containing bare nutrients i.e. primarily minerals and a carbon source. Milk represents a nutrient-rich environment for microorganisms but contains very low levels of free amino acids and peptides. Typically the free amino acid content of milk is less than 100 mg l^{-1} (Alm, 1982). All of the amino acids required by industrial dairy LAB for growth are present in casein, however it is estimated that less than 1% of the total amino acids in casein is required (Kunji *et al.*, 1996). Industrial dairy lactococcal strains exhibit a remarkable reduction in their amino acid biosynthetic capacity. Therefore, in order to satisfy their amino acid requirements, these strains must recruit a proteolytic system that permits the breakdown of caseins and the liberation of amino acids. Such systems have been identified in non-dairy and industrial dairy lactococcal strains alike (Liu *et al.*, 2010).

In cheese manufacturing, proteolysis is considered one of the most important processes in the production of fermented dairy products from a flavour standpoint (Fox, 1989). The proteolytic system of *L. lactis* is composed of 1) a cell envelope proteinase, *prtP* (EC 3.4.21.96), 2) a series of specific peptide transport systems for various peptides and amino acids and 3) various intracellular peptidases that function to degrade peptides to amino acids (Kunji *et al.*, 1996; Pederson *et al.*, 1999). All cell envelope proteinase genes are found on plasmids which range in size from 13.4 kb to 100 kb, however, incidences of chromosomally-encoded proteinase genes have also been reported (Kok, 1990; Nissen-Meyer *et al.*, 1992). Proteinase-like genes have been identified in both industrial dairy and non-dairy strains alike (Liu *et al.*, 2010).

One possible explanation behind this finding is that the *prt*$^+$ phenotype was not as a result of HGT, but arose from changes in the binding site of a native protease enabling casein breakdown (Price *et al.*, 2012).

Once casein has been hydrolysed, free peptides and amino acids are taken into the cell through distinct transport systems. In *L. lactis* three transport systems have been described: two oligopeptide transporters (Opp, Opt) and a transporter for smaller peptides (DtpT). The Opp operon is composed of a membrane lipoprotein (*oppA*), two trans-membrane proteins (*oppB*, *oppC*) and two cytoplasmic ATP binding proteins (*oppD*, oppF) (Tynkkynen *et al.*, 1993). In contrast to the Opp transporter, Opt possesses *optS* and *optA*, two peptide-binding proteins (Lamarque *et al.*, 2011). Adaption to the milk environment is associated with an increased expression of the Opp oligopeptide transport system (Bachmann *et al.*, 2010; Bachmann *et al.*, 2012) and it has been shown previously that the growth rate of numerous peptidase-negative mutants in milk was significantly reduced (Mierau *et al.*, 1996). From an industrial standpoint, once peptides are taken into the cell, multiple peptidases function to liberate free amino acids which are further converted into flavour compounds by amino acid converting enzymes.

1.6.2.2. Lactose utilisation

Lactose is a sugar rarely encountered in the non-dairy environment and is primarily associated with mammals as exemplified by the absence of lactose utilisation genes in *L. lactis* isolates from non-dairy environments (Siezen *et al.*, 2010; Gao *et al.*, 2011). It is widely held that LAB are the dominant group of bacteria in bovine, buffalo, sheep and goat milk prior to pasteurisation (Quigley *et al.*, 2013). In dairy fermentations the rapid ability to generate lactic acid from lactose is a key

phenotypic trait of industrial dairy starters (Marshall, 1991). Lactose uptake by *L. lactis* is encoded by *lacEF* which forms the phosphoenolpyruvate-phosphotransferase system (PEP-PTS). This system functions in the transport of lactose across the cell membrane while simultaneously catalysing the formation of lactose-phosphate. The tagatose-6-phosphate enzymes encoded by *lacABCD*, function inside the cell to produce galactose-6-phosphate and glucose via the enzyme phospho-β-galactosidase (Mills *et al.*, 2006). In contrast to other LAB, *St. thermophilus* equips a galactose lactose antiporter for lactose uptake, located within the *gal-lac* operon. Once in the cell, lactose cleavage by β-galactosidase releases glucose and galactose monomers, with galactose being subsequently excreted from the cell.

1.6.2.3. The curious incidence of *lacE* and *prtP*

Lactose and casein are the principal sugar and protein sources found in milk. Yet strikingly, the genes involved in utilization of these substances can be readily lost from the cell (McKay *et al.*, 1972; McKay and Baldwin, 1974). The gene cluster involved in lactose fermentation and genes involved in casein breakdown are, in most cases, plasmid encoded. These genes can often be on the same plasmid which are normally relatively large (Siezen *et al.*, 2005; Wegmann *et al.*, 2012; Ainsworth *et al.*, 2013). Due to their size, these plasmids are prone to instability within the cell and can be spontaneously lost (McKay and Baldwin, 1974; Bachmann *et al.*, 2012). Plasmids carrying these genes have been found to be largely associated with transposable elements which may contribute to significant deletions and rearrangements leading to further instability of the phenotype (Mills *et al.*, 2006; Wegmann *et al.*, 2012).

A large proportion of the industrial dairy *L. lactis* strains examined by Passerini *et al.* (2010) lacked either genes encoding the extra-cellular proteinase (*prt*P) or those

involved in lactose utilization (*lac*E) or both. Of note, Passerini *et al.* (2010) identified genes involved in lactose fermentation in two *L. lactis* strains isolated from cow skin and hints as a possible route for adaptation of *L. lactis* from plants to milk. *lac*E-lactococci may have overcome the loss of lactose utilization by consuming other sugars released into the milk. The application of non-dairy *L. lactis* strains to mini Gouda-type cheese processing showed an increase in lactococcal cell numbers over ripening (Cavanagh *et al.*, 2014). The authors proposed that the reason for this observation was the utilization of carbohydrates released into the matrix by autolysis of other strains, which has been shown previously in LAB (Adamberg *et al.*, 2005).

Bachmann *et al.* (2011) examined the stabilization of the proteolytic trait in *L. lactis* and the influence that the concentration of peptides surrounding the cell had on the retention of the proteinase phenotype. In high peptide concentrations and low cell densities, the cell was found to maintain the trait; however at high cell densities, strains without the pressure of expressing the phenotype can uptake free peptides liberated by other *prt*$^+$ strains (Fig. 2). This would enable proteinase-negative strains to thrive in the milk environment and may account for the low proportion of strains expressing the proteinase phenotype. Genes encoding oligopeptide uptake and transport, and for the extra cellular proteinase, can be found on separate plasmids (Siezen *et al.*, 2005; Ainsworth *et al.*, 2014a). Despite possessing a plasmid harbouring *prt*P, the cell is unable to use casein as a protein source if a plasmid encoding the Opp operon is absent (Yu *et al.*, 1996). This would suggest that the plasmid encoding oligopeptide transport may have been acquired first and retained in the cell, allowing it to uptake free peptides in milk, released by *prt*$^+$ strains. Genome analysis of the proteolytic system of *L. lactis* identified three strains of plant origin which possess genes associated with the Opp transport system (Liu *et al.*, 2010). Upon chance contamination of milk by

such strains, these organisms may have been readily equipped to utilize free peptides available in the milk (Fig. 2). Although the Opp operon can be plasmid encoded as in strains IL594 and UC509.9 (Górecki *et al*., 2011; Ainsworth *et al*., 2013) over time it may have been integrated into the chromosome which may increase the stability of the *prt* phenotype as in *L. lactis* strains SSL135and IL1403, (Tynkkynen *et al.*, 1993; Bolotin *et al.*, 2001).

Bachmann *et al.* (2012) subsequently investigated the experimental evolution of a non-dairy *L. lactis* ssp. *lactis* strain, KF147, in milk. This strain was transformed with the *prt*$^{+}$ plasmid pNZ521, which was completely lost upon culturing in milk following 1,000 repeated propagations. The absence of an extra-cellular protease is also evident in other species associated with the milk environment; for example, *St. thermophilus* does not normally encode an extra cellular proteinase and as such requires other bacteria to be present with this characteristic. It is probable that the lack of an extra cellular proteinase may be due to the co-evolution of *Lactobacillus bulgaricus* and *St. thermophilus* in milk (Van de Guchte, *et al.* 2006). Descendant cells were subsequently isolated which showed an increased acidification rate in milk, compared to the KF147 parental strain. In industrial dairy strains however, the loss of the proteinase trait is associated with a lower growth and a reduced acidification rate in milk (McKay and Baldwin, 1974). Bachmann *et al.* (2012) suggested that the cell copes with the loss of the proteolytic trait by utilizing the small peptides and free amino acids available in milk more efficiently coupled with *de novo* amino acid biosynthesis. They also postulate that the environment from which the parental strain was isolated may be important in relation to the observed robustness of the strain in milk. Non-dairy strains, such as KF147, are from nutritionally poorer habitats in comparison to milk. Genome sequence analysis of this strain revealed the presence

of genes associated with non-ribosomal peptide or polyketide biosynthesis which may a role in survival and defence outside of milk. Within milk, such systems may enable these strains to cope with the loss of the extra-cellular proteinase and produce acid at a rate similar to *prt*$^{+}$ strains.

1.6.2.4. Amino acid auxotrophy in *Lactococcus*

Optimal growth in an amino acid deficient environment is dependent on the immediate activation and regulation of biosynthetic pathways for amino acids; however, this places a large demand upon the cell and should be suppressed immediately when not required (Chopin, 1993). When grown in milk bacteria can meet their amino acid requirements in a more efficient manner by utilising their proteolytic system. Thus, the repeated culturing of strains in milk has led to the in-activation of biosynthetic pathways for amino acids through mutation and gene loss, as these bacteria become more reliant on their proteolytic system (Delorme *et al.*, 1993; Godon *et al.*, 1993). The requirements of *L. lactis* strains for amino acids is strain dependant and a variation in amino acid requirements has been observed previously between subspecies, with *cremoris* strains showing a lower degree of auxotrophy (Chopin, 1992). In the past, disagreements arose around the amino acid requirements of *L. lactis*, possibly due to the use of defined media with different relative amino acid concentrations (Chopin, 1992). With the development of next generation sequencing technology enabling faster, more economic elucidation of bacterial genomes, the biosynthetic capacity of *L. lactis* can be determined through genome sequence analysis, helping to avoid such disagreements.

The amino acid requirements of *L. lactis* vary according to the environment from which the strain was isolated. Previously, it was determined that *L. lactis*

requires the presence of seven essential amino acids for growth: leucine, isoleucine, valine, serine, threonine, histidine and glutamate; however examination of 'wild-type' *L. lactis* strains revealed that these strains require as little as two essential amino acids for growth (Ayad *et al.*, 1999). These were predominantly glutamic acid and valine but for some strains their requirements also extended to methionine, histidine, leucine, aspartic acid and arginine (Ayad *et al.*, 1999). In order for a cell to synthesise the branched chain amino acids (BCAAs) leucine, isoleucine and valine, four enzymes are required for each amino acid, plus additional enzymes for isoleucine and leucine (Chopin, 1993) (Fig. 3). Previously, non-dairy strains have been found to be prototrophic for leucine and isoleucine, but not for valine. As all genes required for valine are also required for leucine and isoleucine it was proposed that this auxotrophy was attributed to the effect of the other BCAAs on valine biosynthesis (Godon *et al.*, 1993). *L. lactis* is capable of growth in media containing only free amino acids, the secondary transporters of which have recently been characterised (Jensen and Hammer, 1993; Trip *et al.*, 2013). It has been observed that certain strains of *L. lactis* only reach their maximum growth rate with the addition of exogenous amino acids suggesting that these transporters play a role in the up-take of free amino acids (Hugenholtz *et al.*, 1987). One explanation for this observation involves the utilisation of amino acids released by and/ or secreted by other organisms present in milk (Bachmann *et al.*, 2011).

Genome sequence analysis has revealed the dysfunction in genes associated with amino acid biosynthesis in industrial dairy *L. lactis* strains. Typically, industrial dairy *L. lactis* strains are auxotrophic for the BCAAs and histidine (Fig. 3). In the study performed by Bachmann *et al.* (2012), evolution of a non-dairy *lactis* strain to milk showed a decrease in BCAA gene expression which was attributed to a mutation

arising in the terminator/ antiterminator region of the BCAA operon. In strain MG1363, two dysfunctional genes were identified in the leucine biosynthetic pathway (*leu*A and *leu*B) which correlated with leucine auxotrophy (Wegmann *et al.*, 2007). Similar observations were reported in strain IL1403 but not strain SK11 (Makarova *et al.*, 2006). Analysis of industrial dairy *L. lactis* strains revealed that the BCAA locus was non-functional. The extent of this malfunction within the *leu* gene prohibits the return to a state of prototrophy (Godon, *et al.* 1993). This was not observed in lactobacilli, enterococci or pediococci whereby the alteration of a single base pair could lead to a state of prototrophy for numerous biosynthetic pathways. Histidine auxotrophy in *L. lactis* IL1403 was associated with point mutations in the histidine biosynthesis enzymes (Fefer *et al.*, 1998). Despite the appearance of complete, functional genes for methionine biosynthesis in strain MG1363, this strain was identified as auxotrophic for this amino acid (Fernández *et al.*, 2002; Flahaut *et al.*, 2013). Flahaut *et al.* (2013) proposed that discrepancies between model predictions based on genomic data may be caused by low enzymatic activities of the enzymes in question or by mutations arising in genes which are not detected by high-throughput annotation software.

The absence of biosynthetic capabilities for the majority of amino acids has been observed in industrial dairy organisms of other species also such as *Lactobacillus delbrueckii* ssp. *bulgaricus* and *Lactobacillus helveticus* (Van de Guchte *et al.*, 2006; Callanan *et al.*, 2008; Christiansen *et al.*, 2008; Cai *et al.*, 2009). Genome sequence analysis of *Lactobacillus helveticus* CNRZ32, revealed that the observed amino acid auxotrophies were due to the absence of some, or all, genes required for the biosynthesis of those amino acids (Christiansen *et al.*, 2008). *Lactococcus garvieae* was also found to possess a reduced capacity to synthesise amino acids, however this

was attributed to the pathogenic nature of the organism as observed in clostridia (Morita *et al.*, 2011). The high amino acid auxotrophy in *L. lactis* is in stark contrast to the prototrophy in *S. thermophilus* which requires only histidine and methionine or cysteine for growth (Pastink *et al.*, 2009). *In-silico* analysis of two *St. thermophilus* genomes has revealed that, with the exception of histidine, the genes encoding enzymes required for the biosynthesis of all amino acids are present (Iyer *et al.*, 2010). The maintenance of a large number of functional biosynthetic genes for amino acids, demonstrates the ability of this organism to satisfy a large proportion of its own amino acid requirements, in comparison to *L. lactis* which relies heavily on its proteolytic system.

1.6.2.5. Carbohydrate and sugar metabolism

As previously stated, the origin of industrial dairy starters is believed to be plant material (Kelly *et al.*, 2010), an environment which contains an array of carbohydrates grouped into structural (arabinan, xylan, pectin and cellulose) and storage polysaccharides (sucrose, starch and fructan). In contrast, milk contains relatively few carbohydrates with lactose being the primary sugar. Owing to the broad carbohydrate composition of non-dairy environments, it is only to be expected that strains from these particular niches are more metabolically diverse than their industrial dairy counterparts in terms of sugar utilisation. Transcriptome analysis of continuous propagations of *L. lactis* in milk uncovered a down-regulation in genes associated with the metabolism of plant-associated carbon sources (Bachmann *et al.*, 2012). This observation is consistent with the phenotypes of strains domesticated to the dairy environment. Previous analysis of industrial dairy *Lb. casei* strains showed that the domestication of these bacteria to milk was associated with a reduced capacity to

ferment sugars such as D-xylose and D-adonitol which were utilized by strains associated with plant and gastrointestinal (GI) environments (Cai *et al.*, 2009).

Genome analysis of the plant strains KF147 and KF282 show that they possess approximately 200 open reading frames (ORFs) that are absent in *L. lactis* ssp. *lactis* IL1403 and *L. lactis* ssp. cremoris SK11 (Van Hylckama Vlieg *et al.*, 2006). Within the *L. lactis* ssp. *lactis* KF147 genome, these additional genes appear to encode for a range of functions associated with the plant environment, particularly for the utilisation of plant derived carbohydrates. These identified genes were classed as ancestral and not acquired via HGT, and are absent in industrial dairy strains as a consequence of domestication to the milk environment (Siezen *et al.*, 2008). Conversely, IS elements present in *L. lactis* ssp. *lactis* IL1403 are absent in *L. lactis* ssp. *lactis* KF147, including a region encoding hypothetical proteins, phage defence proteins and transposases (Siezen *et al.*, 2008). One transposon was found in *L. lactis* KF147 that encodes α-galactoside utilisation and shared a 92% amino acid identity to a gene cluster found in *L. raffinolactis* (Boucher *et al.*, 2003). α-galactosidase catalyses the breakdown of plant oligosaccharides, releasing D-galactose moieties which may be subsequently used via the Leloir pathway (Van Hylckama Vlieg *et al.*, 2006). After D-xylose, L-Arabinose is the most common pentose monosaccharide found in plant polysaccharides such as pectin and hemicellulose (Seiboth and Metz, 2011). Genes encoding for arabinose metabolism have previously been reported in LAB strains associated with plants but not in industrial dairy lactococci (Van Hylckama Vlieg *et al.*, 2006). Continuous propagation of non-dairy *L. lactis* KF147 resulted in the loss of the transposon Tn*6098*, encoding genes involved in α-galactoside utilisation, which the authors suggested is not required for growth in milk

and the loss of which, may confer an advantage on the cell (Machielsen *et al.*, 2011; Bachmann *et al.*, 2012).

The majority of the six-carbon sugars and disaccharides utilised by lactococci and streptococci are incorporated into the cell via phosphotransferase systems (PTSs) (Price *et al.*, 2012). Genome sequence analysis of other industrial dairy organisms such as *St. thermophilus* has identified that the highest proportion of gene decay occurs in genes associated with transport proteins and the metabolism of carbohydrates. Four of seven PTS transporters were identified as pseudogenes while, in all, *St. thermophilus* strains CNRZ1066 and LMG13811 were found to harbour only a portion of the genes involved in sugar up-take found in other streptococci (Bolotin *et al.*, 2004a). The remaining of the transporters for cellobiose, sucrose and maltose have been identified in another milk associated organism, *Lactobacillus bulgaricus*, which may have allowed this organism to utilise plant associated sugars, prior to inhabiting milk and suggests a possible origin for the strain (Van de Guchte *et al.*, 2006). Siezen *et al.* (2008) investigated the ability of two non-dairy isolates and two industrial dairy strains to utilise an array of mono-, oligo- and poly-saccharides. Both strains IL1403 and SK11 demonstrated a reduced capacity to grow on sugar substrates while only the two non-dairy isolates could grow on L-arabinose, glucuronate, mannitol, gluconate and D-xylose. *L. lactis* A12 displays an increased ability to ferment arabinose and raffinose, sugars not generally considered to be efficiently utilised by lactobacilli and yeasts (Corsetti *et al.*, 2001). The ability of strain A12 to efficiently metabolise these sugars could offer a competitive advantage to this organism in the sourdough environment (Passerini *et al.*, 2013a).

1.7. Applications of 'wild-type' strains in fermented foods

1.7.1. Documented food applications

Non-dairy lactococci possess an array of metabolic characteristics which may be exploited in dairy fermentations to enhance the organoleptic properties or safety of a fermented dairy product. The use of microorganisms that produce metabolites which inhibit the growth of spoilage bacteria dates back thousands of years (Ross *et al.*, 2002). From a food safety stand point, many wild strains have been isolated from different environmental niches which secrete anti-microbial compounds that could prolong the shelf life of some fermented foods. In one study, analysis of the production of anti-microbial compounds by 'wild-type' strains, from dairy and non-dairy origins, found that 32 of 79 isolates inhibited the growth of indicator strains (Ayad *et al.*, 2002). The technological traits of these isolates also showed that these strains could be applied to cheese production. In traditional raw-milk cheeses, 17 strains of *L. lactis* were shown to produce anti-microbial compounds (Alegría *et al.*, 2010).

The flavour forming capability of 'wild-type' (from raw-milk or non-dairy environments) *Lactococcus lactis* strains in dairy fermentations has gained increased interest in recent times due to the diverse aromas such strains may be capable of imparting. Within the cheese environment, liberated amino acids act as precursors to an array of flavour compounds generated by microbial conversion (Yvon and Rijnen, 2001). The conversion of amino acids to aroma compounds begins with an aminotransferase reaction, with the transfer of an amino group to α-ketogluturate and the formation of an α-keto acid. Previously, a raw-milk cheese and non-dairy *Lactococcus* isolates have been shown to produce the enzyme glutamate dehydrogenase (*gdh*) that catalyses the conversion of glutamate to α-ketogluturate

(Tanous *et al.*, 2002; Fallico *et al.*, 2011). This compound has been identified as being rate-limiting in the formation of aroma compounds from amino acids (Tanous *et al.*, 2002). In non-dairy strains, this enzyme is thought to be involved in the biosynthesis of glutamate in the presence of high levels of ammonia (Tanous *et al.*, 2002). When this gene was cloned into a *gdh*$^{-}$ *L. lactis* strain, the formation of volatile flavour compounds was increased, confirming the important role of this enzyme in the formation of aroma compounds from amino acids.

The biosynthesis and degradation of amino acids are intertwined pathways (for reviews see Smit *et al.* 2002 and van Kranenburg *et al.* 2002); for example, cystathionine β-lyase converts methionine to volatile flavour compounds during cheese ripening. Outside of the cheese matrix, the physiological role of this enzyme is to catalyse the conversion of cystathionine to homocysteine (Smit *et al.*, 2005). The amino acid converting ability of LAB varies greatly from strain to strain which is linked to the organism's capacity for the biosynthesis of amino acids (Smit *et al.*, 2005). It has already been mentioned that wild lactococci require significantly fewer amino acids for growth than industrial dairy strains. The maintenance of these metabolic pathways in wild cultures has been associated with the ability of these strains to produce diverse aromas and/ or aroma profiles in cheese models and in pilot scale cheeses (Ayad *et al.*, 1999; Ayad *et al.*, 2000). Analysis of volatile compounds in milk models has shown that non-dairy strains produce a larger number of compounds in comparison to industrial dairy strains and more diverse aroma profiles (Alemayehu *et al.*, 2014; Cavanagh *et al.*, 2014). Although some of the aromas may be considered off-flavours in certain cheese varieties, they may enhance the overall flavour profile when used in combination with established industrial dairy starters.

1.7.2. Potential restrictions on the use of non-dairy LAB in foods

L. lactis possesses generally regarded as safe status (GRAS) which recognises the safety of a substance under its intended use according to the Food, Drug and Cosmetic Act (1958) (FDA, 2010). As fermented dairy foods were widely consumed prior to the establishment of this act, bacteria used in these processes were held to have GRAS status and as such may not be listed on the FDA website (FDA, 2001). Under this framework, the application of particular bacteria to a process is awarded GRAS status and not the bacteria itself (Bourdichon *et al.*, 2012). Therefore, when using a bacterium with a long history of use to a new application or at a considerably higher dosage, GRAS status needs to be established. In the European Union (EU), the European Food Safety Authority (EFSA) has established 'Qualified Presumption of Safety (QPS)', which acts as an overall pre-assessment to assist EFSA's scientific panel in safety risk assessment (EFSA, 2007). In the EFSA QPS update 2013, despite some reported cases of *L. lactis* isolation from clinical samples the QPS recommendation for *L. lactis* was maintained (EFSA, 2013). In this update, recommendations were also put forward for LAB clinical isolates, *L. lactis* in particular, in order to identify any strain-specific virulence factors.

An important issue, with regard to the isolation of novel strains from the environment and their use as food additives, is the potential reservoir of antibiotic resistance genes they may harbour. Genes encoding antibiotic resistance acquired via transposons or plasmids are capable of being transferred to other bacteria, either commensals or pathogens (Nawaz *et al.*, 2011). In a recent report by the World Health Organisation (WHO, 2014) into the surveillance of antimicrobial resistance, the shortcomings in antibiotic resistance monitoring in humans and the requirement of surveillance integration between humans and along the food chain were highlighted.

EFSA standards dictate that an overall qualification for a taxonomic unit for QPS status, is the absence of acquired antibiotic resistance genes which is highlighted as being an issue for LAB. The EFSA QPS update (2013) recommends that antibiotic resistance be established in accordance with international standards and guidelines such as ISO 10932/ IDF 223 (2010) and the Clinical and Laboratory Standards Institute (CLSI, 2007).

Biogenic amines are nitrogenous organic bases formed by regular metabolic activities of microorganisms, humans and animals resulting from the decarboxylation of amino acids (Alvarez and Moreno-Arribas, 2014). When ingested at high concentrations, they can have an deleterious effect on the human body with symptoms including headache, heart palpitations and nausea (Ladero *et al.*, 2010). Ladero *et al.* (2011) found two of 20 *L. lactis* strains capable of producing putrescine via the agamatine deiminase pathway. The gene cluster involved was characterised and appears not to have been acquired recently by HGT. This suggests that the genes involved were present prior to domestication to the milk environment and may be present in non-dairy lactococci also as observed in strain KF147 (Ladero *et al.*, 2011). A PCR-based method has been developed for the rapid identification of putrescine producing strains which may aid in the screening of isolates for this phenotype from the non-dairy environment (Ladero *et al.*, 2011).

1.8. Conclusions

The purpose of this review was to examine the domestication of *Lactococcus* to the milk environment and some of the characteristics associated with this adaptation. In the past a limited number of *L. lactis* genomes were available, and these were primarily associated with cheese production. In the last number of years numerous genome sequences have been elucidated and the isolation of *L. lactis* from more diverse environments and subsequent genome analysis has offered greater insight into the core and dispensable genes of this organism. Industrial dairy *L. lactis* strains have undergone significant genome decay and are considered to be unable to survive outside of this niche (Kelly *et al.*, 2010); however, this adaptation has also been coupled with the acquisition of genes by HGT, such as lactose fermentation and casein utilization, which may allow this organism to thrive in this environment. The development of high-throughput screening methods in conjunction with improved sequencing technology equips scientists with the necessary tools to fully exploit this natural diversity and this could have many applications for the development for tailored fermented products in the future.

1.9. Future perspectives

The diversity observed in *Lactococcus* strains from different ecological niches could contribute greatly to enhancing the safety and flavour profiles of fermented dairy products. If such strains are to be routinely used in processing one must prevent the loss of these natural metabolic traits and the domestication of these microorganisms. One option could be to maintain cultures in minimal media prior to growing the cultures in milk and inoculating vats during cheese production; however, this may significantly increase production costs and may make the use of such strains unfeasible. From a consumer standpoint, acceptability of dairy products made with cultures from diverse environments remains to be established and some regular cheese consumers may not welcome such products. The issue of antibiotic resistance must also be addressed due to the ongoing use of antibiotics for the treatment of bacterial infections in animals and their subtherapeutic use to enhance growth and feed efficiency.

Yet there is potential application of these strains in the production of fermented foods for flavour diversity by the production of unusual aromas and enhanced safety by the production of anti-microbial compounds. Plasmid analysis of *L. lactis* from raw-milk cheese was found to encode for tetracycline resistance (*tetS*) but also for many genes which may be beneficial in dairy fermentations (Fallico *et al.*, 2011). By curing the strain of the antibiotic encoding plasmid without genetic modification, the strain could be used in fermentations without the fear of disseminating antibiotic resistance. *L. lactis* has a long documented safe usage in fermented products and numerous studies have investigated the capacity of this species to inhibit the growth of other bacteria (Parapouli *et al.*, 2013; Alegría *et al.*, 2010; Biscola *et al.*, 2013). Although not considered a normal inhabitant of the human gastrointestinal tract, the

emergence of non-dairy lactococci with potential probiotic properties (capacity to adhere to human colonic epithelial cells, and tolerance to low pH and bile (Kimoto *et al.*, 1999)); such isolates could be used for the production of cheeses or other fermented products with added health benefits. In societies moving towards the increased consumption of convenience foods, the ingestion of organisms naturally found on plants and vegetables may decrease. In many Asian countries fermented plant products are widely consumed and the probiotic potential of lactococci from such niches has been examined (Kimoto *et al.*, 1999; Kimoto *et al.*, 2004). Recently a non-dairy *Lactococcus* isolate was found to possess surface pili with the ability to bind to human intestinal epithelial cells (Meyrand *et al.*, 2013). The addition of organisms from these environments to other, more widely consumed products such as cheese, may help sustain the natural diversity of the gut microbiota. Similarly, the natural anti-microbial compounds produced from non-dairy strains may enhance the shelf life of food products without the addition of additives or preservatives (Biscola *et al.*, 2013; Parapouli *et al.*, 2013).

1.10. Acknowledgements

Thanks to Aidan Casey for assistance in ANI analysis using JSpecies.

1.11. References

Adamberg, K., Antonsson, M., Vogensen, F. K., Nielsen, E. W., Kask, S., Møller, P. L., and Ardö, Y. (2005). Fermentation of carbohydrates from cheese sources by non-starter lactic acid bacteria isolated from semi-hard Danish cheese. *Int. Dairy J.* 15**,** 873-882. DOI: 10.1016/j.idairyj.2004.07.017.

Ainsworth, S., Mahony, J., and Van Sinderen, D. (2014a). The plasmid complement of *Lactococcus lactis* UC509. 9 encodes multiple bacteriophage resistance systems. *Appl. Environ. Microbiol*. (In press) DOI: 10.1128/AEM.01070-01014.

Ainsworth, S., Stockdale, S., Bottacini, F., Mahony, J., and Van Sinderen, D. (2014b). The *Lactococcus lactis* plasmidome: much learnt, yet still lots to discover. *FEMS Microbiol. Rev* 38, 1066-1088. DOI: 10.1111/1574-6976.12074.

Ainsworth, S., Zomer, A., De Jager, V., Bottacini, F., Van Hijum, S. A., Mahony, J., and Van Sinderen, D. (2013). Complete genome of *Lactococcus lactis* subsp. *cremoris* UC509. 9, host for a model lactococcal P335 bacteriophage. *Genome announcements.* 1**,** e00119-00112. DOI: 10.1128/genomeA.00119-12.

Alegría, Á., Delgado, S., Roces, C., López, B., and Mayo, B. (2010). Bacteriocins produced by wild *Lactococcus lactis* strains isolated from traditional, starter-free cheeses made of raw milk. *Int. J. Food Microbiol.* 143**,** 61-66. DOI: 10.1016/j.ijfoodmicro.2010.07.029.

Alemayehu, D., Hannon, J. A., McAuliffe, O., and Ross, R. P. (2014). Characterization of plant-derived lactococci on the basis of their volatile compounds profile when grown in milk. *Int. J. Food Microbiol.* 172**,** 57-61. DOI: 10.1016/j.ijfoodmicro.2013.11.024.

Alm, L. (1982). Effect of fermentation on proteins of Swedish fermented milk products. *J. Dairy Sci.* 65**,** 1696-1704. DOI: 10.3168/jds.S0022-0302(82)82404-1.

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic acids Res.* 25**,** 3389-3402.

Alvarez, M. A., and Moreno-Arribas, M. (2014). The problem of biogenic amines in fermented foods and the use of potential biogenic amine-degrading microorganisms as a solution. *Trends Food Sci. Tech.* DOI: 10.1016/j.tifs.2014.07.007.

Asensi, G. F., De Sales, N. F. F., Dutra, F. F., Feijó, D. F., Bozza, M. T., Ulrich, R. G., Miyoshi, A., De Morais, K., De Carvalho Azevedo, V. A., and Silva, J. T. (2013). Oral immunization with *Lactococcus lactis* secreting attenuated recombinant staphylococcal enterotoxin B induces a protective immune response in a murine model. *Microbial Cell Fact.* 12**,** 32. DOI: 10.1186/1475-2859-12-32.

Ayad, E., Verheul, A., Wouters, J., and Smit, G. (2000). Application of wild starter cultures for flavour development in pilot plant cheese making. *Int. Dairy J.* 10**,** 169-179. DOI: 10.1016/s0958-6946(00)00041-8.

Ayad, E. H. E., Verheul, A., De Jong, C., Wouters, J. T. M., and Smit, G. (1999). Flavour forming abilities and amino acid requirements of *Lactococcus lactis* strains isolated from artisanal and non-dairy origin. *Int. Dairy J.* 9**,** 725-735.10.1016/s0958-6946(99)00140-5.

Ayad, E. H. E., Verheul, A., Wouters, J. T. M., and Smit, G. (2002). Antimicrobial-producing wild lactococci isolated from artisanal and non-dairy origins. *Int. Dairy J.* 12**,** 145-150. DOI: 10.1016/s0958-6946(01)00133-9.

Bachmann, H., De Wilt, L., Kleerebezem, M., and Van Hylckama Vlieg, J. E. (2010). Time resolved genetic responses of *Lactococcus lactis* to a dairy environment. *Environ. Microbiol.* 12**,** 1260-1270. DOI: 10.1111/j.1462-2920.2010.02168.x.

Bachmann, H., Molenaar, D., Kleerebezem, M., and Van Hylckama Vlieg, J. E. (2011). High local substrate availability stabilizes a cooperative trait. *ISME J.* 5**,** 929-932. DOI: 10.1038/ismej.2010.179.

Bachmann, H., Starrenburg, M. J., Molenaar, D., Kleerebezem, M., and Van Hylckama Vlieg, J. E. (2012). Microbial domestication signatures of *Lactococcus lactis* can be reproduced by experimental evolution. *Genome Res.* 22**,** 115-124.

Baltrus, D. A. (2013). Exploring the costs of horizontal gene transfer. *Trends Ecol. Evol.* 28**,** 489-495. DOI: 10.1016/j.tree.2013.04.002.

Beimfohr, C., Ludwig, W., and Schleifer, K.-H. (1997). Rapid genotypic differentiation of *Lactococcus lactis* subspecies and Biovar. *Syst. Appl. Microbiol.* 20**,** 216-221. DOI: 10.1016/S0723-2020(97)80068-9.

Biscola, V., Todorov, S., Capuano, V., Abriouel, H., Gálvez, A., and Franco, B. (2013). Isolation and characterization of a nisin-like bacteriocin produced by a *Lactococcus lactis* strain isolated from charqui, a Brazilian fermented, salted and dried meat product. *Meat Sci.* 93**,** 607-613. DOI: 10.1016/j.fm.2014.06.025.

Bolotin, A., Quinquis, B., Ehrlich, S. D., and Sorokin, A. (2012). Complete genome sequence of *Lactococcus lactis* subsp. *cremoris* A76. *J. Bacteriol.* 194**,** 1241-1242. DOI: 10.1128/JB.06629-11.

Bolotin, A., Quinquis, B., Renault, P., Sorokin, A., Ehrlich, S. D., Kulakauskas, S., Lapidus, A., Goltsman, E., Mazur, M., and Pusch, G. D. (2004a). Complete sequence and comparative genome analysis of the dairy bacterium *Streptococcus thermophilus*. *Nat. Biotechnol.* 22**,** 1554-1558. DOI: 10.1038/nbt1034.

Bolotin, A., Quinquis, B., Sorokin, A., and Ehrlich, D. S. (2004b). Recent genetic transfer between *Lactococcus lactis* and enterobacteria. *J. Bacteriol.* 186**,** 6671-6677. DOI: 10.1128/JB.186.19.6671-6677.2004.

Bolotin, A., Wincker, P., Mauger, S., Jaillon, O., Malarme, K., Weissenbach, J., Ehrlich, S. D., and Sorokin, A. (2001). The complete genome sequence of the lactic acid bacterium *Lactococcus lactis* ssp. *lactis* IL1403. *Genome Res.* 11**,** 731-753. DOI: 10.1101/gr.169701.

Bongers, R. S., Hoefnagel, M. H., Starrenburg, M. J., Siemerink, M. A., Arends, J. G., Hugenholtz, J., and Kleerebezem, M. (2003). IS*981*-mediated adaptive evolution recovers lactate production by *ldhB* transcription activation in a lactate dehydrogenase-deficient strain of *Lactococcus lactis*. *J. Bacteriol.* 185**,** 4499-4507. DOI: 10.1128/JB.185.15.4499-4507.2003.

Boucher, I., Vadeboncoeur, C., and Moineau, S. (2003). Characterization of genes involved in the metabolism of α-galactosides by *Lactococcus raffinolactis*. *Appl. Environ. Microbiol.* 69**,** 4049-4056. DOI: 10.1128/AEM.69.7.4049-4056.2003.

Bourdichon, F., Casaregola, S., Farrokh, C., Frisvad, J. C., Gerds, M. L., Hammes, W. P., Harnett, J., Huys, G., Laulund, S., and Ouwehand, A. (2012). Food fermentations: microorganisms with technological beneficial use. *Int. J. Food Microbiol.* 154**,** 87-97. DOI: 10.1016/j.ijfoodmicro.2011.12.030.

Cai, H., Thompson, R., Budinich, M. F., Broadbent, J. R., and Steele, J. L. (2009). Genome sequence and comparative genome analysis of *Lactobacillus casei*: insights into their niche-associated evolution. *Genome Biol. Evol.* 1**,** 239-257. DOI: 10.1093/gbe/evp019.

Cai Y, Yang J, Pang H & Kitahara M (2011) *Lactococcus fujiensis* sp nov., a lactic acid bacterium isolated from vegetable. *Int. J. Syst. Evol. Microbiol.* 61, 1590–1594.

Callanan, M., Kaleta, P., O'Callaghan, J., O'Sullivan, O., Jordan, K., McAuliffe, O., Sangrador-Vegas, A., Slattery, L., Fitzgerald, G. F., and Beresford, T. (2008). Genome sequence of *Lactobacillus helveticus*, an organism distinguished by selective gene loss and insertion sequence element expansion. *J. Bacteriol.* 190**,** 727-735. DOI: 10.1128/JB.01295-07.

Campbell-Platt, G. (1994). Fermented foods- a world perspective. *Food Res. Int.* 27**,** 253-257. DOI: 10.1016/0963-9969(94)90093-0.

Cavanagh, D., Kilcawley, K. N., O'Sullivan, M. G., Fitzgerald, G. F., and McAuliffe, O. (2014). Assessment of wild non-dairy lactococcal strains for flavour diversification in a mini-Gouda type cheese modeL. *Food Res. Int.* 62**,** 432-440. DOI: 10.1016/j.foodres.2014.03.043.

Chen, Y.-S., Chang, C.-H., Pan, S.-F., Wang, L.-T., Chang, Y.-C., Wu, H.-C., and Yanagida, F. (2013). *Lactococcus taiwanensis* sp. *nov*., a lactic acid bacterium isolated from fresh cummingcordia. *Int. J. Syst. Evol. Microbiol.* 63**,** 2405-2409. DOI: 10.1099/ijs.0.045757-0.

Chen, Y. S., Otoguro, M., Lin, Y. H., Pan, S. F., Ji, S.H., Yu, C. R., Liou, M. S., Chang, Y. C., Wu, H.C., and Yanagida, F. (2014). *Lactococcus formosensis* sp. *nov*., a lactic acid bacterium isolated from *yan-tsai-shin* (fermented broccoli stems). *Int. J. Syst. Evol. Microbiol.* 64**,** 146-151. DOI: 10.1099/ijs.0.052811-0.

Cho S. L., Nam S. W., Yoon J. H., Lee J. S., Sukhoom A. & Kim W. (2008) *Lactococcus chungangensis* sp. nov., a lactic acid bacterium isolated from activated sludge foam. *Int. J. Syst. Evol. Microbiol*. 58, 1844–1849.

Chopin, A. (1993). Organization and regulation of genes for amino acid biosynthesis in lactic acid bacteria. *FEMS Microbiol. Rev.* 12**,** 21-37. DOI: 10.1111/j.1574-6976.1993.tb00011.x.

Christiansen, J. K., Hughes, J. E., Welker, D. L., Rodríguez, B. T., Steele, J. L., and Broadbent, J. R. (2008). Phenotypic and genotypic analysis of amino acid auxotrophy in *Lactobacillus helveticus* CNRZ 32. *Appl. Environ. Microbiol.* 74**,** 416-423. DOI: 10.1128/AEM.01174-07.

Clair, E., Linn, L., Travert, C., Amiel, C., Séralini, G.-E., and Panoff, J.-M. (2012). Effects of Roundup® and glyphosate on three food microorganisms: *Geotrichum candidum*, *Lactococcus lactis* subsp. *cremoris* and *Lactobacillus delbrueckii* subsp. *bulgaricus*. *Curr. Microbiol.* 64**,** 486-491. DOI: 10.1007/s00284-012-0098-3.

CLSI, (2007). Clinical and Laboratory Standards Institute (CLSI). Methods for antimicrobial dilution and disk susceptibility testing of infrequently isolated or fastidious bacteria; approved guideline. CLSI document M45-A (ISBN 1-56238-607-7). CLSI, Wayne, Pennsylvania.

Coakley, M., Fitzgerald, G., and Ross, R. P. (1997). Application and evaluation of the phage resistance-and bacteriocin-encoding plasmid pMRC01 for the improvement of dairy starter cultures. *Appl. Environ. Microbiol.* 63**,** 1434-1440.

Corsetti, A., Lavermicocca, P., Morea, M., Baruzzi, F., Tosti, N., and Gobbetti, M. (2001). Phenotypic and molecular identification and clustering of lactic acid bacteria and yeasts from wheat (species *Triticum durum* and *Triticum aestivum*) sourdoughs of Southern Italy. *Int. J. Food Microbiol.* 64**,** 95-104. DOI: 10.1016/S0168-1605(00)00447-5.

Costa, N., Hannon, J., Guinee, T., Auty, M., Mcsweeney, P., and Beresford, T. (2010). Effect of exopolysaccharide produced by isogenic strains of *Lactococcus lactis* on half-fat Cheddar cheese. *J. Dairy Sci.* 93**,** 3469-3486.

Cunin, R., Glansdorff, N., Pierard, A., and Stalon, V. (1986). Biosynthesis and metabolism of arginine in bacteria. *Microbiol. Rev.* 50**,** 314.

Curioni, P., and Bosset, J. (2002). Key odorants in various cheese types as determined by gas chromatography-olfactometry. *Int. Dairy J.* 12**,** 959-984. DOI: 10.1016/S0958-6946(02)00124-3.

De Visser, J. A. G., Akkermans, A. D., Hoekstra, R. F., and De Vos, W. M. (2004). Insertion-sequence-mediated mutations isolated during adaptation to growth and starvation in *Lactococcus lactis*. *Genetics.* 168**,** 1145-1157. DOI: 10.1534/genetics.104.032136.

Delorme, C., Godon, J.-J., Ehrlich, S. D., and Renault, P. (1993). Gene inactivation in *Lactococcus lactis*: histidine biosynthesis. *J. Bacteriol.* 175**,** 4391-4399.

Douillard, F. P., Ribbera, A., Kant, R., Pietilä, T. E., Järvinen, H. M., Messing, M., Randazzo, C. L., Paulin, L., Laine, P., and Ritari, J. (2013). Comparative genomic and functional analysis of 100 *Lactobacillus rhamnosus* strains and their comparison with strain GG. *PLoS genetics.* 9**,** e1003683. DOI: 10.1371/journal.pgen.1003683.

Drici, H., Gilbert, C., Kihal, M., and Atlan, D. (2010). Atypical citrate-fermenting *Lactococcus lactis* strains isolated from dromedary's milk. *J. Appl. Microbiol.* 108**,** 647-657. DOI: 10.1111/j.1365-2672.2009.04459.x.

Drider, D., Bekal, S., and Prévost, H. (2004). Genetic organization and expression of citrate permease in lactic acid bacteria. *Genet. Mol. Res* 3**,** 271-281.

Du, Y., Song, L., Feng, W., Pei, G., Zheng, P., Yu, Z., Sun, J., and Qiao, J. (2013). Draft genome sequence of *Lactococcus lactis* subsp. *lactis* strain YF11. *Genome Announc.* 1**,** e00599-00513. DOI: 10.1128/genomeA.00599-13.

Dufour, A., Rincé, A., Uguen, P., and Le Pennec, J.-P. (2000). IS*1675*, a novel lactococcal insertion element, forms a transposon-like structure including the lacticin 481 lantibiotic operon. *J. Bacteriol.* 182**,** 5600-5605. DOI: 10.1128/JB.182.19.5600-5605.2000.

EFSA, (2007). Opinion of the Scientific Committee on a request from EFSA on the introduction of a Qualified Presumption of Safety (QPS) approach for assessment of selected microorganisms referred to EFSA. *The EFSA Journal*. 587, 1-16.

EFSA BIOHAZ Panel (EFSA Panel on Biological Hazards), (2013). Scientific Opinion on the maintenance of the list of QPS biological agents intentionally added to food and feed (2013 update). *The EFSA Journal*. 2013;11(11): 3449, 108 pp. DOI:10.2903/j.efsa.2013.3449.

Erkus, O., De Jager, V. C., Spus, M., Van Alen-Boerrigter, I. J., Van Rijswijck, I. M., Hazelwood, L., Janssen, P. W., Van Hijum, S. A., Kleerebezem, M., and Smid, E. J. (2013). Multifactorial diversity sustains microbial community stability. *ISME J.* 7**,** 2126-2136. DOI: 10.1038/ismej.2013.108.

Falentin, H., Naquin, D., Loux, V., Barloy-Hubler, F., Loubière, P., Nouaille, S., Lavenier, D., Le Bourgeois, P., François, P., and Schrenzel, J. (2014). Genome sequence of *Lactococcus lactis* subsp. *lactis* bv. *diacetylactis* LD61. *Genome Announc.* 2**,** e01176-01113. DOI: 10.1128/genomeA.01176-13.

Fallico, V., McAuliffe, O., Fitzgerald, G. F., and Ross, R. P. (2011). Plasmids of raw milk cheese isolate *Lactococcus lactis* subsp. *lactis* biovar. *diacetylactis* DPC3901 suggest a plant-based origin for the strain. *Appl. Environ. Microbiol.* 77**,** 6451-6462. DOI: 10.1128/AEM.00661-11.

Fallico, V., Ross, R. P., Fitzgerald, G., and McAuliffe, O. (2012). Novel conjugative plasmids from the natural isolate *Lactococcus lactis* subspecies *cremoris* DPC3758: A repository of genes for the potential improvement of dairy starters. *J. Dairy Sci.* 95**,** 3593-3608.

FDA, (2001). Partial list of microorganisms and microbial-derived ingredients that are used in foods. FDA http://www.fda.gov/Food/FoodingredientsPackaging/ucm078956.htm. [Accessed 7 June 2014].

FDA, (2010). Generally Recognised as Safe (GRAS) notifications. FDA http://www.fda.gov/ANimalVeterinary/Products/AnimalFoodFeeds/GenerallyRecognisedasSafeGRASNotifications/default.htm.

Fefer, J., Ratzan, K., Sharp, S., and Saiz, E. (1998). *Lactococcus garvieae* endocarditis: report of a case and review of the literature. *Diagn. Micr. Infec. Dis.* 32**,** 127-130. DOI: 10.1016/S0732-8893(98)00065-0.

Fernández, E., Alegría, Á., Delgado, S., Martín, M. C., and Mayo, B. (2011). Comparative phenotypic and molecular genetic profiling of wild *Lactococcus lactis* subsp. l*actis* strains of the *L. lactis* subsp. *lactis* and *L. lactis* subsp. *cremoris* genotypes, isolated from starter-free cheeses made of raw milk. *Appl. Environ. Microbiol.* 77**,** 5324-5335. DOI: 10.1128/AEM.02991-10.

Fernández, M., Kleerebezem, M., Kuipers, O. P., Siezen, R. J., and Van Kranenburg, R. (2002). Regulation of the metC-cysK operon, involved in sulfur metabolism in *Lactococcus lactis*. *J. Bacteriol.* 184**,** 82-90. DOI: 10.1128/JB.184.1.82-90.2002.

Ferrario, C., Ricci, G., Milani, C., Lugli, G. A., Ventura, M., Eraclio, G., Borgo, F., and Fortina, M. G. (2013). *Lactococcus garvieae*: where is it from? A first approach to explore the evolutionary history of this emerging pathogen. *PloS one.* 8**,** e84796. DOI: 10.1371/journal.pone.0084796.

Flahaut, N. A., Wiersma, A., Van De Bunt, B., Martens, D. E., Schaap, P. J., Sijtsma, L., Dos Santos, V. a. M., and De Vos, W. M. (2013). Genome-scale metabolic model for *Lactococcus lactis* MG1363 and its application to the analysis of flavor formation. *Appl. Microbiol. Biotechnol.* 97**,** 8729-8739. DOI: 10.1007/s00253-013-5140-2.

Forde, A., and Fitzgerald, G. F. (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.

Forde, A., and Fitzgerald, G. F. (2003). Molecular organization of exopolysaccharide (EPS) encoding genes on the lactococcal bacteriophage adsorption blocking plasmid, pCI658. *Plasmid.* 49**,** 130-142.

Fox, P. (1989). Proteolysis during cheese manufacture and ripening. *J. Dairy Sci.* 72**,** 1379-1400. DOI: 10.3168/jds.S0022-0302(89)79246-8.

Gao, Y., Lu, Y., Teng, K.-L., Chen, M.-L., Zheng, H.-J., Zhu, Y.-Q., and Zhong, J. (2011). Complete genome sequence of *Lactococcus lactis* ssp. *lactis* CV56, a probiotic strain isolated from the vaginas of healthy women. *J. Bacteriol.* 193**,** 2886-2887. DOI: 10.1128/JB.00358-11.

Garde, S., Babin, M., Gaya, P., Nuñez, M., and Medina, M. (1999). PCR amplification of the gene *acmA* differentiates *Lactococcus lactis* subsp. *lactis* and *L. lactis* subsp. *cremoris*. *Appl. Environ. Microbiol.* 65**,** 5151-5153.

Garvie, E. I., and Farrow, J. A. (1982). Notes: *Streptococcus lactis* subsp. *cremoris* (Orla-Jensen) comb. nov. and *Streptococcus lactis* subsp. *diacetilactis* (Matuszewski et al.) nom. rev., comb. nov. *Int. J. Syst. Bacteriol.* 32**,** 453-455.

Gasson, M. J., and Shearman, C. A. (2003). "Plasmid biology, conjugation, and transposition," in *Genetics of Lactic Acid Bacteria* eds. W.B.J. B. & W.P. J. (Dordrecht, The Netherlands: Kluwer Academic Publishers), 25-44.

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.

Godon, J.-J., Delorme, C., Ehrlich, S. D., and Renault, P. (1992). Divergence of genomic sequences between *Lactococcus lactis* subsp. *lactis* and *Lactococcus lactis* subsp. *cremoris*. *Appl. Environ. Microbiol.* 58**,** 4045-4047.

Górecki, R. K., Koryszewska-Bagińska, A., Gołębiewski, M., Żylińska, J., Grynberg, M., and Bardowski, J. K. (2011). Adaptative potential of the *Lactococcus lactis* IL594 strain encoded in its 7 plasmids. *PloS one.* 6**,** e22238. DOI: 10.1371/journal.pone.0022238.

Gruzza, M., Fons, M., Ouriet, M., Duval-Iflah, Y., and Ducluzeau, R. (1994). Study of gene transfer *in vitro* and in the digestive tract of gnotobiotic mice from *Lactococcus lactis* strains to various strains belonging to human intestinal flora. *Microb. Releases.* 2**,** 183-189.

Gruzza, M., Langella, P., Duval-Iflah, Y., and Ducluzeau, R. (1993). Gene transfer from engineered *Lactococcus lactis* strains to *Enterococcus faecalis* in the digestive tract of gnotobiotic mice. *Microb. Releases.* 2**,** 121-125.

Hols, P., Hancy, F., Fontaine, L., Grossiord, B., Prozzi, D., Leblond-Bourget, N., Decaris, B., Bolotin, A., Delorme, C., and Dusko Ehrlich, S. (2005). New insights in the molecular biology and physiology of *Streptococcus thermophilus* revealed by

comparative genomics. *FEMS Microbiol. Rev.* 29**,** 435-463. DOI: 10.1016/j.fmrre.2005.04.008.

Huang, D. C., Novel, M., and Novel, G. (1991). A transposon-like element on the lactose plasmid of *Lactococcus lactis* subsp. *lactis* Z270. *FEMS Microbiol. Lett.* 77**,** 101-106.

Hugenholtz, J., Dijkstra, M., and Veldkamp, H. (1987). Amino acid limited growth of starter cultures in milk. *FEMS Microbiol. Lett.* 45**,** 191-198.

ISO/ DIS 10932/IDF223, (2010). Milk and milk products - Determination of the minimal inhibitory concentration (MIC) of antibiotics applicable to bifidobacteria and non-enterococcal lactic acid bacteria (LAB). International Organisation for Standardisation, Geneva, Switzerland.

Iyer, R., Tomar, S., Uma Maheswari, T., and Singh, R. (2010). *Streptococcus thermophilus* strains: Multifunctional lactic acid bacteria. *Int. Dairy J.* 20**,** 133-141. DOI: 10.1016/j.idairyj.2009.10.005.

Jarvis, A. W., and Jarvis, B. D. (1981). Deoxyribonucleic acid homology among lactic streptococci. *Appl. Environ. Microbiol.* 41**,** 77-83.

Jensen, P. R., and Hammer, K. (1993). Minimal requirements for exponential growth of *Lactococcus lactis*. *Appl. Environ. Microbiol.* 59**,** 4363-4366.

Kandler, O., and Weiss, N. (1986). "Genus *Lactobacillus* Beijerinck 1901," in *Bergey's manual of systematic bacteriology,* eds. P.H.A. Sneath, N.S. Mair, M.E. Sharpe & J.G. Holt. (Baltimore, MD: The Williams and Wilkins Co.), 1209-1234.

Kato, H., Shiwa, Y., Oshima, K., Machii, M., Araya-Kojima, T., Zendo, T., Shimizu-Kadota, M., Hattori, M., Sonomoto, K., and 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. DOI: 10.1128/JB.00074-12.

Kelly, W., and Ward, L. (2002). Genotypic vs. phenotypic biodiversity in *Lactococcus lactis*. *Microbiol.* 148**,** 3332-3333.

Kelly, W. J., Altermann, E., Lambie, S. C., and Leahy, S. C. (2013). Interaction between the genomes of *Lactococcus lactis* and phages of the P335 species. *Front. Microbiol.* 4**,** 1-9. DOI: 10.3389/fmicb.2013.00257.

Kelly, W. J., Davey, G. P., and Ward, L. J. (1998a). Conjugative transfer of raffinose metabolism in *Lactococcus lactis*. *FEMS Microbiol. Lett.* 167**,** 145-149. DOI: 10.1111/j.1574-6968.1998.tb13220.x.

Kelly, W. J., Davey, G. P., and Ward, L. J. H. (1998b). Characterization of lactococci isolated from minimally processed fresh fruit and vegetables. *Int. J. Food Microbiol.* 45**,** 85-92.

Kelly, W. J., Ward, L. J. H., and Leahy, S. C. (2010). Chromosomal diversity in *Lactococcus lactis* and the origin of dairy starter cultures. *Genome Biol. Evol.* 2**,** 729-744. DOI: 10.1093/gbe/evq056.

Kempler, G., and Mckay, L. (1981). Biochemistry and genetics of citrate utilization in *Streptococcus lactis* ssp. *diacetylactis*. *J. Dairy Sci.* 64**,** 1527-1539.

Kimoto, H., Kurisaki, J., Tsuji, N., Ohmomo, S., and Okamoto, T. (1999). Lactococci as probiotic strains: adhesion to human enterocyte-like Caco-2 cells and tolerance to low pH and bile. *Lett. Appl. Microbiol.* 29**,** 313-316. DOI: 10.1046/j.1365-2672.1999.00627.x.

Kimoto, H., Nomura, M., Kobayashi, M., Okamoto, T., and Ohmomo, S. (2004). Identification and probiotic characteristics of *Lactococcus* strains from plant materials. *Japan agricultural research quarterly.* 38**,** 111-118.

Kleerebezem, M., and Hugenholtz, J. (2003). Metabolic pathway engineering in lactic acid bacteria. *Curr. Opin. Biotechnol.* 14**,** 232-237. DOI: 10.1016/S0958-1669(03)00033-8.

Kok, J. (1990). Genetics of the proteolytic system of lactic acid bacteria. *FEMS Microbiol. Lett.* 87**,** 15-41. DOI: 10.1111/j.1574-6968.1990.tb04877.x.

Kunji, E. R., Mierau, I., Hagting, A., Poolman, B., and Konings, W. N. (1996). The proteotytic systems of lactic acid bacteria. *Antonie van Leeuwenhoek.* 70**,** 187-221. DOI: 10.1007/978-94-011-1340-3_4.

Kurtz, S., Phillippy, A., Delcher, A. L., Smoot, M., Shumway, M., Antonescu, C., and Salzberg, S. L. (2004). Versatile and open software for comparing large genomes. *Genome Biol.* 5**,** R12.

Kütahya, O. E., Starrenburg, M. J., Rademaker, J. L., Klaassen, C. H., Van Hylckama Vlieg, J. E., Smid, E. J., and Kleerebezem, M. (2011). High-resolution AFLP typing of *Lactococcus lactis* strains enables identification of genetic markers for subspecies related phenotypes. *Appl. Environ. Microbiol.* 77**,** 5192-5198. DOI: 10.1128/AEM.00518-11.

Ladero, V., Calles-Enriquez, M., Fernández, M., and A Alvarez, M. (2010). Toxicological effects of dietary biogenic amines. *Curr. Nutr. Food Sci.* 6**,** 145-156. DOI: http://dx.doi.org/10.2174/157340110791233256.

Ladero, V., Rattray, F. P., Mayo, B., Martín, M. C., Fernández, M., and Alvarez, M. A. (2011). Sequencing and transcriptional analysis of the biosynthesis gene cluster of putrescine-producing *Lactococcus lactis*. *Appl. Environ. Microbiol.* 77**,** 6409-6418.

Lamarque, M., Aubel, D., Piard, J.-C., Gilbert, C., Juillard, V., and Atlan, D. (2011). The peptide transport system Opt is involved in both nutrition and environmental sensing during growth of *Lactococcus lactis* in milk. *Microbiol.* 157**,** 1612-1619.

Lambie, S. C., Altermann, E., Leahy, S. C., and Kelly, W. J. (2014). Draft genome sequence of *Lactococcus lactis* subsp. *cremoris* HPT, the first defined-strain dairy starter culture bacterium. *Genome Announc.* 2**,** e00107-00114. DOI: 10.1128/genomeA.00107-14.

Le, D. T. L., Tran, T.-L., Duviau, M.-P., Meyrand, M., Guérardel, Y., Castelain, M., Loubiere, P., Chapot-Chartier, M.-P., Dague, E., and Mercier-Bonin, M. (2013). Unraveling the role of surface mucus-binding protein and pili in muco-adhesion of *Lactococcus lactis*. *PloS one.* 8**,** e79850. DOI: 10.1371/journal.pone.0079850.

Licht, T. R., Christensen, B. B., Krogfelt, K. A., and Molin, S. (1999). Plasmid transfer in the animal intestine and other dynamic bacterial populations: the role of community structure and environment. *Microbiol.* 145**,** 2615-2622.

Linares, D. M., Kok, J., and Poolman, B. (2010). Genome sequences of *Lactococcus lactis* MG1363 (revised) and NZ9000 and comparative physiological studies. *J. Bacteriol.* 192**,** 5806-5812. DOI: 10.1128/JB.00533-10.

Liu, M., Bayjanov, J. R., Renckens, B., Nauta, A., and Siezen, R. J. (2010). The proteolytic system of lactic acid bacteria revisited: a genomic comparison. *BMC genomics.* 11**,** 36. DOI: 10.1186/1471-2164-11-36.

Lorenzo-Díaz, F., Fernández-López, C., Pilar Garcillán-Barcia, M., and Espinosa, M. (2014). Bringing them together: plasmid pMV158 rolling circle replication and conjugation under an evolutionary perspective. *Plasmid.* 74**,** 15-31. DOI: 10.1016/j.plasmid.2014.05.004.

Machielsen, R., Siezen, R. J., Van Hijum, S. A., and Van Hylckama Vlieg, J. E. (2011). Molecular description and industrial potential of Tn6098 conjugative transfer conferring alpha-galactoside metabolism in *Lactococcus lactis*. *Appl. Environ. Microbiol.* 77**,** 555-563. DOI: 10.1128/AEM.02283-10.

Makarova, K., Slesarev, A., Wolf, Y., Sorokin, A., Mirkin, B., Koonin, E., Pavlov, A., Pavlova, N., Karamychev, V., and Polouchine, N. (2006). Comparative genomics of the lactic acid bacteria. *PNAS.* 103**,** 15611-15616. DOI: 10.1073/pnas.0607117103.

Mallet, A., Guéguen, M., Kauffmann, F., Chesneau, C., Sesboué, A., and Desmasures, N. (2012). Quantitative and qualitative microbial analysis of raw milk reveals substantial diversity influenced by herd management practices. *Int. Dairy J.* 27**,** 13-21. DOI: DOI: 10.1016/j.idairyj.2012.07.009.

Mangin, I., Corroler, D., Reinhardt, A., and Gueguen, M. (1999). Genetic diversity among dairy lactococcal strains investigated by polymerase chain reaction with three arbitrary primers. *J. Appl. Microbiol.* 86**,** 514-520. DOI: 10.1046/j.1365-2672.1999.00699.x.

Marshall, V. M. (1991). Inoculated ecosystem in a milk environment. *J. Appl. Bacteriol.* 73**,** 127S-135S. DOI: 10.1111/j.1365-2672.1992.tb03632.x.

Mckay, L.L., and Baldwin, K. (1974). Simultaneous loss of proteinase-and lactose-utilizing enzyme activities in *Streptococcus lactis* and reversal of loss by transduction. *Appl. Microbiol.* 28**,** 342-346.

Mckay, L.L., Baldwin, K., and Zottola, E. (1972). Loss of lactose metabolism in lactic streptococci. *Appl. Microbiol.* 23**,** 1090-1096.

McNulty, N. P., Yatsunenko, T., Hsiao, A., Faith, J. J., Muegge, B. D., Goodman, A. L., Henrissat, B., Oozeer, R., Cools-Portier, S., and Gobert, G. (2011). The impact of a consortium of fermented milk strains on the gut microbiome of gnotobiotic mice and monozygotic twins. *Sci. Transl. Med.* 3**,** 106ra106. DOI: 10.1126/scitranslmed.3002701.

Meneses, C. H., Rouws, L. F., Simões-Araújo, J. L., Vidal, M. S., and Baldani, J. I. (2011). Exopolysaccharide production is required for biofilm formation and plant colonization by the nitrogen-fixing endophyte *Gluconacetobacter diazotrophicus*. *Mol. Plant-Microbe In.* 24**,** 1448-1458.

Meyrand, M., Guillot, A., Goin, M., Furlan, S., Armalyte, J., Kulakauskas, S., Cortes-Perez, N. G., Thomas, G., Chat, S., and Pechoux, C. (2013). Surface proteome analysis of a natural isolate of *Lactococcus lactis* reveals the presence of pili able to bind human intestinal epithelial cells. *Mol. Cell. Proteomics.* 12**,** 3935-3947. DOI: 10.1074/mcp.M113.029066.

Michel, C., Pelletier, C., Boussaha, M., Douet, D.-G., Lautraite, A., and Tailliez, P. (2007). Diversity of lactic acid bacteria associated with fish and the fish farm environment, established by amplified rRNA gene restriction analysis. *Appl. Environ. Microbiol.* 73**,** 2947-2955. DOI: 10.1128/AEM.01852-06.

Mierau, I., Kunji, E., Leenhouts, K. J., Hellendoorn, M. A., Haandrikman, A. J., Poolman, B., Konings, W. N., Venema, G., and Kok, J. (1996). Multiple-peptidase mutants of *Lactococcus lactis* are severely impaired in their ability to grow in milk. *J. Bacteriol.* 178**,** 2794-2803.

Mills, S., McAuliffe, O. E., Coffey, A., Fitzgerald, G. F., and Ross, R. P. (2006). Plasmids of lactococci–genetic accessories or genetic necessities? *FEMS Microbiol. Rev.* 30**,** 243-273. DOI: 10.1111/j.1574-6976.2005.00011.x.

Mills, S., O'Suillvan, O., Hill, C., Fitzgerald, G., and Ross, R. P. (2010). The changing face of dairy starter culture research: from genomics to economics. *Int. J. Dairy Technol.* 63**,** 149-170. DOI: 10.1111/j.1471-0307.2010.00563.x.

Monsallier, F., Verdier-Metz, I., Agabriel, C., Martin, B., and Montel, M.-C. (2012). Variability of microbial teat skin flora in relation to farming practices and individual dairy cow characteristics. *Dairy Sci. Technol.* 92**,** 265-278. DOI: 10.1007/s13594-012-0064-7.

Morita, H., Toh, H., Oshima, K., Yoshizaki, M., Kawanishi, M., Nakaya, K., Suzuki, T., Miyauchi, E., Ishii, Y., and Tanabe, S. (2011). Complete genome sequence and comparative analysis of the fish pathogen *Lactococcus garvieae*. *PLoS One.* 6**,** e23184. DOI: 10.1371/journal.pone.0023184.

Nardi, M., Renault, P., and Monnet, V. (1997). Duplication of the *pepF* gene and shuffling of DNA fragments on the lactose plasmid of *Lactococcus lactis*. *J. Bacteriol.* 179**,** 4164-4171.

Nawaz, M., Wang, J., Zhou, A., Ma, C., Wu, X., Moore, J. E., Millar, B. C., and Xu, J. (2011). Characterization and transfer of antibiotic resistance in lactic acid bacteria from fermented food products. *Curr. Microbiol.* 62**,** 1081-1089. DOI: 10.1007/s00284-010-9856-2.

Nissen-Meyer, J., Lillehaug, D., and Nes, I. F. (1992). The plasmid-encoded lactococcal envelope-associated proteinase is encoded by a chromosomal gene in *Lactococcus lactis* subsp. *cremoris* BC101. *Appl. Environ. Microbiol.* 58**,** 750-753.

Nomura, M., Kimoto, H., Someya, Y., and Suzuki, I. (1999). Novel characteristic for distinguishing *Lactococcus lactis* subsp. *lactis* from subsp. *cremoris*. *Int. J. Syst. Bacteriol.* 49**,** 163-166. DOI: 10.1099/00207713-49-1-163.

Nomura, M., Kobayashi, M., Narita, T., Kimoto-Nira, H., and Okamoto, T. (2006). Phenotypic and molecular characterization of *Lactococcus lactis* from milk and plants. *J. Appl. Microbiol.* 101**,** 396-405.

Nomura, M., Kobayashi, M., Ohmomo, S., and Okamoto, T. (2000). Inactivation of the glutamate decarboxylase gene in *Lactococcus lactis* subsp. *cremoris*. *Appl. Environ. Microbiol.* 66**,** 2235-2237. DOI: 10.1128/AEM.66.5.2235-2237.2000.

Nomura, M., Kobayashi, M., and Okamoto, T. (2002). Rapid PCR-based method which can determine both phenotype and genotype of *Lactococcus lactis* subspecies. *Appl. Environ. Microbiol.* 68**,** 2209-2213. DOI: 10.1128/AEM.68.5.2209-2213.2002.

Obis, D., Guillot, A., and Mistou, M. Y. (2001). Tolerance to high osmolality of *Lactococcus lactis* subsp. *lactis* and *cremoris* is related to the activity of a betaine transport system. *FEMS Microbiol. Lett.* 202**,** 39-44. DOI: 10.1111/j.1574-6968.2001.tb10777.x.

Orla-Jensen, S. (1919). *The lactic acid bacteria.* Copenhagen, Denmark: Høst and son.

Parapouli, M., Delbès-Paus, C., Kakouri, A., Koukkou, A.-I., Montel, M.-C., and Samelis, J. (2013). Characterization of a wild, novel nisin A-producing *Lactococcus* strain with an *L. lactis* subsp. *cremoris* genotype and an *L. lactis* subsp. *lactis* phenotype, isolated from Greek raw milk. *Appl. Environ. Microbiol.* 79**,** 3476-3484. DOI: 10.1128/AEM.00436-13.

Passerini, D., Beltramo, C., Coddeville, M., Quentin, Y., Ritzenthaler, P., Daveran-Mingot, M.-L., and Le Bourgeois, P. (2010). Genes but not genomes reveal bacterial domestication of *Lactococcus lactis*. *PLoS One.* 5**,** e15306. DOI: 10.1371/journal.pone.0015306.

Passerini, D., Coddeville, M., Le Bourgeois, P., Loubière, P., Ritzenthaler, P., Fontagné-Faucher, C., Daveran-Mingot, M.-L., and Cocaign-Bousquet, M. (2013a). The carbohydrate metabolism signature of *Lactococcus lactis* strain A12 reveals its sourdough ecosystem origin. *Appl. Environ. Microbiol.* 79**,** 5844-5852. DOI: 10.1128/AEM.01560-13.

Passerini, D., Laroute, V., Coddeville, M., Le Bourgeois, P., Loubiere, P., Riztenthaler, P., Cocaign-Bousquet, M., and Daveran-Mingot, M.-L. (2013b). New insights into *Lactococcus lactis* diacetyl-and acetoin-producing strains isolated from diverse origins. *Int. J. Food Microbiol.* 160**,** 329-336. DOI: 10.1016/j.ijfoodmicro.2012.10.023.

Pastink, M. I., Teusink, B., Hols, P., Visser, S., De Vos, W. M., and Hugenholtz, J. (2009). Genome-scale model of *Streptococcus thermophilus* LMG18311 for metabolic comparison of lactic acid bacteria. *Appl. Environ. Microbiol.* 75**,** 3627-3633. DOI: 10.1128/AEM.00138-09.

Pedersen, M. B., Iversen, S. L., Sørensen, K. I., and Johansen, E. (2005). The long and winding road from the research laboratory to industrial applications of lactic acid bacteria. *FEMS Microbiol. Rev.* 29**,** 611-624. DOI: 10.1016/j.fmrre.2005.04.001.

Pederson, J. A., Steele, J. L., Christensen, J. E., and Dudley, E. G. (1999). Peptidases and amino acid catabolism in lactic acid bacteria. *Antonie van Leeuwenhoek.* 76**,** 217-246. DOI: 10.1023/A:1002001919720.

Pérez, T., Balcázar, J. L., Peix, A., Valverde, A., Velázquez, E., De Blas, I., and Ruiz-Zarzuela, I. (2011). *Lactococcus lactis* subsp. *tructae* subsp. *nov*. isolated from the intestinal mucus of brown trout (*Salmo trutta*) and rainbow trout (*Oncorhynchus mykiss*). *Int. J. Syst. Evol. Microbiol.* 61**,** 1894-1898. DOI: 10.1099/ijs.0.023945-0.

Polzin, K. M., and Mckay, L. L. (1991). Identification, DNA sequence, and distribution of IS*981*, a new, high-copy-number insertion sequence in lactococci. *Appl. Environ. Microbiol.* 57**,** 734-743.

Price, C. E., Zeyniyev, A., Kuipers, O. P., and Kok, J. (2012). From meadows to milk to mucosa–adaptation of *Streptococcus* and *Lactococcus* species to their nutritional environments. *FEMS Microbiol. Rev.* 36**,** 949-971. DOI: 10.1111/j.1574-6976.2011.00323.x.

Pu, Z., Dobos, M., Limsowtin, G., and Powell, I. (2002). Integrated polymerase chain reaction-based procedures for the detection and identification of species and subspecies of the Gram-positive bacterial genus *Lactococcus*. *J. Appl. Microbiol.* 93**,** 353-361. DOI: 10.1046/j.1365-2672.2002.01688.x.

Quigley, L., O'Sullivan, O., Stanton, C., Beresford, T. P., Ross, R. P., Fitzgerald, G. F., and Cotter, P. D. (2013). The complex microbiota of raw milk. *FEMS Microbiol. Rev.* 37**,** 664-698. DOI: 10.1111/1574-6976.12030.

Rademaker, J. L. W., Herbet, H., Starrenburg, M. J. C., Naser, S. M., Gevers, D., Kelly, W. J., Hugenholtz, J., Swings, J., and Van Hylckama Vlieg, J. E. T. (2007). Diversity analysis of dairy and nondairy *Lactococcus lactis* isolates, using a novel

multilocus sequence analysis scheme and (GTG) 5-PCR fingerprinting. *Appl. Environ. Microbiol.* 73**,** 7128-7137.

Richter, M., and Rosselló-Móra, R. (2009). Shifting the genomic gold standard for the prokaryotic species definition. *PNAS.* 106**,** 19126-19131.

Romero, D. A., and Klaenhammer, T. R. (1990). Abortive phage infection and restriction/modification activities directed by pTR2030 determinants are enhanced by recombination with conjugal elements in lactococci. *J. Gen. Microbiol.* 136**,** 1817-1824. DOI: 10.1099/00221287-136-9-1817.

Romero, D. A., and Klaenhammer, T. R. (1993). Transposable elements in lactococci: a review. *J. Dairy Sci.* 76**,** 1-19. DOI: 10.3168/jds.S0022-0302(93)77318-X.

Ross, R. P., Morgan, S., and Hill, C. (2002). Preservation and fermentation: past, present and future. *Int. J. Food Microbiol.* 79**,** 3-16. DOI: 10.1016/S0168-1605(02)00174-5.

Ross, R. P., Stanton, C., Hill, C., F Fitzgerald, G., and Coffey, A. (2000). Novel cultures for cheese improvement. *Trends Food Sci. Technol.* 11**,** 96-104. DOI: 10.1016/S0924-2244(00)00057-1.

Sakala, R., Hayashidani, H., Kato, Y., Kaneuchi, C., and Ogawa, M. (2002). Isolation and characterization of *Lactococcus piscium* strains from vacuum-packaged refrigerated beef. *J. Appl. Microbiol.* 92**,** 173-179. DOI: 10.1046/j.1365-2672.2002.01513.x.

Salama, M., Sandine, W., and Giovannoni, S. (1991). Development and application of oligonucleotide probes for identification of *Lactococcus lactis* subsp. *cremoris*. *Appl. Environ. Microbiol.* 57**,** 1313-1318.

Salama, M. S., Musafija-Jeknic, T., Sandine, W. E., and Giovannoni, S. J. (1995). An ecological study of lactic acid bacteria: isolation of new strains of *Lactococcus*

including *Lactococcus lactis* subspecies *cremoris*. *J. Dairy Sci.* 78**,** 1004-1017. DOI: 10.3168/jds.S0022-0302(95)76716-9.

Salama, M. S., Sandine, W. E., and Giovannoni, S. J. (1993). Isolation of *Lactococcus lactis* subsp. *cremoris* from nature by colony hybridization with rRNA probes. *Appl. Environ. Microbiol.* 59**,** 3941-3945.

Sanders, J. W., Leenhouts, K., Burghoorn, J., Brands, J. R., Venema, G., and Kok, J. (1998). A chloride-inducible acid resistance mechanism in *Lactococcus lactis* and its regulation. *Mol. Microbiol.* 27**,** 299-310.

Schleifer, K., Kraus, J., Dvorak, C., Kilpper-Bälz, R., Collins, M., and Fischer, W. (1985). Transfer of *Streptococcus lactis* and related streptococci to the genus *Lactococcus* gen. nov. *Syst. Appl. Microbiol.* 6**,** 183-195.

Schneider, D., and Lenski, R. E. (2004). Dynamics of insertion sequence elements during experimental evolution of bacteria. *Res. Microbiol.* 155**,** 319-327. DOI: 10.1016/j.resmic.2003.12.008.

Seiboth, B., and Metz, B. (2011). Fungal arabinan and L-arabinose metabolism. *Appl. Microbiol. Biotechnol.* 89**,** 1665-1673. DOI: 10.1007/s00253-010-3071-8.

Serna Cock, L., and Rodríguez De Stouvenel, A. (2006). Lactic acid production by a strain of *Lactococcus lactis* subsp. *lactis* isolated from sugar cane plants. *Electron. J. Biotechnol.* 9**,** 0-0. DOI: 10.2225/vol9-issue1-fulltext-10.

Settanni, L., and Corsetti, A. (2008). Application of bacteriocins in vegetable food biopreservation. *Int. J. Food Microbiol.* 121**,** 123-138. DOI: 10.1016/j.ijfoodmicro.2007.09.001.

Siezen, R. J., Bayjanov, J., Renckens, B., Wels, M., Van Hijum, S. A., Molenaar, D., and Van Hylckama Vlieg, J. E. (2010). Complete genome sequence of *Lactococcus lactis* subsp. *lactis* KF147, a plant-associated lactic acid bacterium. *J. Bacteriol.* 192**,** 2649-2650. DOI: 10.1128/JB.00276-10.

Siezen, R. J., Renckens, B., Van Swam, I., Peters, S., Van Kranenburg, R., Kleerebezem, M., and De Vos, W. M. (2005). Complete sequences of four plasmids of *Lactococcus lactis* subsp. *cremoris* SK11 reveal extensive adaptation to the dairy environment. *Appl. Environ. Microbiol.* 71**,** 8371-8382. DOI: 10.1128/AEM.71.12.8371-8382.2005.

Siezen, R. J., Starrenburg, M. J., Boekhorst, J., Renckens, B., Molenaar, D., and Van Hylckama Vlieg, J. E. (2008). Genome-scale genotype-phenotype matching of two *Lactococcus lactis* isolates from plants identifies mechanisms of adaptation to the plant niche. *Appl. Environ. Microbiol.* 74**,** 424-436.

Smit, G., Smit, B. A., and Engels, W. J. (2005). Flavour formation by lactic acid bacteria and biochemical flavour profiling of cheese products. *FEMS Microbiol. Rev.* 29**,** 591-610. DOI: 10.1016/j.fmrre.2005.04.002.

Smit, G., Van Hylckama Vlieg, J. E., Smit, B. A., Ayad, E. H., and Engels, W. J. (2002). Fermentative formation of flavour compounds by lactic acid bacteria. *Aust. J. Dairy Technol.* 57**,** 61-68.

Smokvina, T., Wels, M., Polka, J., Chervaux, C., Brisse, S., Boekhorst, J., Van Hylckama Vlieg, J. E., and Siezen, R. J. (2013). *Lactobacillus paracasei* comparative genomics: towards species pan-genome definition and exploitation of diversity. *PloS one.* 8**,** e68731. DOI: 10.1371/journal.pone.0068731.

Stackebrandt, E., Frederiksen, W., Garrity, G. M., Grimont, P. A., Kämpfer, P., Maiden, M. C., Nesme, X., Rosselló-Mora, R., Swings, J., and Trüper, H. G. (2002). Report of the *ad hoc* committee for the re-evaluation of the species definition in bacteriology. *Int. J. Syst. Evol. Microbiol.* 52**,** 1043-1047.

Steidler, L., and Rottiers, P. (2006). Therapeutic drug delivery by genetically modified *Lactococcus lactis*. *Ann. N. Y. Acad. Sci.* 1072**,** 176-186. DOI: 10.1196/annals.1326.031.

Stiles, M. E., and Holzapfel, W. H. (1997). Lactic acid bacteria of foods and their current taxonomy. *Int. J. Food Microbiol.* 36**,** 1-29.

Taïbi, A., Dabour, N., Lamoureux, M., Roy, D., and Lapointe, G. (2010). Evaluation of the genetic polymorphism among *Lactococcus lactis* subsp. *cremoris* strains using comparative genomic hybridization and multilocus sequence analysis. *Int. J. Food Microbiol.* 144**,** 20-28. DOI: 10.1016/j.ijfoodmicro.2010.02.020.

Tailliez, P., Tremblay, J., Ehrlich, S., and Chopin, A. (1998). Molecular diversity and relationship within *Lactococcus lactis*, as revealed by randomly amplified polymorphic DNA (RAPD). *Syst. Appl. Microbiol.* 21**,** 530-538. DOI: 10.1016/S0723-2020(98)80065-9.

Tan-a-Ram, P., Cardoso, T., Daveran-Mingot, M.-L., Kanchanatawee, S., Loubière, P., Girbal, L., and Cocaign-Bousquet, M. (2011). Assessment of the diversity of dairy *Lactococcus lactis* subsp. *lactis* isolates by an integrated approach combining phenotypic, genomic, and transcriptomic analyses. *Appl. Environ. Microbiol.* 77**,** 739-748. DOI: 10.1128/AEM.01657-10.

Tanigawa, K., Kawabata, H., and Watanabe, K. (2010). Identification and typing of *Lactococcus lactis* by matrix-assisted laser desorption ionization-time of flight mass spectrometry. *Appl. Environ. Microbiol.* 76**,** 4055-4062. DOI: 10.1128/AEM.02698-09.

Tanous, C., Kieronczyk, A., Helinck, S., Chambellon, E., and Yvon, M. (2002). Glutamate dehydrogenase activity: a major criterion for the selection of flavour-producing lactic acid bacteria strains. *Antonie van Leeuwenhoek.* 82**,** 271-278. DOI: 10.1023/a:1020616506703.

Teeling, H., Meyerdierks, A., Bauer, M., Amann, R., and Glöckner, F. O. (2004). Application of tetranucleotide frequencies for the assignment of genomic fragments. *Environ. Microbiol.* 6**,** 938-947.

Teixeira, L. M., Merquior, V. L. C., Vianni, M. D. C. E., Carvalho, M. D. G. S., El Fracalanzza, S., Steigerwalt, A. G., Brenner, D. J., and Facklam, R. R. (1996). Phenotypic and genotypic characterization of atypical *Lactococcus garvieae* strains isolated from water buffalos with subclinical mastitis and confirmation of *L. garvieae* as a senior subjective synonym of *Enterococcus seriolicida*. *Int. J. Syst. Bacteriol.* 46**,** 664-668.

Tomás, A. F., Karakashev, D., and Angelidaki, I. (2013). *Thermoanaerobacter pentosaceus* sp. nov., an anaerobic, extremely thermophilic, high ethanol-yielding bacterium isolated from household waste. *Int. J. Syst. Evol. Micro.* 63**,** 2396-2404.

Trip, H., Mulder, N. L., and Lolkema, J. S. (2013). Cloning, expression, and functional characterization of secondary amino acid transporters of *Lactococcus lactis*. *J. Bacteriol.* 195**,** 340-350. DOI: 10.1128/JB.01948-12.

Tynkkynen, S., Buist, G., Kunji, E., Kok, J., Poolman, B., Venema, G., and Haandrikman, A. (1993). Genetic and biochemical characterization of the oligopeptide transport system of *Lactococcus lactis*. *J. Bacteriol.* 175**,** 7523-7532.

Urbach, E., Daniels, B., Salama, M. S., Sandine, W. E., and Giovannoni, S. J. (1997). The ldh phylogeny for environmental isolates of *Lactococcus lactis* is consistent with rRNA genotypes but not with phenotypes. *Appl. Environ. Microbiol.* 63**,** 694-702.

Vacheyrou, M., Normand, A.-C., Guyot, P., Cassagne, C., Piarroux, R., and Bouton, Y. (2011). Cultivable microbial communities in raw cow milk and potential transfers from stables of sixteen French farms. *Int. J. Food Microbiol.* 146**,** 253-262. DOI: 10.1016/j.ijfoodmicro.2011.02.033.

Van De Guchte, M., Penaud, S., Grimaldi, C., Barbe, V., Bryson, K., Nicolas, P., Robert, C., Oztas, S., Mangenot, S., and Couloux, A. (2006). The complete genome sequence of *Lactobacillus bulgaricus* reveals extensive and ongoing reductive evolution. *PNAS.* 103**,** 9274-9279. DOI: 10.1073/pnas.0603024103.

Van De Guchte, M., Serror, P., Chervaux, C., Smokvina, T., Ehrlich, S. D., and Maguin, E. (2002). Stress responses in lactic acid bacteria. *Antonie Van Leeuwenhoek.* 82**,** 187-216.

Van Hylckama Vlieg, J. E., Rademaker, J. L., Bachmann, H., Molenaar, D., Kelly, W. J., and Siezen, R. J. (2006). Natural diversity and adaptive responses of *Lactococcus lactis*. *Curr. Opin. Biotechnol.* 17**,** 183-190. DOI: 10.1016/j.copbio.2006.02.007.

Van Kranenburg, R., Kleerebezem, M., Van Hylckama Vlieg, J., Ursing, B. M., Boekhorst, J., Smit, B. A., Ayad, E. H., Smit, G., and Siezen, R. J. (2002). Flavour formation from amino acids by lactic acid bacteria: predictions from genome sequence analysis. *Int. Dairy J.* 12**,** 111-121. DOI: 10.1016/S0958-6946(01)00132-7.

Verraes, C., Van Boxstael, S., Van Meervenne, E., Van Coillie, E., Butaye, P., Catry, B., De Schaetzen, M.-A., Van Huffel, X., Imberechts, H., and Dierick, K. (2013). Antimicrobial resistance in the food chain: a review. *Int. J. Environ. Res. Public Health.* 10**,** 2643-2669. DOI: 10.3390/ijerph10072643.

Vinh, D. C., Nichol, K. A., Rand, F., and Embil, J. M. (2006). Native-valve bacterial endocarditis caused by *Lactococcus garvieae*. *Diagn. Microbiol. Infect. Dis.* 56**,** 91-94. DOI: 10.1016/j.diagmicrobio.2006.02.010.

Wang, C., Shie, H. S., Chen, S. C., Huang, J. P., Hsieh, I. C., Wen, M. S., Lin, F. C., and Wu, D. (2007). *Lactococcus garvieae* infections in humans: possible association with aquaculture outbreaks. *Int. J. Clin. Pract.* 61**,** 68-73. DOI: 10.1111/j.1742-1241.2006.00855.x.

Weerkamp, A., Klijn, N., Neeter, R., and Smit, G. (1996). Properties of mesophilic lactic acid bacteria from raw milk and naturally fermented raw milk products. *Nederlands melk en Zuiveltijdschrift.* 50**,** 319-332.

Wegmann, U., O'Connell-Motherway, M., Zomer, A., Buist, G., Shearman, C., Canchaya, C., Ventura, M., Goesmann, A., Gasson, M. J., and Kuipers, O. P. (2007).

Complete genome sequence of the prototype lactic acid bacterium *Lactococcus lactis* subsp. *cremoris* MG1363. *J. Bacteriol.* 189**,** 3256-3270. DOI: 10.1128/JB.01768-06.

Wegmann, U., Overweg, K., Jeanson, S., Gasson, M., and Shearman, C. (2012). Molecular characterization and structural instability of the industrially important composite metabolic plasmid pLP712. *Microbiol.* 158**,** 2936-2945. DOI: 10.1099/mic.0.062554-0.

Wells, J. M., and Mercenier, A. (2008). Mucosal delivery of therapeutic and prophylactic molecules using lactic acid bacteria. *Nat. Rev. Microbiol.* 6**,** 349-362. DOI: 10.1038/nrmicro1840.

WHO, (2014). Antimicrobial Resistance: Global Report on Surveillance.

Wouters, J., Ayad, E. H., Hugenholtz, J., and Smit, G. (2002). Microbes from raw milk for fermented dairy products. *Int. Dairy J.* 12**,** 91-109. DOI: 10.1016/S0958-6946(01)00151-0.

Xu, H., Sun, Z., Liu, W., Yu, J., Song, Y., Lv, Q., Zhang, J., Shao, Y., Menghe, B., and 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. DOI: 10.3168/jds.2013-7738.

Yang, X., Wang, Y., and Huo, G. (2013). Complete Genome Sequence of *Lactococcus lactis* subsp. *lactis* KLDS 4.0325. *Genome Announc.* 1**,** e00962-00913. DOI: 10.1128/genomeA.00962-13.

Yarlagadda, A., Wilkinson, M., O'Sullivan, M., and Kilcawley, K. (2014). Utilisation of microfluidisation to enhance enzymatic and metabolic potential of lactococcal strains as adjuncts in Gouda type cheese. *Int. Dairy J.* 38**,** 1-10. DOI: 10.1016/j.idairyj.2014.01.007.

Yiu, K.-H., Siu, C.-W., To, K. K.-W., Jim, M.-H., Lee, K. L.-F., Lau, C.-P., and Tse, H.-F. (2007). A rare cause of infective endocarditis; *Lactococcus garvieae*. *Int. J. Cardio.* 114**,** 286-287. DOI: 10.1016/j.ijcard.2005.11.092.

Yu, W., Gillies, K., Kondo, J. K., Broadbent, J. R., and Mckay, L. L. (1996). Loss of plasmid-mediated oligopeptide transport system in lactococci: another reason for slow milk coagulation. *Plasmid.* 35**,** 145-155.

Yvon, M., and Rijnen, L. (2001). Cheese flavour formation by amino acid catabolism. *Int. Dairy J.* 11**,** 185-201. DOI: 10.1016/S0958-6946(01)00049-8.

Table 1. List of available *Lactococcus lactis* genomes. Data was collected from Genbank (http://www.ncbi.nlm.nih.gov/genome/?term=Lactococcus+lactis; accessed 2nd July 2014).

Strain	Year	Genotype	Origin	Size (mb)	Proteins	Plasmid	G/ C content (%)	Reference
L. lactis IL1403	2001	*lactis*	dairy	2.37	2, 277	0	35.3	(Bolotin *et al.*, 2001)
L lactis KF147	2010	*lactis*	mung bean sprouts	2.60	2, 578	1	34.9	(Siezen *et al.*, 2010)
L. lactis KF282	2010	*lactis*	mustard and cress	-	-	-	-	(Siezen, Bayjanov *et al.*2010)
L. lactis CV56	2011	*lactis*	vaginal flora	2.40	2, 301	5	35.2	(Gao *et al.*, 2011)
L. lactis CNCMI-1631	2011	*lactis*	fermented milk	2.51	2, 579	-	34.9	(McNulty *et al.*, 2011)
L. lactis IO-1	2012	*lactis*	drain water	2.42	2, 224	-	35.1	(Kato *et al.*, 2012)
L. lactis YF11	2013	*lactis*	fermented corn	2.53	2, 531	0	34.8	(Du *et al.*, 2013)
L. lactis NCDO2118	2013	*lactis*	frozen peas	2.81	n/a	-	35.0	(unpublished)
L. lactis TIFN2	2013	*lactis*	dairy- cheese	2.51	2, 521	-	35.1	(Erkus *et al.*, 2013)
L. lactis TIFN4	2013	*lactis*	dairy- cheese	2.55	2, 598	-	35.0	(Erkus *et al.*, 2013)
L. lactis KLDS 4.0325	2013	*lactis*	fermented horse milk	2.59	2, 587	3	35.4	(Yang *et al.*, 2013)
L. lactis Dephy 1	2013	*lactis*	undefined	2.60	2, 686	-	35.1	(unpublished)
L. lactis A12	2013	*lactis*	sourdough bread	2.70	2, 725	-	35.3	(Passerini *et al.*, 2013a)
L. lactis LD61	2014	*lactis*	dairy-cheese	2.60	2, 601	6	36.4	(Falentin *et al.*, 2014)
L. lactis SK11	2006	*cremoris*	dairy-cheese	2.44	2, 381	4	35.9	(Makarova *et al.*, 2006)
L. lactis MG1363	2007	*cremoris*	dairy	2.53	2, 434	0	35.7	(Wegmann *et al.*, 2007)
L. lactis NZ9000	2010	*cremoris*	dairy	2.53	2, 510	0	35.7	(Linares *et al.*, 2010)
L. lactis A76	2012	*cremoris*	dairy- cheese	2.45	2, 643	4	35.9	(Bolotin *et al.*, 2012)
L. lactis KW2	2013	*cremoris*	fermented corn	2.43	2,268	0	35.7	(Kelly *et al.*, 2013)
L. lactis TIFN1	2013	*cremoris*	dairy- cheese	2.68	2, 754	-	35.5	(Erkus *et al.*, 2013)
L. lactis TIFN3	2013	*cremoris*	dairy- cheese	2.73	2, 891	-	35.5	(Erkus *et al.*, 2013)
L. lactis TIFN5	2013	*cremoris*	dairy- cheese	2.54	2, 232	-	35.5	(Erkus *et al.*, 2013)
L. lactis TIFN7	2013	*cremoris*	dairy- cheese	2.63	2, 505	-	35.6	(Erkus *et al.*, 2013)
L. lactis UC509.9	2013	*cremoris*	dairy	2.25	2, 208	8	35.9	(Ainsworth *et al.*, 2013)
L. lactis HP^T	2014	*cremoris*	dairy	2.27	2, 374	7	36.7	(Lambie *et al.*, 2014)

Table 2. Methods available for subspecies identification of *cremoris* and *lactis* genotypes using different techniques.

Gene (s) encoding	Technique	Application	Identification	Reference
N-acetylmuramidase	PCR[a]	Specific primer sets were designed for each subspecies and used in reaction mixtures.	Formation of 1,131 bp amplicon for both subspecies plus an additional 700 bp product for subspecies *lactis.*	(Garde *et al.*, 1999)
16S rRNA	PCR	Reaction mixtures contained a species specific reverse primer and a subspecies specific forward primer.	Formation of an amplicon using either a subspecies *lactis* or *cremoris* forward primer and common reverse primer.	(Pu *et al.*, 2002)
histidine biosynthesis	PCR	Specific primer sets were designed for each subspecies and used in reaction mixtures.	Subspecies *cremoris* formed a 1149 bp amplicon while subspecies *lactis* formed a 934 bp amplicon.	(Beimfohr *et al.*, 1997)
glutamate decarboxylase	PCR-RFLP[b]	Formation of a PCR amplicon followed by endonuclease digestion with *Ase*I.	Digested PCR amplicons formed 190 and 410 bp fragments for subspecies *lactis*, and 190 and 370 bp fragments or 560 bp fragment for subspecies *cremoris*.	(Nomura *et al.*, 2002)
n/a	RAPD[c]-PCR	Generation of PCR amplicons forming distinct patterns using primers composed of random sequences.	Subspecies were identified by computer based analysis of banding patterns from RAPD fingerprints.	(Mangin *et al.*, 1999)
16S rRNA	northern blot hybridisation	*Cremoris* specific oligonucleotide probes were tested for binding to RNAs and whole cell lysates.	Strains that hybridise to the probe were designated subspecies *cremoris*.	(Salama *et al.*, 1991)
branched chain amino acid biosynthesis	southern blot hybridisation	Bacterial DNA was digested with *Eco*RI and hybridised with a 18.5 kb fragment.	Following hybridisation, the observation of fragment patterns common to those found in either subspecies *lactis* or	(Godon *et al.*, 1992)

			subspecies *cremoris* are used for identification.	
multi-locus sequence typing (MLST)	PCR	Seven housekeeping genes were amplified by PCR and sequenced in forward and reverse orientation.	Sequence data is used in downstream phylogenetic analysis which shows subspecies lactis and subspecies cremoris genotype strains on separate branches.	(Rademaker *et al.*, 2007)

[a] polymerase chain reaction
[b] polymerase chain reaction-restriction fragment length polymorphism
[c] randomly amplified polymorphic DNA

Table 3. Determination of average nucleotide identity (ANI) and correlation of tetranucleotide frequencies[a] for chromosomal DNA of *L. lactis* type-strains IL1403 and MG1363 and other selected *Lactococcus lactis* strains.

Strain	Subspecies	ANIb[b] (%)	ANIm[c] (%)	TETRA
Reference IL1403				
MG1363	*cremoris*	87.0	88.6	0.994
A76	*cremoris*	86.88	88.6	0.991
SK11	*cremoris*	86.68	88.28	0.991
UC509.9	*cremoris*	86.88	88.45	0.991
NZ9000	*cremoris*	87.0	88.6	0.994
KW2	*cremoris*	86.16	87.71	0.994
KF147	*lactis*	98.58	98.57	0.999
CV56	*lactis*	99.52	99.47	0.999
NCDO 2118	*lactis*	98.42	98.53	0.999
KLDS 4.0325	*lactis*	99.29	99.4	0.998
IO-1	*lactis*	97.84	98.18	0.999
Reference MG1363				
IL1403	*lactis*	86.86	88.6	0.994
A76	*cremoris*	97.79	98.01	0.998
SK11	*cremoris*	97.79	98.07	0.998
UC509.9	*cremoris*	97.67	98	0.997
NZ9000	*cremoris*	100	99.79	1.0
KW2	*cremoris*	98.09	98.33	0.998
KF147	*lactis*	86.69	88.26	0.994
CV56	*lactis*	86.81	88.46	0.995
NCDO 2118	*lactis*	86.57	88.25	0.994
KLDS 4.0325	*lactis*	86.77	88.62	0.995
IO-1	*lactis*	86.71	88.28	0.994

[a] ANI and correlations of tetranucleotide frequencies were calculated using JSpecies (Richter and Rosselló-Móra, 2009) with default parameter settings.

[b] Results generated using BLAST algorithm.

[c] Results generated using MUMmer algorithm.

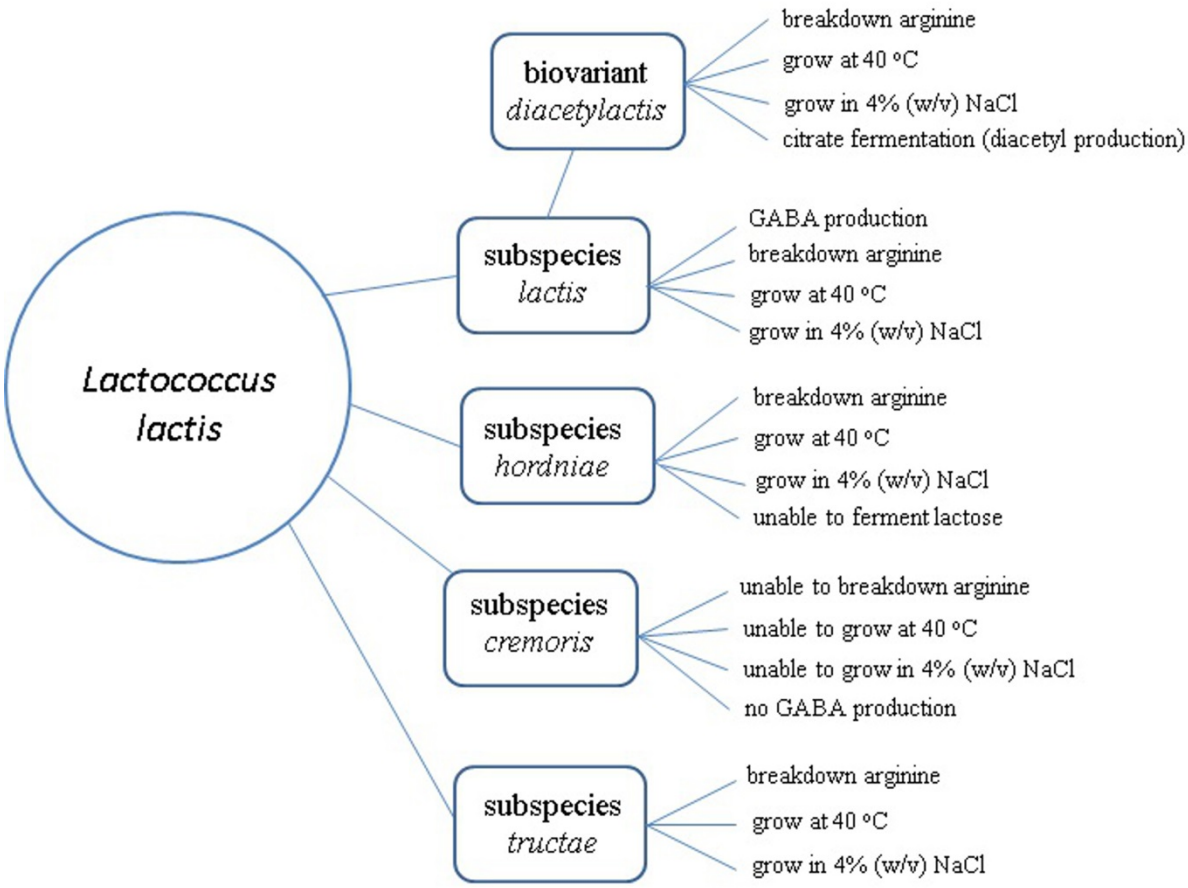


Fig. 1. Phenotypic traits for *Lactococcus lactis* subspecies identification.

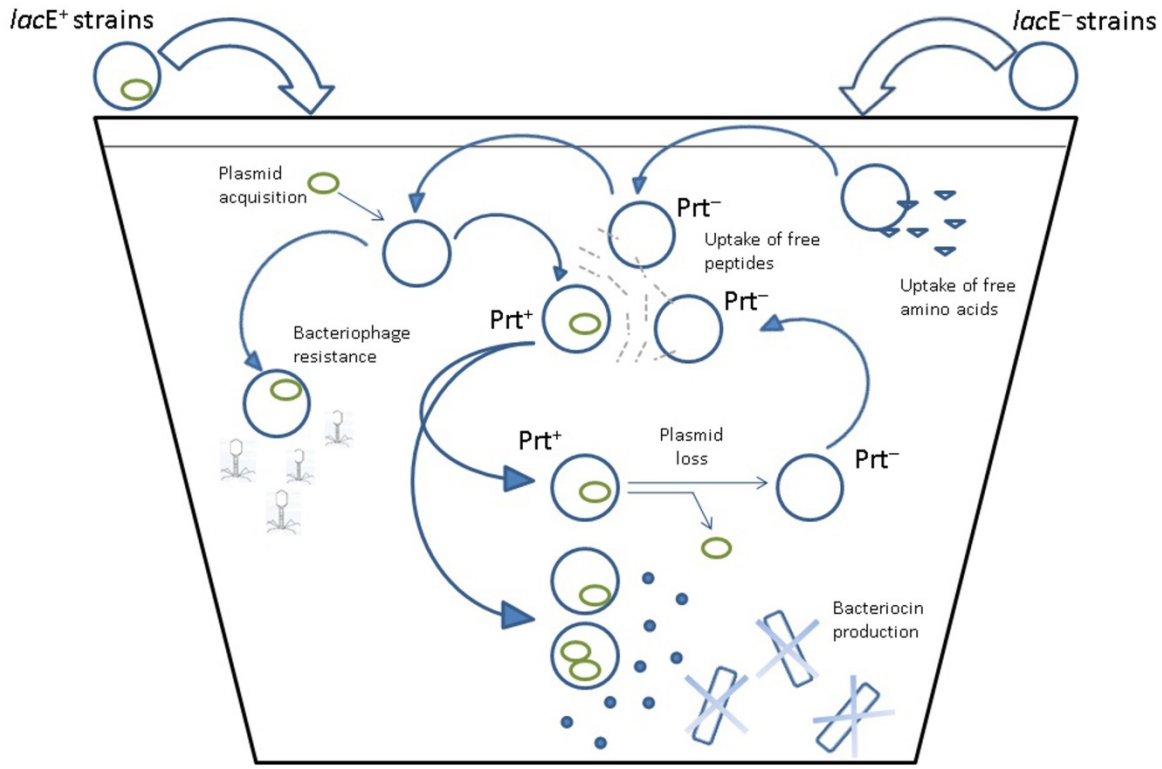


Fig. 2. Events which may have taken place upon initial contamination of milk by non-dairy lactococci which facilitated their adaptation to this environment. Some of these strains may or may not have been capable of utilising lactose (*lacE+* and *lacE-* strains respectively). Following introduction to milk, these strains could have utilised the free peptides released by prt^+ organisms and free amino acids present in the solution. Over time these organisms may have acquired plasmids carrying proteinase genes which, due to the metabolic burden on the cell, may have been readily lost. *L. lactis* may also have acquired other plasmids giving them an advantage such as bacteriocin production and/ or bacteriophage resistance; however, it is more likely that these traits were acquired prior to the introduction to milk.

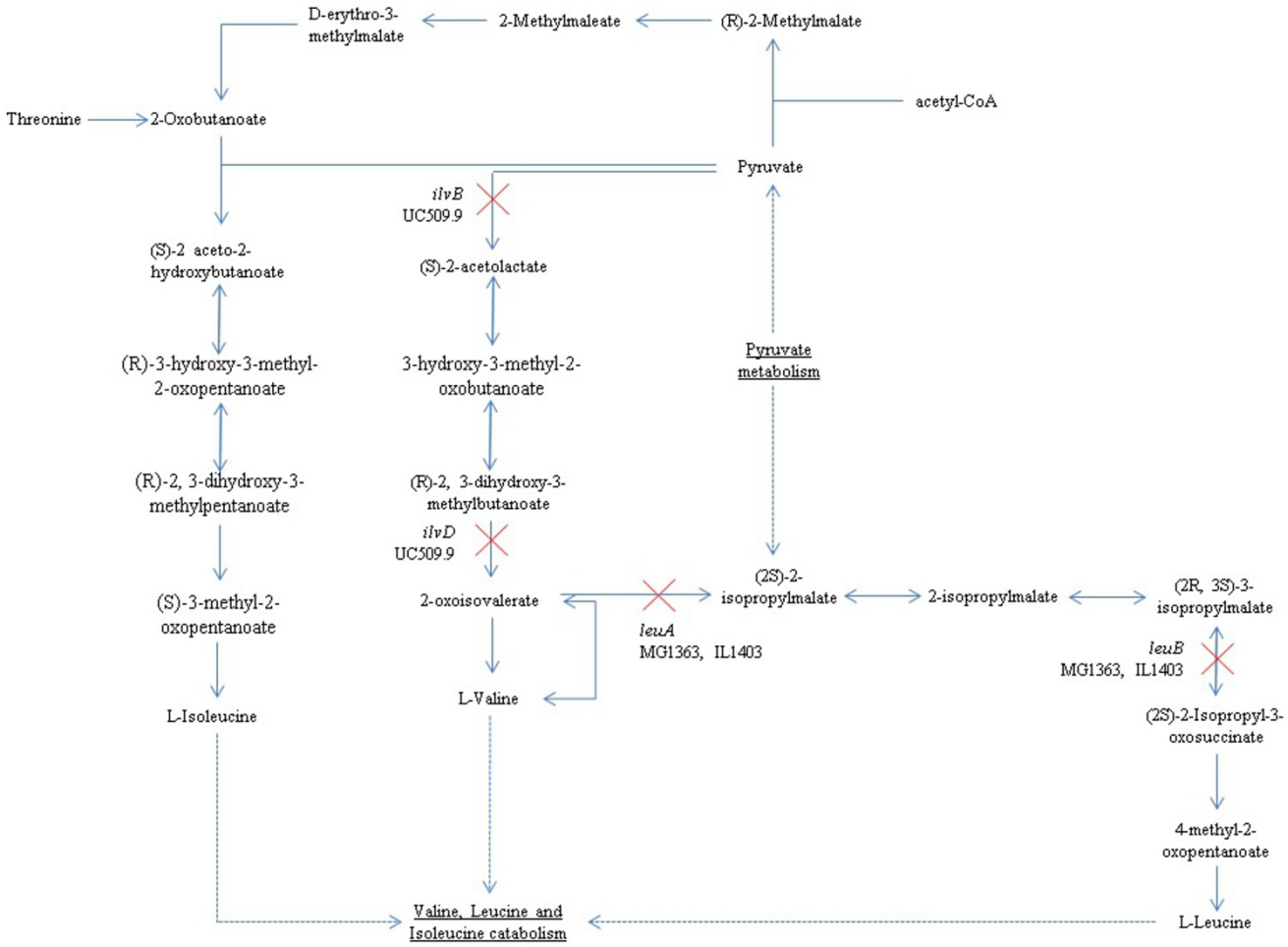


Fig. 3. Biosynthetic pathway for branched chain amino acids (BCAAs) in *Lactococcus lactis*. The crosses denote pseudogenes identified in the dairy strains IL1403, UC509.9 and MG1363.

Chapter 2

Evaluation of non-dairy *Lactococcus lactis* with potential dairy applications reveals extensive phenotype-genotype disparity

All of this chapter has been published in Applied and Environmental Microbiology

(Manuscript in publication)

2.1. Abstract

Lactococcus lactis is predominantly associated with dairy fermentations but evidence suggests that the domesticated organism originated from a plant niche. *L. lactis* possesses an unusual taxonomic structure whereby strain phenotypes and genotype often do not correlate, which in turn has led to confusion in *L. lactis* classification. A bank of *L. lactis* strains was isolated from various non-dairy niches (grass, vegetables, bovine rumen) and were further characterised on the basis of key technological traits including growth in milk and key enzymatic activities. Phenotypic analysis revealed all non-dairy strains to possess a *lactis* phenotype; however, seven of these strains possessed a *cremoris* genotype, determined by two separate PCR assays. Multi-locus sequence typing (MLST) showed that *lactis* and *cremoris* genotypes clustered together regardless of habitat, but highlighted the increased diversity that exists amongst 'wild' strains. Calculation of average nucleotide identity (ANI) and tetranucleotide frequency correlation coefficients (TETRA), using the JSpecies software tool, revealed *cremoris* and *lactis* subspecies differ in ANI values by ~14%; below the threshold set out for species circumscription. Further analysis of strain TIFN3 and strains from non-industrial backgrounds, revealed TETRA values of <0.99 in addition to ANI values <95%, implicating that these 2 groups are separate species. These findings suggest a need for a revision of *L. lactis* taxonomy and highlights the genetic diversity of 'wild' subspecies lactis isolates. As a result of the phenotypes described, the cultures characterised in this study may have applications for adjunct use in cheese manufacturing.

2.2. Introduction

Lactococcus lactis is a member of the lactic acid bacteria (LAB), a group of organisms used worldwide in the production of fermented dairy products. Three *Lactococcus lactis* subspecies exist: *cremoris*, *lactis* and *hordniae*. Many strains of *L. lactis* ssp. *lactis* and *L. lactis* ssp. *cremoris* are typically associated with dairy fermentations but evidence suggests that these organisms originated from a plant niche, and are now considered to be 'domesticated' compared to their so-called 'wild' counterparts. Conversely, *L. lactis* ssp. *hordniae* is unable to utilise lactose and has not been isolated previously from the dairy environment (Schleifer *et al.*, 1985). A citrate metabolising bio-variety also exists, *diacetylactis*, which is capable of imparting buttery aromas in dairy fermentations (Kempler and McKay, 1981). Information regarding subspecies classification is important for the application of cultures in dairy fermentations, particularly cheese production, as subspecies *cremoris* strains are often considered more suitable due to their association with cheeses free from off-flavours (Taïbi *et al.*, 2011).

Before the advent of molecular methods, subspecies classification of *L. lactis* was based on the possession of a number of phenotypic traits. The ability to grow in the presence of 4% NaCl, at 40 ^{o}C, at pH 9.2 and the ability to degrade arginine were assigned as traits for *lactis* subspecies (Orla-Jensen, 1919), while subspecies *cremoris* do not exhibit these characteristics. The ability to ferment maltose is also considered a phenotypic trait of the *lactis* subspecies (Rademaker *et al.*, 2007). In more recent years, genotypic characterisation has largely replaced phenotypic characterisation in subspecies designation of new isolates. Genotypic characterisation based on 16S rRNA is commonly used; however, the use of 16S rRNA sequencing alone can lead to discrepancies in subspecies identification (Xu *et al.*, 2014). This may be due to *L.*

lactis ssp. *lactis* and *cremoris* exhibiting differences in the 16S rRNA gene by as little as 0.7% despite differing up to 15% at the whole genome level (Wegmann *et al.*, 2007). In some instances, the phenotype and genotype do not correlate whereby a strain with a *lactis* genotype possesses a *cremoris* phenotype (Wouters *et al.*, 2002) and conversely for *cremoris* genotypes (Jarvis and Jarvis, 1981; Parapouli *et al.*, 2013). In an attempt to accurately identify subspecies of *L. lactis*, various assays have been developed (Godon *et al.*, 1992; Beimfohr *et al.*, 1997; Pu *et al.*, 2002). These assays employ different molecular techniques such as PCR restriction fragment length polymorphism (PCR-RFLP) and southern blot hybridisation, targeting genes encoding branched chain amino acid biosynthesis and glutamate decarboxylase amongst others (Godon *et al.*, 1992; Nomura *et al.*, 2002).

In particular, two such methods have been developed for rapid and simple species genotype identification by PCR, based on the 16S rRNA gene and the histidine biosynthesis operon (Beimfohr *et al.*, 1997; Pu *et al.*, 2002). The method designed by Pu *et al*. (2002), targets a 19 bp region which differs by 5 bp between subspecies. Beimfohr *et al.* (1997) exploited a 200 bp insertion in the *hisZ* gene to differentiate subspecies, with *cremoris* and *lactis* genotypes forming amplicons of different size. With advances in sequencing technology and the development of genome analytical software tools, faster, more straightforward approaches for accurate species classification are emerging. One such software tool, JSpecies (Richter and Rosselló-Móra, 2009), can be used in defining species based on average nucleotide identity (ANI) and tetranucleotide frequency correlation coefficients (TETRA). What has become obvious upon analysis of the increasing array of *L. lactis* genomes from diverse sources is that *L. lactis* possesses an unusual taxonomy with different genotypes and phenotypes thereof. Indeed, in order to correct for the observed

incongruence between phenotype and genotype, Kelly and Ward (2002) proposed that genotype and phenotype should be stated when referring to the subspecies of a strain.

In this chapter, we describe the phenotypic and genotypic characterisation of a bank of novel *L. lactis* isolates from non-dairy environments and highlight the extensive phenotype-genotype disparity, or mismatching at the subspecies level. Previous analysis of these strains has demonstrated their diverse metabolic capacity, forming different volatile profiles in milk to dairy *L. lactis* strains and their capacity to alter flavour in a Gouda-type cheese model (Cavanagh *et al.*, 2013). Our analysis shows that genome level comparison of these wild strains with the long-domesticated dairy strains further complicates traditional classification and we suggest that a revision of the current species classification of the *L. lactis* species is warranted. In addition to underlining the shortcomings in subspecies definition of *L. lactis,* this work also demonstrates the phenotypic and genotypic diversity that exists between *L. lactis* from different environmental niches and the potential of lactococci from non-dairy niches to be used in dairy fermentations.

2.3. Materials and Methods

2.3.1. Strains and growth conditions

Dairy *L. lactis* strains were provided by the DPC culture collection (Teagasc Food Research Centre, Moorepark, Cork, Ireland). *Lactococcus lactis* strains were propagated in M17 media (Oxoid, Hampshire, England) containing 0.5% lactose monohydrate (VWR, Leuven, Belgium) (wt./ vol.) (LM17) at 30 ^{o}C. *L. lactis* ssp. *lactis* IL1403 and *L. lactis* ssp. *cremoris* MG1363 and HP were cultured propagated in M17 media (Difco) containing 0.5% glucose (Sigma Aldrich, Dublin, Ireland) (GM17).

2.3.2. *Lactococcus* isolation from non-dairy sources

Environmental samples used in this study for the isolation of novel lactococci were sourced at the Teagasc Animal and Grassland Research and Innovation Centre (Moorepark, Fermoy, Cork, Ireland). Grass samples were gathered from allotments assigned to animal grazing. Bovine rumen samples were collected from a canulated cow, fed on a grass-based diet while vegetable samples (baby corn, fresh green peas) were collected from a local supermarket. Fresh sample material (10 g) was added to 100 ml of maximum recovery diluent (MRD) (Oxoid Ltd., England) and homogenised using a stomacher (Seaward, UK). The resulting suspensions were diluted in MRD, plated on LM17 media, and incubated for 48 hrs at 30 °C. Individual colonies were tested for catalase activity (degradation of 3% hydrogen peroxide) and esculin hydrolysis (formation of black) colonies on kanamycin esculin azide agar (KAA) (Oxoid Ltd.). Individual strains were subsequently tested for their ability to grow in milk. This was achieved by twice washing and re-suspending an overnight culture in sterile water and adding a 1.5% inoculum to 10% reconstituted semi-skimmed milk

(RSM) (Kerry Foods, Ireland). Each sample was incubated at 30 ^{o}C and the pH was measured at 5 and 24 hrs.

2.3.3. Subspecies phenotype identification

Lactococcus cultures were tested for their ability to grow at 40 ^{o}C and in 4% NaCl (wt./ vol.) for 24 hrs in LM17 broth. Likewise, growth at 8-10 ^{o}C, 45 ^{o}C and 6.5% NaCl (wt./ vol.) (Sigma) was examined over two days. Growth was determined by measuring absorbance using a Synergy 2 plate reader (BioTek Instruments Inc., Vermont, USA) in comparison to strains IL1403 (*lactis* phenotype) and SK11 (*cremoris* phenotype). Arginine utilisation was assessed using the media described by Beimfohr *et al.* (1997) with the addition of bromocresol purple (0.001% wt./ vol.) in place of phenol red. Acetoin (acetylmethylcarbinol) production from glucose was determined using the Voges-Poskauer test (Dalynn Biologicals) over a period of 72 hrs. Citrate fermentation was assessed using Kempler and McKay agar (Kempler and McKay, 1981) incubated for up to 5 days at 30 ^{o}C under aerobic conditions. *L. lactis* ssp. *lactis* IL1403 and *L. lactis* ssp. *lactis* biovar. *diacetylactis* DRC3, and *L. lactis* ssp. *cremoris* SK11 and MG1363 were used as reference strains for the particular subspecies strains for all phenotypic assays. Each phenotypic assay was performed in two independent experiments in duplicate.

2.3.4. Identification of *Lactococcus* species and subspecies genotypes

Genomic DNA was extracted from overnight cultures using the Genelute Bacterial Genomic DNA kit (Sigma). Using primers employed by Alander *et al*. (1999), 16S rRNA analysis was initially performed to identify putative lactococci (Table 1). Genotypic subspecies designation of *Lactococcus* strains was determined

using two separate PCR reactions targeting the 16S rRNA gene and the histidine biosynthesis operon (Table 1) (Beimfohr *et al.*, 1997; Pu *et al.*, 2002). PCR reactions were carried out in 50 µl reactions using the platinum hi-fidelity PCR Supermix (Invitrogen) containing 50 ng of DNA per reaction. PCR amplicons for 16S rRNA analysis were purified using the high pure PCR clean up micro kit (Roche Diagnostics, Germany) with sequence analysis performed by GATC Biotech in forward and reverse orientation (Cologne, Germany). Sequences were aligned and trimmed using the MegAlign software program (DNAstar Lasergene, Madison, WI) and a consensus sequence was compared to those available in the NCBI BLASTn database.

2.3.5. Multi-locus sequence typing (MLST)

PCR primers were used to amplify partial sequences of seven genes of *L. lactis* for the purposes of MLST (Table 1). The genes selected encode three housekeeping genes: the ATP synthase alpha subunit (*atpA*), the phenylalanine tRNA synthase alpha subunit (*pheS*) and the RNA polymerase alpha subunit (*rpoA*); and three genes involved in flavour formation during fermentation; the lysl aminopeptidase N (*pepN*), the post proline di-peptidyl aminopeptidase (*pepX*) and the branched-chain amino acid transferase (*bcaT*). A small sub-unit (SSU) of the 16S rRNA region was also included in the analysis. DNA was extracted using the GenElute Bacterial Genomic DNA kit (Sigma-Aldrich) according to the manufacturer's instructions for Gram-positive bacteria. PCR amplicons were generated using Platinum Hi-Fi PCR Supermix (Invitrogen) and sequenced in forward and reverse by GATC Biotech (Konstanz, Germany). Sequences were assembled and a consensus sequence generated using the Lasergene software package and the Seqman assembly program (DNAstar Inc., Madison, USA). Sequences were aligned and trimmed using the Bioedit biological

sequence alignment editor (available at: http://www.mbio.ncsu.edu/bioedit/bioedit.html) and entered into the PubMLST non-redundant database (NRDB) in order to identify similar genotypes over a batch of sequence (Gish and States, 1993). Subsequently, a dataset table was created in which the sequence type (ST) was identified and uploaded to the sequence type analysis and recombinational tests program (START2) (Jolley *et al.*, 2001). Sequence analysis was performed using MEGA version 4 using the neighbor-joining method with 1,000 bootstrap replications (Tamura *et al.*, 2007).

2.3.6. Average nucleotide identity (ANI) analysis

Based on results of MLST analysis, three non-dairy lactococcal strains which branched separately were selected for draft sequencing. Genomic libraries were prepared using the Rapid Library Preparation method recommended by Roche (Roche Diagnostics Ltd., West Sussex, UK). Emulsion PCR and 454 sequencing were performed in the Teagasc Sequencing Centre (Moorepark, Fermoy, Cork) on a 454 FLX Sequencer using standard protocols from the manufacturer (Roche Diagnostics Ltd., West Sussex, UK). The single-end reads were assembled using the SeqMan NGen application of the DNAStar Lasergene Genomics Suite (DNASTAR Inc., Madison, USA), with the software specified to trim any poor quality read sequences which fall below an average quality score of 20 over a window of five base pairs. Average nucleotide identity was performed with the BLAST algorithm (ANIb) using the JSpecies software program (Richter and Rosselló-Móra, 2009). In tandem with ANI, this program also allowed for the calculation of tetranucleotide frequency correlation coefficients (TETRA).

2.3.7. Biosample accession numbers

The complete nucleotide sequences of *L. lactis* genomes have been submitted to the GenBank database under the Biosample accession numbers SAMN03396773, SAMN03396924 and SAMN03396976 for strains DPC6853, DPC6856 and DPC6860 respectively.

2.3.8. Assessment of key technological traits in non-dairy isolates

2.3.8.1. Lactose utilisation

Milk acidification kinetics were assessed over a 17 hr period for individual cultures in real-time using the iCinac pH monitoring system (Alliance instruments, France) with readings taken every minute. *Lactococcus* cultures were washed twice in sterile MRD and added to 10% (wt./ vol.) RSM (Kerry Foods), to generate a 1.5% inoculum and incubated at 30 ^{o}C. Pearce test analysis was performed as described previously (O'Donovan *et al.*, 1996).

2.3.8.2. Amino acid transferase activity (methionine and phenylalanine)

The amino acid transferase activity of dairy and non-dairy strains was determined for the sulphur containing amino acid, methionine, and the aromatic amino acid, phenylalanine. Both reaction mixtures contained 50 µM pyridoxal phosphate, 5 mM α-keto Glutaric acid, 0.5 mM sodium arsenate and 50 mM sodium tetra-borate buffer (pH 8.5) with either 5 mM L-phenylalanine or methionine. Standards were prepared at concentrations of 100, 200, 300, 400, 500 and 1000 mM for phenylpyruvate and α-ketomethylthiobutyrate. For the blank preparation 1 ml of reaction mixture was mixed with 100 µl of distilled water and incubated at 30^{o}C for 30 minutes. One ml of TCA was added and the solution centrifuged at 12,000 x *g* for

2 minutes. Analysis was performed using a Cary 100 Bio UV-Vis spectrophotometer (Varian, Netherlands) with the wavelength set to 300 nm. Test mixtures were analysed as per the blank with the addition of 100 µl of bacterial culture grown to pH 5.7 and read against the standard curve. The instrument was zeroed prior to each measurement and each sample was assayed in triplicate. Amino acid transferase activity was expressed as µmoles/ min/ mg of protein.

2.3.8.3. Antibiotic resistance and biogenic amine formation

Minimum inhibitory concentrations (MICs) for antibiotics were evaluated using the VetMIC system (National Veterinary Institute of Sweden, Uppsala, Sweden) for LAB. Each micro-titre plate contained 2-fold serial dilutions of 16 antibiotics. Following growth in LAB susceptibility test broth, each culture was re-suspended in MRD as per the manufacturer's instructions to create the desired cell density (~3 x 10^5 CFU/ ml). One hundred µl of cell suspension was added to each well on the micro-titre plate and incubated for 48 hrs at 28 ^oC. The lowest antibiotic concentration at which no growth occurred was defined as the minimum inhibitory concentration for each antibiotic. Antibiotic resistance was examined in duplicate.

The presence of histidine and tyrosine decarboxylase genes were determined by PCR using the primer sets outlined by Coton and Coton (2005) and Coton *et al*. (2004). *Lactobacillus brevis* DPC6660 and *Lactobacillus buchneri* DPC6666 were used as tyrosine and histamine decarboxylase positive controls respectively with *Lactococcus lactis* ssp. *lactis* DRC3 used a negative control.

2.4. Results

2.4.1. Species confirmation of non-dairy isolates

A bank of novel non-dairy strains isolated in the course of this study from grass, vegetables and the bovine rumen was further characterised. Individual colonies were first examined for catalase activity (to eliminate catalase-positive organisms) and esculin hydrolysis on KAA (to eliminate *Enterococcus* species). Isolates which were negative for both traits were examined by Gram reaction and for growth in milk. Gram positive isolates capable of growth in milk were subsequently identified by 16S rRNA analysis. Using the primer set described by Alander *et al*. (1999), a 1.5 kb amplicon was generated for all non-dairy strains targeting the 16S rRNA gene. Subsequent analysis of the sequences confirmed a number of isolates as *Lactococcus lactis*. All novel *L. lactis* strains were deposited in the DPC culture collection, Teagasc Food Research Centre Moorepark, Co. Cork, Ireland.

2.4.2. Subspecies phenotype identification

Subspecies *lactis* phenotypes are identified by the ability to grow in 4% NaCl, at 40 ^{o}C, and the capability to degrade arginine. In addition, the biovariant *diacetylactis* has the ability to ferment citrate to produce diacetyl. Subspecies *cremoris* phenotypes are negative for each of these attributes. To establish subspecies phenotypes of the isolated non-dairy lactococci, each isolate was examined for the aforementioned characteristics (Table 2) and compared to a bank of well-characterised dairy lactococci which are commonly used in the dairy industry. All of the non-dairy lactococci isolated in this study were able to grow at 40 ^{o}C and in media containing 4% NaCl. Seven of the eight non-dairy lactococcal strains were able to hydrolyse arginine. Strain DPC6855 was unable to hydrolyse arginine but could grow in 4%

NaCl and at 40 ºC, and was therefore assigned as subspecies *lactis*. Only the dairy strain, DRC3, was capable of fermenting citrate which is the hallmark of biovariant *diacetylactis* strains. Based on these results, all non-dairy lactococci were assigned *lactis* phenotypes.

2.4.3. Subspecies genotype identification

The non-dairy *L. lactis* isolates were analysed at the genotypic level to accurately identify subspecies, and were again compared to the bank of dairy lactococci. Two separate PCR-based assays were performed in order to identify the subspecies genotype of each strain. The 16S rRNA gene exhibits a 0.07% variance depending on the strain (Salama *et al.*, 1991). Primers designed by Pu *et al.* (2002) were employed to discriminate between *lactis* and *cremoris* genotypes on the basis of this difference. The primers sets comprise a common reverse primer, LacreR, and a subspecies-specific forward primer, LacF, for subspecies *lactis* and CreF for subspecies *cremoris* to generate a 163 bp amplicon. In all cases, amplicons were generated with one or other of the subspecies-specific forward primers, but not for both (Table 3). To further distinguish between subspecies *lactis* and *cremoris*, Biemfohr *et al.* (1997) designed a single primer set to amplify specific regions of the histidine biosynthesis operon. Between *lactis* and *cremoris* genotypes, the average sequence divergence from positions 480-700 within the histidine biosynthesis operon equates to 45%. In addition, members of the *cremoris* subspecies possess a 200 bp insertion at the end of *hisZ* gene within this operon which encodes a phosphoribosyltransferase regulatory subunit (Beimfohr *et al.*, 1997). Using the primer set designed by Biemfohr *et al.* (1997), amplification of subspecies *cremoris* results in a 1149 bp product, and subspecies *lactis* a 934 bp product (Table 3). Subspecies genotypes determined using

both assays correlated with one another, verifying the genotype of the strains. In all, seven *cremoris* genotype, *lactis* phenotype strains were isolated from grass and rumen samples and one *lactis* genotype, *lactis* phenotype strain was isolated from corn (Table 3).

2.4.4. Genetic diversity of *L. lactis* strains from different environments

2.4.4.1. Clustering of *lactis* and *cremoris* genotypes

MLST is a comprehensive technique compared to other typing methods as it provides information on aspects of the population organization and evolution of the species in question (Maiden, 2006; Turner and Feil, 2007). In contrast, multi-locus sequence analysis (MLSA) concatenates the DNA sequence of the selected housekeeping genes and using this information examines the phylogenetic relationship between strains. Eight dairy and eight non-dairy strains were examined for their genetic diversity using comparative sequence analysis of seven genes including the 16S rRNA gene. The primer sets used in MLST analysis (Table 1) do not generate products for *L. lactis* ssp. *hordniae* or other *Lactococcus* species (Rademaker *et al.*, 2007). Amplicons were generated for segments of each of these genes for all 16 *Lactococcus* strains tested and sequenced in forward and reverse orientation and a consensus sequence generated. Sequence data for each of the seven housekeeping genes were also extracted from the complete genome sequences of *L. lactis* strains from different habitats available on the NCBI database. Sequence types (STs) were created for all strains using the non-redundant database program (Gish and States, 1993). For the non-dairy lactococci isolated in this study, four STs were identified: (1) DPC6853; (2) DPC6854 and DPC6855; (3) DPC6856; and (4) DPC6857, DPC6858, DPC6859 and DPC6860. MLST analysis, including our non-dairy bank

and sequenced *L. lactis* genomes from the NCBI database, identified 20 different STs (eight *lactis* genotypes; 12 *cremoris* genotypes). The frequency of polymorphic sites ranged from two in the 16S rRNA gene to 85 in the *pepX* gene (Table 4). The sequence divergence at these sites was identified as not resulting in a change in the amino acid, as determined by the low *dN*/ *dS* ratio identified for all alleles. Collective analysis of all seven loci, and each of the genes individually, showed the presence of two separate clusters corresponding to *cremoris* and *lactis* genotype strains which corresponds to the findings of other reports using the same scheme (Fig. 1; 2) (Rademaker *et al.*, 2007, Fernández *et al.*, 2011). *L. lactis* NZ9000 is a derivative of *L. lactis* MG1363 and both strains possessed the same ST and consistently grouped with each other throughout the analysis, thereby confirming comparative analysis of the gene sequences (Linares *et al.*, 2010). Overall, results of our MLST analysis clearly show strains of subspecies *lactis* and *cremoris* grouping separately, irrespective of the environment from which they were isolated.

2.4.4.2. ANI: Revision of *Lactococcus* subspecies as separate species

To further examine the genetic diversity of the non-dairy lactococcal bank, ANI was calculated for *L. lactis* strains DPC6853 (corn), DPC6856 (bovine rumen) and DPC6860 (grass) using draft genome sequences, and compared to *L. lactis* genomes available on the NCBI database. ANI is an alternative to DNA-DNA hybridisation, using genome sequences, as a means of defining a species (Richter and Rosselló-Móra, 2009). Table 5 shows ANIb (ANI calculated using the BLAST algorithm) values for *L. lactis* strains available on the NCBI database plus strains DPC6856, DPC6853 and DPC6860 isolated in this study, calculated using the JSpecies software tool (Richter and Rosselló-Móra, 2009). Firstly, as expected, derivatives of the same strain share a

100% ANIb, as in the case of *L. lactis* ssp. *cremoris* NZ9000 and *L. lactis* ssp. *cremoris* MG1363 (Kuipers *et al.*, 1998). Secondly, all strains belonging to the same subspecies possessed an ANIb value between 96.53-99.96%. Contrastingly, strains reported as different subspecies shared ANIb values of 85.54-87.45%, below the cut-off for species circumscription (<95%). Based on ANIb values alone, the authors propose that *lactis* and *cremoris* genotypes be reclassified as different species

In conjunction with an ANI of <95%, a tetranucleotide frequency correlation coefficient (TETRA) of <0.99 is also used in tandem with ANI for species circumscription (Richter and Rosselló-Móra, 2009). When compared to strain *L. lactis* ssp. *cremoris* TIFN3, used in cheese production, the non-dairy strains DPC6860 and DPC6856 (genotype *cremoris*, phenotype *lactis*) possessed lower ANIb values of 96.16% and 96.83% respectively compared to other strains designated as *cremoris* subspecies (~97%). Similarly, strain DPC6853 possessed a lower ANIb value of 96.39-97.24% in comparison to other *L. lactis* ssp. *lactis* strains. This suggests an increased genetic diversity amongst strains isolated in this study in comparison to previously sequenced strains of the same genotype. In addition this data also highlights a reduced genetic diversity amongst dairy starter strains as identified previously (Passerini *et al*., 2010).

For the vast majority of comparisons between subspecies, a below threshold ANIb value was coupled with a TETRA value >0.99 which classifies them as the same species (Table 6). However, the dairy strain, *L. lactis* ssp. *cremoris* TIFN3 possessed ANIb and TETRA values below the threshold for species circumscription when compared to strain DPC6853. Furthermore when compared to 'wild' *L. lactis* ssp. *lactis* strains and strain IL1403, similar ANIb and TETRA values were also observed; which were not found between strain TIFN3 and dairy subspecies *lactis* strains. The

phenotypes of some of these 'wild' strains are not defined in reports; however, it is tempting to speculate that all of these strains possess *lactis* phenotypes owing to the environment from which they were isolated. These findings suggest that *L. lactis* TIFN3 is a separate species to 'wild' subspecies *lactis* strains and IL1403, which may possess *lactis* phenotypes. The genome sequence status of TIFN3 is insufficient to support this reclassification however; future sequence analysis of the TIFN3 genome may provide evidence to support these findings.

2.4.5. Key technological traits

2.4.5.1. Starter activity: milk acidification

The fast acidification of milk is a crucial trait of dairy cultures for use in the dairy industry (Marshall, 1991). All strains isolated in this study possessed good acidification activity when grown in RSM at 30 ^{o}C (Fig. 3). Strain DPC6856 (isolated from the bovine rumen) showed the slowest production of acid, reaching pH 4.75 after 17 hours compared to 4.15 for other strains. This is in comparison to the dairy strains, the fastest of which strain ML8 reached a final pH of 4.34 in 17 hours. The Pearce test simulates conditions that starter cultures are be exposed to in Cheddar cheese making, which in turn allows for the evaluation of strains for use as starter cultures (Pearce, 1969). Pearce test analysis (Fig 4) showed that non-dairy strains would be unsuitable for use as starters as they are unable to reach the desired pH under processing conditions (as mimicked by the Pearce test). Although these strains are unsuitable as starters they are capable of growth in milk without the use of supplementation and on this basis they were used as adjuncts.

2.4.5.2. Enzyme activity analysis

Aminotransferases function in converting amino acids to their respective α-keto acids which are further converted to flavour compounds during cheese ripening (Yvon *et al.,* 1997; Rijnen *et al.,* 2003; Smit *et al.,* 2004). Dairy and non-dairy cultures were examined for their amino acid transferase activity for the aromatic amino acid, phenylalanine, and the sulphur amino acid, methionine in duplicate experiments (Fig. 5). To enable the comparative assessment of amino acid transferase activity of strains, laboratory media was used for culturing cells prior to performing enzyme assays. Non-dairy isolates displayed increased activity for phenylalanine compared to decreased activity for methionine relative to dairy strains. *L. lactis* H88M1 showed the lowest transferase activity for phenylalanine, and *L. lactis* DRC3 and ML8 showed the highest activity for methionine. Conversely, strains DPC6854 and DPC6855 had the lowest activity for methionine while DPC6858 had the highest activity for phenylalanine.

2.4.5.3. Antibiotic resistance

The antibiotic resistance of bacteria destined for use in food, needs to be carefully assessed to prevent dissemination of these genes to other bacteria along the food chain. Therefore, the antibiotic resistance of all non-dairy strains and three representative dairy strains were determined using VetMIC Lact-1 and Lact-2 plates analysing a total of 16 antibiotics. Few differences were observed in the MIC profiles between dairy and non-dairy strains (Table 7). Non-dairy isolates showed a lower MIC for ciprofloxacin (DPC6854, DPC6855), vancomycin (DPC6855, DPC6857, DPC6858, DPC6859, DPC6860) ampicillin (DPC6855) and clindamycin (DPC6854). All strains tested were highly resistant to trimethoprim (Tm), which corresponds to previous reports on *Lactococcus* strains isolated from raw milk cheeses. Strain DPC6853 possessed a much higher MIC for tetracycline than any of the other strains

tested. With the exception of DPC6853, all non-dairy strains isolated in this study were below the microbiological cut-off values for antimicrobial resistance set out by EFSA (2012). No histidine or tyrosine decarboxylase genes were identified by PCR in any of the non-dairy lactococci isolated in this study.

2.5. Discussion

With the advancement of molecular tools and high throughput screening methods, the classification of *L. lactis* subspecies has come under much scrutiny and has sparked debate as to how the subspecies should be identified. Phenotypic tests have been used in the past to distinguish subspecies; however in some cases these tests may not account for related organisms that show unusual phenotypic traits (Centeno *et al.,* 1996; Teixeira *et al.,* 1996; Delgado and Mayo, 2004). Numerous reports have identified subspecies genotypes with mismatching phenotypes (Salama *et al.,* 1993; Fernández *et al*, 2011; Parapouli *et al*., 2013). Salama *et al.* (1993) examined *L. lactis* strains, isolated from natural environments, which were identified by colony hybridisation using species and subspecies specific probes. Fernández *et al.* (2011) examined the phenotypic and genetic diversity of *L. lactis* strains from raw milk cheeses while Parapouli *et al.* (2013) characterised an *L. lactis* strain isolated from raw milk, which produced a novel Nisin A bacteriocin. In this study, seven of the eight strains isolated from various non-dairy niches were found to possess *cremoris* genotypes with one *lactis* genotype isolated from corn. All of these strains possessed a *lactis* phenotype, since seven strains were capable of hydrolysing arginine and growing in 4% NaCl and at 40 ^{o}C, with one strain unable to hydrolyse arginine. Previously *L. lactis* strains from alfalfa, radish sprouts, and mung bean sprouts were incapable of hydrolysing arginine but overall possessed a *lactis* phenotype (Rademaker *et al.,* 2007). As a result of adaptation to milk, *L. lactis* is held to have undergone a process of reductive evolution forfeiting processes no longer required in the milk environment (Kelly *et al.,* 2010). The rarity of *cremoris* phenotypes outside of milk suggests that growth at 40 ^{o}C, in 4% NaCl and capacity to utilize arginine are not required in this environment and it has been hypothesised that the *lactis* phenotype is

required for growth in diverse environments outside of milk (Klijn *et al.*, 1995; Kelly and Ward, 2002). Our findings support this hypothesis whereby seven subspecies *cremoris* strains from non-dairy niches were isolated, all of which possessed *lactis* phenotypes.

It has been estimated, according to divergence in the 16S rRNA gene, that *lactis* and *cremoris* genotypes separated approximately 17 million years ago (Bolotin *et al.*, 2004). Diversity analysis using the composite data set of the seven loci employed in this study, clearly highlighted this separation with *cremoris* and *lactis* subspecies grouping separately irrespective of the environment from which they were isolated. Previously, phylogenetic analysis of dairy and non-dairy lactococci suggested the relatively recent emergence of dairy lactococcal strains domesticated to milk (Rademaker *et al.*, 2007, Passerini *et al.*, 2010). Therefore, although these *lactis* and *cremoris* genotypes may be phylogenetically close, they are following different directions with respect to their evolution. This may account for the separate grouping of environmental *lactis* genotypes with respect to dairy *lactis* genotypes during neighbour joining cluster analysis of individual genes (with the exception of the 16S rRNA gene and *rpoA*) and the seven loci data set. *L. lactis* CV56, isolated from human vaginal samples, clustered with strains of dairy origin hinting that this organism may have originated from a dairy environment (Gao *et al.*, 2011). Similarly, DPC6856 isolated from the bovine rumen, clustered with grass strains suggesting that it is derived from a grass niche.

ANI is an alternative to DNA-DNA hybridisation as a means of species circumscription, using full or partial genome sequences (Richter and Rosselló-Móra, 2009). An ANI value of 95% is considered the threshold for species definition, corresponding to a DDH similarity value of 70% (Richter and Rosselló-Móra, 2009).

It is important to note that software tools such as JSpecies allow for pair-wise comparisons only using ANI. With regards to future development of such software tools, there is a call to create a feature where ANI for a cluster of strains (i.e. *cremoris* strains) can be compared to another cluster i.e. (*lactis* strains); similar to that as described by Altermann (2012) (in this analysis ORFs are compared rather than average nucleotide sequence). Comparison of ANI values showed that subspecies *cremoris* and *lactis* possessed approximately 86% DNA similarity, in agreement with the figure set out by Wegmann *et al.* (2007) between strains IL1403 and MG1363. Based on our analysis and under the guideline set-out by Richter and Rosselló-Móra (2009), *L. lactis* ssp. *lactis* strains are a different species to subspecies *cremoris* strains possessing an ANI value of below 95%. Contrastingly, Fernández *et al.* (2011) proposed that *cremoris* and *lactis* genotype strains are members of the same species, based on the species concept for prokaryotes (Rosselló-Mora and Amann, 2001). In addition Fernández *et al.* (2011) also put forward that both *cremoris* and *lactis* subspecies genotypes with *lactis* phenotypes represent true subspecies. In comparison to species, subspecies definition is quite vague whereby no established molecular cut-off values exist, and subspecies classification is often at the judgement of the taxonomist (Konstantinidis and Stackebrandt, 2013) i.e. no threshold values for subspecies classification using ANIb or DDH are established. The analysis performed in this study was based on genomic sequences only and did not take into account any of the physical traits of the strains. Rosselló-Mora and Amann (2001) stated that a species can consist of different genomic groups, and with in-depth phenotypic analysis better circumscription can be obtained. Therefore in conjunction with ANI and TETRA, further analysis of both the phenotype and chemotaxonomic markers of

analysed strains are required to support the proposed revision of Lactococcus subspecies as separate species.

Recently, the isolation of novel lactic acid bacteria strains from diverse ecological niches has gained renewed interest as it has been shown that such strains possess more diverse metabolic traits than established dairy cultures (Alemayehu *et al.*, 2014). In the present study we examined technological traits important in the production of fermented dairy products. All non-dairy strains grew well in milk and were capable of acidifying milk at a similar rate to previously isolated lactococci from dairy and non-dairy origins (Nomura *et al.*, 2006). Amino acid transferase activity clearly showed that non-dairy isolates possess an increased capacity to trans-animate phenylalanine in contrast to the sulphur derived amino acid, methionine. Sulphur compounds derived from methionine impart onion, garlic, and cabbage flavour notes to cheeses such as found in Cheddar and Camembert (McSweeney and Sousa, 2000; Singh *et al.*, 2003). Contrastingly aroma compounds originating from aromatic amino acids can contribute to off-flavours in cheeses such as flowery, bitter almond and rosy aromas (van Kranenburg *et al.*, 2002; Marilley and Casey, 2004). The conversion products of phenylalanine have been found in hard- and soft-type cheeses (van Kranenburg *et al.*, 2002) and in small amounts, the production of these compounds may be beneficial in diversifying the flavour profiles of semi-hard cheeses.

Antibiotics can persist for an extended period of time in soil and water based environments, thus the widespread use of agricultural antibiotics may select for resistant strains of bacteria in these habitats (Thiele-Bruhn, 2003; Segura *et al.*, 2009; Allen *et al.*, 2010). Therefore, if non-dairy or 'wild' isolates such as those identified in this study are to be used as cultures in food processing, their antibiotic resistance must be carefully assessed and fall within guidelines set out by EFSA (2012). These

guidelines state the cut-off values for *L. lactis* for ampicillin (2 mg/ L), vancomycin (4 mg/ L), gentamycin (32 mg/ L), kanamycin (64 mg/ L), streptomycin (32 mg/ L), erythromycin (1 mg/ L), clindamycin (1 mg/ L), tetracycline (4 mg/ L) and chloramphenicol (8 mg/ L). *L. lactis* DPC6853 possessed a much higher MIC for tetracycline, an antibiotic used commonly in the treatment of human and animal diseases (Chopra and Roberts, 2001). Increased resistance to this compound has previously been identified in *L. lactis* strains from raw milk cheeses and was found to be plasmid encoded (Fallico *et al.*, 2011). This may raise some doubt about the suitability of DPC6853 for use in food production; however, if this resistance is plasmid-encoded, sensitivity of this strain to this antimicrobial could be restored (Fallico *et al.*, 2011).

2.6. Conclusions

All non-dairy *L. lactis* strains isolated in this study demonstrated some key technological traits for application in dairy fermentations and were found to possess *lactis* phenotypes, with seven out of eight isolates possessing *cremoris* genotypes. MLST clearly separated *lactis* and *cremoris* genotypes but highlighted the diversity which exists between 'wild' isolates and their dairy counterparts. Further examination of a number of these strains by ANI and TETRA suggests that the classification of *L. lactis* requires revision in light of the many non-dairy lactococci being isolated and sequenced. With the increased mining of diverse environments for novel *L. lactis* strains, a standard approach for species/ subspecies classification using analysis such as ANI is required. Similarly, it is also crucial to accurately identify the subspecies phenotype and other chemotaxonomic markers. Our analysis suggests the feasibility of reclassifying the *L. lacti*s subspecies *lactis* and *cremoris* into two separate species, *L. lactis* and *L. cremoris*.

2.7. Acknowledgements

Thanks to Aidan Casey for genome assembly, annotation and assistance in ANI analysis.

Thanks to Paul Cotter for genome sequencing.

2.8. References

Alander, M., Satokari, R., Korpela, R., Saxelin, M., Vilpponen-Salmela, T., Mattila-Sandholm, T., and Von Wright, A. (1999). Persistence of colonization of human colonic mucosa by a probiotic strain, *Lactobacillus rhamnosus* GG, after oral consumption. *Appl. Environ. Microbiol.* 65**,** 351-354.

Alemayehu, D., Hannon, J. A., McAuliffe, O., and Ross, R. P. (2014). Characterization of plant-derived lactococci on the basis of their volatile compounds profile when grown in milk. *Int. J. Food Microbiol.* 172**,** 57-61. DOI: 10.1016/j.ijfoodmicro.2013.11.024.

Allen, H. K., Donato, J., Wang, H. H., Cloud-Hansen, K. A., Davies, J., and Handelsman, J. (2010). Call of the wild: antibiotic resistance genes in natural environments. *Nature Rev. Microbiol.* 8**,** 251-259.

Altermann, E. (2012). Tracing lifestyle adaptation in prokaryotic genomes. *Front. Microbiol*. 3, 1-17.

Beimfohr, C., Ludwig, W., and Schleifer, K.-H. (1997). Rapid genotypic differentiation of *Lactococcus lactis* subspecies and biovar. *Syst. Appl. Microbiol.* 20**,** 216-221.

Centeno, J., Menéndez, S., and Rodriguez-Otero, J. (1996). Main microbial flora present as natural starters in Cebreiro raw cow's-milk cheese (Northwest Spain). *Int. J. Food Microbiol.* 33**,** 307-313.

Chopra, I., and Roberts, M. (2001). Tetracycline antibiotics: mode of action, applications, molecular biology, and epidemiology of bacterial resistance. *Microbiol. Mol. Biol. Rev.* 65**,** 232-260.

Coton, E., and Coton, M. (2005). Multiplex PCR for colony direct detection of Gram-positive histamine-and tyramine-producing bacteria. *J. Microbiol. Meth.* 63**,** 296-304.

Coton, M., Coton, E., Lucas, P., and Lonvaud, A. (2004). Identification of the gene encoding a putative tyrosine decarboxylase of *Carnobacterium divergens* 508. Development of molecular tools for the detection of tyramine-producing bacteria. *Food Microbiol.* 21**,** 125-130.

Delgado, S., and Mayo, B. (2004). Phenotypic and genetic diversity of Lactococcus lactis and Enterococcus spp. strains isolated from Northern Spain starter-free farmhouse cheeses. *Int. J. Food Microbiol.* 90**,** 309-319.

EFSA Panel on Additives and Products or Substances used in Animal Feed (FEEDAP). (2012). Guidance on the assessment of bacterial susceptibility to antimicrobials of human and veterinary importance. *EFSA Journal*. 10 (6), 2740. [10 pp.] DOI: 10.2903/j.efsa.2012.2740.

Fallico, V., McAuliffe, O., Fitzgerald, G. F., and Ross, R. P. (2011). Plasmids of raw milk cheese isolate *Lactococcus lactis* subsp. *lactis* biovar. *diacetylactis* DPC3901 suggest a plant-based origin for the strain. *Appl. Environ. Microbiol.* 77**,** 6451-6462. DOI: 10.1128/AEM.00661-11.

Fernández, E., Alegría, Á., Delgado, S., Martín, M. C., and Mayo, B. (2011). Comparative phenotypic and molecular genetic profiling of wild *Lactococcus lactis* ssp. l*actis* strains of the *L. lactis* ssp. *lactis* and *L. lactis* ssp. cremoris genotypes, isolated from starter-free cheeses made of raw milk. *Appl. Environ. Microbiol.* 77**,** 5324-5335.

Gao, Y., Lu, Y., Teng, K.-L., Chen, M.-L., Zheng, H.-J., Zhu, Y.-Q., and Zhong, J. (2011). Complete genome sequence of *Lactococcus lactis* ssp. *lactis* CV56, a probiotic strain isolated from the vaginas of healthy women. *J. Bacteriol.* 193**,** 2886-2887. DOI: 10.1128/JB.00358-11.

Gish, W., and States, D. J. (1993). Identification of protein coding regions by database similarity search. *Nat. Genet.* 3**,** 266-272.

Godon, J.-J., Delorme, C., Ehrlich, S. D., and Renault, P. (1992). Divergence of genomic sequences between *Lactococcus lactis* subsp. *lactis* and *Lactococcus lactis* subsp. *cremoris*. *Appl. Environ. Microbiol.* 58**,** 4045-4047.

Jarvis, A. W., and Jarvis, B. D. (1981). Deoxyribonucleic acid homology among lactic streptococci. *Appl. Environ. Microbiol.* 41**,** 77-83.

Jolley, K. A., Feil, E., Chan, M.-S., and Maiden, M. C. J. (2001). Sequence type analysis and recombinational tests (START). *Bioinformatics.* 17**,** 1230-1231.

Kelly, W., and Ward, L. (2002). Genotypic vs. phenotypic biodiversity in *Lactococcus lactis. Microbiol.* 148**,** 3332-3333.

Kelly, W. J., Ward, L. J. H., and Leahy, S. C. (2010). Chromosomal diversity in *Lactococcus lactis* and the origin of dairy starter cultures. *Genome Biol. Evol.* 2**,** 729.

Kempler, G., and McKay, L. (1981). Biochemistry and genetics of citrate utilization in *Streptococcus lactis* ssp. *diacetylactis*. *J. Dairy Sci.* 64**,** 1527-1539.

Klijn, N., Weerkamp, A. H., and De Vos, W. M. (1995). Detection and characterization of lactose-utilizing *Lactococcus* spp. in natural ecosystems. *Appl. Environ. Microbiol.* 61**,** 788-792.

Konstantinidis, K. T., and Stackebrandt, E. (2013). "Defining taxonomic ranks," in *The prokaryotes*. (Germany: Springer), 229-254.

Kuipers, O. P., De Ruyter, P. G., Kleerebezem, M., and De Vos, W. M. (1998). Quorum sensing-controlled gene expression in lactic acid bacteria. *J. Biotechnol.* 64**,** 15-21.

Maiden, M. C. (2006). Multilocus sequence typing of bacteria. *Annu. Rev. Microbiol.* 60**,** 561-588.

Marilley, L., and Casey, M. (2004). Flavours of cheese products: metabolic pathways, analytical tools and identification of producing strains. *Int. J. Food Microbiol.* 90**,** 139-159.

Marshall, V. M. (1991). Inoculated ecosystem in a milk environment. *J. Appl. Bacteriol.* 73**,** 127S-135S.

McSweeney, and Sousa, M. J. (2000). Biochemical pathways for the production of flavour compounds in cheeses during ripening: A review. *Le Lait.* 80**,** 293-324.

Nomura, M., Kobayashi, M., and Okamoto, T. (2002). Rapid PCR-based method which can determine both phenotype and genotype of *Lactococcus lactis* subspecies. *Appl. Environ. Microbiol.* 68**,** 2209-2213. DOI: 10.1128/AEM.68.5.2209-2213.2002.

Nomura, M., Kobayashi, M., Narita, T., Kimoto-Nira, H., and Okamoto, T. (2006). Phenotypic and molecular characterization of *Lactococcus lactis* from milk and plants. *J. Appl. Microbiol.* 101**,** 396-405. DOI: 10.1111/j.1365-2672.2006.02949.x.

O'Donovan, C. M., Wilkinson, M. G., Guinee, T. P., and Fox, P. F. (1996). An investigation of the autolytic properties of three lactococcal strains during cheese ripening. *Int. Dairy J.* 6**,** 1149-1165.

Orla-Jensen, S. (1919). *The lactic acid bacteria.* Copenhagen, Denmark: Høst and son.

Parapouli, M., Delbès-Paus, C., Kakouri, A., Koukkou, A.-I., Montel, M.-C., and Samelis, J. (2013). Characterization of a wild, novel nisin A-producing *Lactococcus* strain with an *L. lactis* subsp. *cremoris* genotype and an *L. lactis* subsp. *lactis* phenotype, isolated from Greek raw milk. *Appl. Environ. Microbiol.* 79**,** 3476-3484. DOI: 10.1128/AEM.00436-13.

Passerini, D., Beltramo, C., Coddeville, M., Quentin, Y., Ritzenthaler, P., Daveran-Mingot, M.-L., and Le Bourgeois, P. (2010). Genes but not genomes reveal bacterial domestication of *Lactococcus lactis*. *PLoS One.* 5**,** e15306. DOI: 10.1371/journal.pone.0015306.

Pearce, L. E. (1969). Activity tests for cheese starter cultures. *New Zealand J. Dairy Technol.* 4**,** 246-247.

Pu, Z., Dobos, M., Limsowtin, G., and Powell, I. (2002). Integrated polymerase chain reaction based procedures for the detection and identification of species and subspecies of the Gram positive bacterial genus *Lactococcus*. *J. Appl. Microbiol.* 93**,** 353-361.

Rademaker, J. L. W., Herbet, H., Starrenburg, M. J. C., Naser, S. M., Gevers, D., Kelly, W. J., Hugenholtz, J., Swings, J., and Van Hylckama Vlieg, J. E. T. (2007). Diversity analysis of dairy and nondairy *Lactococcus lactis* isolates, using a novel multilocus sequence analysis scheme and (GTG) 5-PCR fingerprinting. *Appl. Environ. Microbiol.* 73**,** 7128-7137. DOI: 10.1128/AEM.01017-07.

Richter, M., and Rosselló-Móra, R. (2009). Shifting the genomic gold standard for the prokaryotic species definition. *PNAS.* 106**,** 19126-19131.

Rijnen, L., Yvon, M., Van Kranenburg, R., Courtin, P., Verheul, A., Chambellon, E., and Smit, G. (2003). Lactococcal aminotransferases AraT and BcaT are key enzymes for the formation of aroma compounds from amino acids in cheese. *Int. Dairy J.* 13**,** 805-812.

Rosselló-Mora, R., and Amann, R. (2001). The species concept for prokaryotes. *FEMS Microbiol. Rev.* 25**,** 39-67.

Salama, M., Sandine, W., and Giovannoni, S. (1991). Development and application of oligonucleotide probes for identification of *Lactococcus lactis* subsp. *cremoris*. *Appl. Environ. Microbiol.* 57**,** 1313-1318.

Salama, M. S., Sandine, W. E., and Giovannoni, S. J. (1993). Isolation of *Lactococcus lactis* subsp. *cremoris* from nature by colony hybridization with rRNA probes. *Appl. Environ. Microbiol.* 59**,** 3941-3945.

Schleifer, K., Kraus, J., Dvorak, C., Kilpper-Bälz, R., Collins, M., and Fischer, W. (1985). Transfer of *Streptococcus lactis* and Related Streptococci to the Genus Lactococcus gen. nov. *Syst. Appl. Microbiol.* 6**,** 183-195.

Segura, P. A., François, M., Gagnon, C., and Sauvé, S. (2009). Review of the occurrence of anti-infectives in contaminated wastewaters and natural and drinking waters. *Environ. Health Persp.* 117**,** 675.

Singh, T., Drake, M., and Cadwallader, K. (2003). Flavor of Cheddar cheese: a chemical and sensory perspective. *Comprehensive Reviews in Food Science and Food Safety.* 2**,** 166-189.

Smit, B., Engels, W., Wouters, J., and Smit, G. (2004). Diversity of L-leucine catabolism in various microorganisms involved in dairy fermentations, and identification of the rate-controlling step in the formation of the potent flavour component 3-methylbutanal. *Appl. Microbiol. Biotechnol.* 64**,** 396-402.

Staley, J. T., and Krieg, N. R. (1984). "Bacterial classification I. Classification of prokaryotic organisms: an overview" in *Bergey's manual of systematic bacteriology*, eds. S.T. Williams, M.E. Sharpe & J.G. Holt. (Baltimore: Williams & Wilkins), 1-4.

Taïbi, A., Dabour, N., Lamoureux, M., Roy, D., and Lapointe, G. (2011). Comparative transcriptome analysis of *Lactococcus lactis* subsp. *cremoris* strains under conditions simulating Cheddar cheese manufacture. *Int. J. Food Microbiol.* 146**,** 263-275.

Tamura, K., Dudley, J., Nei, M., and Kumar, S. (2007). MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* 24**,** 1596-1599.

Teixeira, L. M., Merquior, V. L. C., Vianni, M. D. C. E., Carvalho, M. D. G. S., El Fracalanzza, S., Steigerwalt, A. G., Brenner, D. J., and Facklam, R. R. (1996). Phenotypic and genotypic characterization of atypical *Lactococcus garvieae* strains isolated from water buffalos with subclinical mastitis and confirmation of *L. garvieae*

as a senior subjective synonym of *Enterococcus seriolicida*. *Int. J. Syst. Bacteriol.* 46**,** 664-668.

Thiele-Bruhn, S. (2003). Pharmaceutical antibiotic compounds in soils–a review. *J. Plant Nutr. Soil Sc.*166**,** 145-167.

Turner, K. M., and Feil, E. J. (2007). The secret life of the multilocus sequence type. *Int. J. Antimicrob. Ag.* 29**,** 129-135.

Van Kranenburg, R., Kleerebezem, M., Van Hylckama Vlieg, J., Ursing, B. M., Boekhorst, J., Smit, B. A., Ayad, E. H. E., Smit, G., and Siezen, R. J. (2002). Flavour formation from amino acids by lactic acid bacteria: predictions from genome sequence analysis. *Int. Dairy J.* 12**,** 111-121.

Wegmann, U., O'Connell-Motherway, M., Zomer, A., Buist, G., Shearman, C., Canchaya, C., Ventura, M., Goesmann, A., Gasson, M. J., and Kuipers, O. P. (2007). Complete genome sequence of the prototype lactic acid bacterium *Lactococcus lactis* subsp. *cremoris* MG1363. *J. Bacteriol.* 189**,** 3256-3270. DOI: 10.1128/JB.01768-06.

Wouters, J., Ayad, E. H., Hugenholtz, J., and Smit, G. (2002). Microbes from raw milk for fermented dairy products. *Int. Dairy J.* 12**,** 91-109. DOI: 10.1016/S0958-6946(01)00151-0.

Xu, H., Sun, Z., Liu, W., Yu, J., Song, Y., Lv, Q., Zhang, J., Shao, Y., Menghe, B., and 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. DOI: 10.3168/jds.2013-7738.

Yvon, M., Thirouin, S., Rijnen, L., Fromentier, D., and Gripon, J. C. (1997). An aminotransferase from *Lactococcus lactis* initiates conversion of amino acids to cheese flavor compounds. *Appl. Environ. Microbiol.* 63**,** 414-419.

Table 1. Oligonucleotide primer sequences used in PCR reactions for species/ subspecies identification and multi-locus sequence typing (MLST).

Primer	Sequence	Target region	Reference
Species identification			
16S_F[a]	5'-AGAGTTTGATCCTGGCTCAGG-3'	16S rRNA	Alander *et al.*, (1999)
16S_R	5'-ACGGCAACCTTGTTACGAGTT-3'	16S rRNA	
Subspecies identification			
LacF	5'-GTACTTGTACCGACTGGAT-3'	16S rRNA	Pu *et al*., (2002)
CreF	5'-GTGCTTGCACCGATTTGAA-3'	16S rRNA	
LacreR	5'-GGGATCATCTTTGAGTGAT-3'	16S rRNA	
Lhis5F	5'-CTTCGTTATGATTTTACA-3'	Histidine operon	Biemfohr *et al*., (1997)
Lhis6R	5'-AATATCAACAATTCCATG-3'	Histidine operon	
MLST[b]			
rpoA_F	5'-ATGATYGARTTTGAAAAACC-3'	RNA polymerase	Rademaker *et al*., (2007)
rpoA_R	5'-ACHGTRTTRATDCCDGCRCG-3'	RNA polymerase	
atpA_F	5'-TAYRTYGGKGAYGGDATYGC-3'	ATP synthase	
atpA_R	5'- CCRCGRTTHARYTTHGCYTG-3'	ATP synthase	
bcaT_F	5'-TTTKSHRTGCCDGTWGG-3'	BCAA aminotransferase	
bcaT_R	5'-GGWCCHACTTCYGTYTC-3'	BCAA aminotransferase	
pepN_F	5'-ATKTCTTAYGCWGAYRTYGT-3'	aminopeptidase N	
pepN_R	5'-TTKCTTCAAGSMAWGSCC-3'	aminopeptidase N	
pepX_F	5'-TTTGGGTTGAAAGTCCAGT-3'	X-prolyl peptidase	
pepX_R	5'-CCAAGAAGAAATTCCAGC-3'	X-prolyl peptidase	
pheS_F	5'-CAYCCNGCHCGYGAYATGC-3'	Phe tRNA synthetase	
pheS_R	5'-CCWARVCCRAARGCAAARCC-3'	Phe tRNA synthetase	
SSU rRNA_F	5'-GCGGCGTGCCTAATACATGC-3'	16S rRNA	
SSU rRNA_R	5'-ATCTACGCATTTCACCGCTAC-3'	16S rRNA	
Biogenic amine production			
HDC3_F	5'- GATGGTATTGTTTCKTATGA-3'	Histidine decarboxylase	Coton and Coton, 2005
HDC4_R	5' CAAACACCAGCATCTTC-3'		
TD5_F	'5- CAAATGGAAGAAGAAGTAGG-3'	Tyrosine decarboxylase	Coton *et al*., 2004
TD2_R	'5- ACATAGTCAACCATRTTGAA-3'		

[a] F= forward; R= reverse

[b] R= A/ G, Y= C/ T, M= A/ C, K= G/ T, S= C/ G, W= A/ T, H= A/ C/ T, B= C/ G/ T/, V= A/ C/ G, D= A/ G/ T, N= A/ C/ G/ T

Table 2. Phenotypic characterisation of dairy and non-dairy lactococcal strains.

	Genotype	Phenotype	Strain	Growth at		Arginine hydrolysis	Fermentation of		
				40 °C[a]	NaCl[a]		Maltose	Lactose	Citrate
Non-dairy									
Corn	*lactis*	*lactis*	DPC6853	+	+	+	+	+	–
Grass	*cremoris*	*lactis*	DPC6857	+	+	+	+	+	–
Grass	*cremoris*	*lactis*	DPC6858	+	+	+	+	+	–
Grass	*cremoris*	*lactis*	DPC6855	+	+	–	+	+	–
Grass	*cremoris*	*lactis*	DPC6859	+	+	+	+	+	–
Grass	*cremoris*	*lactis*	DPC6860	+	+	+	+	+	–
Bovine rumen	*cremoris*	*lactis*	DPC6856	+	+	+	+	+	–
Grass	*cremoris*	*lactis*	DPC6854	+	+	+	+	+	–
Dairy									
Cheese	*lactis*	*lactis*	DRC3	+	+	+	+	+	+
Starter culture	*lactis*	*lactis*	303	+	+	+	+	+	–
Cheese	*lactis*	*lactis*	ML8	+	+	+	–	+	–
Cheese	*lactis*	*lactis*	229	+	+	+	+	+	–
Raw milk	*cremoris*	*cremoris*	H88M1	+	–	–	+	+	–
Cheese	*cremoris*	*cremoris*	AM1	–	–	–	–	+	–
Cheese	*cremoris*	*cremoris*	310	–	–	–	–	+	–
Cheese	*cremoris*	*cremoris*	HP	–	–	–	–	+	–
Cheese	*cremoris*	*lactis*	SK1G	+	+	–	+	+	–
Type strains	*cremoris*	*cremoris*	SK11	–	–	–	–	+	–
	lactis	*lactis*	IL1403	+	+	+	+	–	–

[a] Growth at 40 °C and in 4% NaCl was monitored for 48 hours to determine subspecies phenotype.

Table 3. Differentiation of *L. lactis* subspecies by PCR, using the primers designed by Pu *et al*. (2002) and Biemfohr *et al*. (1997) targeting the 16S rRNA gene and the histidine biosynthesis operon respectively.

Target region:	16S rRNA gene		Histidine operon
	LacF[a]	CreF[a]	Amplicon size[b] (bp)
L. lactis DPC6853	+	–	934
L. lactis DPC6855	–	+	1149
L. lactis DPC6860	–	+	1149
L. lactis DPC6859	–	+	1149
L. lactis DPC6854	–	+	1149
L. lactis DPC6856	–	+	1149
L. lactis DPC6858	–	+	1149
L. lactis DPC6857	–	+	1149
L. lactis ML8	+	–	934
L. lactis DRC3	+	–	934
L. lactis 303	+	–	934
L. lactis 229	+	–	934
L. lactis AM1	–	+	1149
L. lactis H88M1	–	+	1149
L. lactis 310	–	+	1149
L. lactis SK1G	–	+	1149

[a] (+) denotes the formation of an amplicon; (–) denotes the formation of no amplicon

[b] subspecies *cremoris* generates a 1149 bp product; subspecies *lactis* generates a 934 bp product

Table 4. Genetic diversity for 26 *L. lactis* strains at seven loci.

Locus	Length (bp)			% GC Content	No. of polymorphic sites	d_n/d_s ratio	Unique alleles
	Gene	Amplicon	Sequence analysed				
atpA	1,503	1,141	393	41.52	32	0.008	10
bcaT	1,047	493	320	39.47	45	0.029	8
pepN	1,023	482	491	33.53	52	0.255	8
pepX	2,269	602	401	39.35	85	0.076	9
rpoA	939	721	531	40.19	13	0.092	12
pheS	2,533	618	361	41.51	36	0.010	9
16S rRNA	1,548	1,465	531	49.30	2	-	5

Table 5. ANI % calculated with the BLAST algorithm using the JSpecies software tool[a] (Richter and Rosselló-Móra, 2009).

	KW2*	SK11*	IL1403	TIFN2	UC509.9*	LD61	TIFN4	YF11	DPC6856*	DPC6860*	MG1363*	Dephy 1	JCM 5805	A12	A76*	KLDS 4.0325	NCDO2118	I-O1	TIFN3*	CV56	DPC6853	HP*	GE214*	NZ9000*	CNCM I-1631	KF147
*KW2		98.1	86.07	86.07	98.02	86.03	86.05	86.09	98.03	97.89	98.13	86.08	85.99	86	98.05	86.06	86.03	85.92	97.45	85.98	85.62	98.26	98.48	98.13	85.99	86.02
*SK11	98.23		87.18	87.21	99.27	87.12	87.03	86.5	97.92	97.45	97.85	86.45	86.9	87	99.36	86.85	86.33	86.54	98.96	87.12	86.5	99.02	98.26	97.86	86.79	86.53
IL1403	86.06	86.61		99.89	86.83	99.91	99.89	98.64	86.29	86.7	86.96	98.61	98.62	98.59	86.82	99.29	98.42	97.78	87.34	99.46	97.04	86.39	86.6	86.96	99.41	98.46
TIFN2	85.87	86.39	99.78		86.57	99.74	99.96	98.51	86.71	86.95	86.85	98.37	98.39	98.13	86.64	99.11	98.32	97.67	87.59	99.28	96.88	86.68	86.86	86.85	99.14	98.24
*UC509.9	98.13	99.27	87.35	87.34		87.33	87.11	86.82	97.87	97.46	97.67	86.72	87.1	87.24	99.35	86.96	86.61	86.72	99.18	87.47	86.91	98.97	98.14	97.67	87.04	86.74
LD61	85.82	86.3	99.71	99.63	86.56		99.63	98.33	86.68	87.02	86.95	98.34	98.34	98.03	86.59	99.01	98.23	97.53	87.8	99.15	96.53	86.78	86.99	86.95	99	98.12
TIFN4	85.82	86.37	99.8	99.94	86.54	99.72		98.32	86.68	86.97	86.86	98.31	98.36	98.03	86.6	99.09	98.27	97.63	87.67	99.25	96.79	86.65	86.85	86.86	98.97	98.22
YF11	85.87	86.05	98.6	98.51	86.42	98.49	98.52		86.02	86.37	86.58	98.32	98.22	98.35	86.28	98.38	98.34	97.74	87.09	98.51	96.8	86.3	86.47	86.6	98.4	98.25
*DPC6856	98.01	97.6	86.24	86.66	97.45	86.83	86.74	86.41		97.74	98.89	86.36	86.41	86.19	97.31	86.23	85.93	86.06	97.11	86.22	86.48	97.52	99.52	98.89	86.56	85.95
*DPC6860	98.02	97.4	86.49	86.88	97.4	87.04	86.85	86.82	98.14		98.98	86.72	86.61	86.52	97.41	86.63	86.39	86.23	97.01	86.47	86.68	97.5	98.57	98.99	86.62	86.35
*MG1363	98.05	97.62	86.89	87.03	97.56	87.12	87	86.77	98.87	98.56		86.86	86.77	86.96	97.6	86.77	86.57	86.69	97.22	86.77	86.5	97.73	99.23	100	86.77	86.66
Dephy 1	85.85	86.02	98.41	98.33	86.21	98.34	98.24	98.19	86.09	86.25	86.68		99.15	97.9	86.21	98.45	97.88	97.68	87.18	98.43	96.39	86.41	86.57	86.68	98.28	97.93
JCM 5805	85.76	86.22	98.48	98.48	86.3	98.47	98.47	98.29	86.15	86.34	86.55	99.46		98.19	86.36	98.17	98.12	97.84	87.26	98.48	96.72	86.58	86.6	86.55	98.4	98.07
A12	85.75	86.07	98.39	98.27	86.3	98.28	98.2	98.44	85.72	86.19	86.58	98.11	98.16		86.45	98.23	98.12	97.39	86.86	98.29	96.54	86.3	86.17	86.58	98.12	98.04
*A76	98.06	99.28	87.38	87.39	99.27	87.41	87.28	86.59	97.45	97.08	97.69	86.57	86.92	87.22		87.13	86.38	86.38	98.95	87.05	86.67	99.03	97.92	97.69	86.84	86.61
KLDS 4.0325	86.08	86.58	99.15	99.1	86.66	99.07	99.08	98.39	86.23	86.61	86.82	98.49	98.17	98.35	86.92		98.12	97.91	87.31	99.01	96.96	86.52	86.51	86.82	98.98	98.18
NCDO2118	85.92	86.11	98.25	98.2	86.38	98.18	98.14	98.34	85.76	86.29	86.5	98.02	98.1	98.1	86.19	98.1		97.74	87.01	98.23	96.93	86.06	86.29	86.52	98.13	99.82
I-O1	85.92	86.16	97.72	97.61	86.4	97.55	97.61	97.84	85.94	86.27	86.64	97.7	97.87	97.55	86.18	97.81	97.78		86.96	97.79	97.24	85.99	86.2	86.66	97.63	97.76
*TIFN3	97.31	98.55	87.66	88.02	98.68	88.31	88.12	87.19	96.83	96.16	97.24	87.15	87.45	87.26	98.62	87.46	86.77	86.84		87.56	87.66	98.07	97.31	97.24	87.45	86.87
CV56	86.04	86.45	99.43	99.26	86.81	99.27	99.26	98.51	86.22	86.52	86.76	98.6	98.59	98.35	86.5	99.05	98.3	97.88	87.18		97.06	86.36	86.45	86.78	99.64	98.29
DPC6853	85.88	86.16	97.53	97.59	86.4	97.54	97.57	97.81	86.48	86.73	86.74	97.38	97.34	97.5	86.3	97.6	97.67	97.6	87.15	97.6		86.5	86.67	86.74	97.5	97.64
*HP	98.3	99.12	86.45	86.68	99.02	86.9	86.66	86.53	97.76	97.31	97.94	86.54	86.6	86.38	99.11	86.55	86.31	86.15	98.82	86.4	86.25		98.27	97.94	86.68	86.39
*GE214	97.93	97.55	86.26	86.71	97.5	86.91	86.72	86.49	99.15	97.63	98.84	86.5	86.48	86.2	97.35	86.2	86.19	86	97.28	86.27	86.34	97.6		98.84	86.56	86.15
*NZ9000	98.02	97.7	86.93	87.06	97.58	87.14	87.05	86.77	98.83	98.53	100	86.84	86.71	86.92	97.63	86.81	86.53	86.7	97.23	86.81	86.33	97.78	99.21		86.83	86.61
CNCM I-1631	85.76	86.2	99.23	99.11	86.46	99.1	99.07	98.37	86.29	86.43	86.49	98.32	98.32	98.12	86.28	98.93	98.09	97.62	87.21	99.46	96.84	86.54	86.56	86.49		98.03
KF147	85.79	86.08	98.36	98.25	86.34	98.25	98.21	98.29	85.74	86.18	86.51	98.05	98.06	98.01	86.16	98.18	99.77	97.75	86.98	98.2	96.79	86.14	86.32	86.52	98.07	

[a] Subspecies *cremoris* genotype strains are identified by *

Values are coloured as follows: 100%; 95-99%; 90-94%; 85-89%.

Table 6. Tetranucleotide frequency correlation coefficient (TETRA) calculated using the JSpecies software tool[a] (Richter and Rosselló-Móra 2009) with values below the threshold TETRA value (<0.99) highlighted in green.

	KW2*	SK11*	IL1403	TIFN2	UC509.9*	LD61	TIFN4	YF11	DPC6856*	DPC6860*	MG1363*	Dephy 1	JCM 5805	A12	A76*	KLDS 4.0325	NCDO2118	I-O1	TIFN3*	CV56	DPC6853	HP*	GE214*	NZ9000*	CNCM I-1631	KF147
*KW2		0.997	0.993	0.993	0.996	0.993	0.993	0.993	0.996	0.998	0.998	0.994	0.994	0.993	0.996	0.994	0.994	0.994	0.990	0.995	0.994	0.998	0.997	0.998	0.994	0.994
*SK11	0.997		0.991	0.992	0.999	0.992	0.992	0.990	0.996	0.997	0.998	0.992	0.992	0.991	0.999	0.992	0.991	0.990	0.996	0.992	0.991	0.997	0.996	0.998	0.991	0.991
IL1403	0.993	0.991		0.999	0.991	0.998	0.999	0.998	0.991	0.994	0.994	0.998	0.998	0.997	0.991	0.998	0.999	0.999	0.986	0.999	0.998	0.994	0.993	0.994	0.999	0.999
TIFN2	0.993	0.992	0.999		0.991	0.999	1.000	0.998	0.993	0.995	0.995	0.999	0.999	0.997	0.992	0.998	0.998	0.997	0.990	0.998	0.998	0.995	0.995	0.995	0.999	0.998
*UC509.9	0.996	0.999	0.991	0.991		0.991	0.991	0.990	0.994	0.995	0.997	0.991	0.991	0.991	0.998	0.991	0.991	0.991	0.994	0.992	0.991	0.996	0.994	0.997	0.991	0.991
LD61	0.993	0.992	0.998	0.999	0.991		1.000	0.998	0.994	0.995	0.995	0.999	0.999	0.997	0.992	0.998	0.998	0.997	0.990	0.998	0.999	0.994	0.995	0.995	0.999	0.998
TIFN4	0.993	0.992	0.999	1.000	0.991	1.000		0.998	0.994	0.995	0.995	0.999	0.998	0.997	0.992	0.998	0.997	0.997	0.990	0.998	0.998	0.995	0.995	0.995	0.999	0.998
YF11	0.993	0.990	0.998	0.998	0.990	0.998	0.998		0.992	0.994	0.994	0.999	0.999	0.998	0.990	0.998	0.999	0.999	0.985	0.999	0.999	0.993	0.994	0.994	0.999	0.999
*DPC6856	0.996	0.996	0.991	0.993	0.994	0.994	0.994	0.992		0.998	0.998	0.994	0.994	0.992	0.996	0.994	0.991	0.991	0.994	0.992	0.993	0.997	0.999	0.998	0.994	0.992
*DPC6860	0.998	0.997	0.994	0.995	0.995	0.995	0.995	0.994	0.998		0.999	0.996	0.996	0.995	0.996	0.995	0.994	0.994	0.993	0.995	0.995	0.999	0.999	0.999	0.996	0.994
*MG1363	0.998	0.998	0.994	0.995	0.997	0.995	0.995	0.994	0.998	0.999		0.995	0.996	0.994	0.998	0.995	0.994	0.994	0.994	0.995	0.995	0.998	0.999	1.000	0.995	0.994
Dephy 1	0.994	0.992	0.998	0.999	0.991	0.999	0.999	0.999	0.994	0.996	0.995		0.999	0.998	0.991	0.999	0.998	0.998	0.988	0.999	0.999	0.995	0.995	0.995	0.999	0.998
JCM 5805	0.994	0.992	0.998	0.999	0.991	0.999	0.998	0.999	0.994	0.996	0.996	0.999		0.998	0.992	0.999	0.998	0.998	0.988	0.999	0.998	0.995	0.996	0.996	0.999	0.998
A12	0.993	0.991	0.997	0.997	0.991	0.997	0.997	0.998	0.992	0.995	0.994	0.998	0.998		0.991	0.998	0.997	0.997	0.988	0.997	0.997	0.995	0.994	0.994	0.998	0.997
*A76	0.996	0.999	0.991	0.992	0.998	0.992	0.992	0.990	0.996	0.996	0.998	0.991	0.992	0.991		0.992	0.990	0.990	0.997	0.991	0.991	0.996	0.996	0.998	0.991	0.990
KLDS 4.0325	0.994	0.992	0.998	0.998	0.991	0.998	0.998	0.998	0.994	0.995	0.995	0.999	0.999	0.998	0.992		0.998	0.997	0.989	0.998	0.998	0.994	0.996	0.995	0.999	0.998
NCDO2118	0.994	0.991	0.999	0.998	0.991	0.998	0.997	0.999	0.991	0.994	0.994	0.998	0.998	0.997	0.990	0.998		0.999	0.984	0.999	0.999	0.993	0.993	0.994	0.998	1.000
I-O1	0.994	0.990	0.999	0.997	0.991	0.997	0.997	0.999	0.991	0.994	0.994	0.998	0.998	0.997	0.990	0.997	0.999		0.984	0.999	0.999	0.993	0.993	0.994	0.998	0.999
*TIFN3	0.990	0.996	0.986	0.990	0.994	0.990	0.990	0.985	0.994	0.993	0.994	0.988	0.988	0.988	0.997	0.989	0.984	0.984		0.986	0.987	0.993	0.994	0.994	0.988	0.985
CV56	0.995	0.992	0.999	0.998	0.992	0.998	0.998	0.999	0.992	0.995	0.995	0.999	0.999	0.997	0.991	0.998	0.999	0.999	0.986		0.999	0.995	0.994	0.995	0.999	0.999
DPC6853	0.994	0.991	0.998	0.998	0.991	0.999	0.998	0.999	0.993	0.995	0.995	0.999	0.998	0.997	0.991	0.998	0.999	0.999	0.987	0.999		0.995	0.995	0.995	0.999	0.999
*HP	0.998	0.997	0.994	0.995	0.996	0.994	0.995	0.993	0.997	0.999	0.998	0.995	0.995	0.995	0.996	0.994	0.993	0.993	0.993	0.995	0.995		0.998	0.998	0.995	0.993
*GE214	0.997	0.996	0.993	0.995	0.994	0.995	0.995	0.994	0.999	0.999	0.999	0.995	0.996	0.994	0.996	0.996	0.993	0.993	0.994	0.994	0.995	0.998		0.999	0.995	0.994
*NZ9000	0.998	0.998	0.994	0.995	0.997	0.995	0.995	0.994	0.998	0.999	1.000	0.995	0.996	0.994	0.998	0.995	0.994	0.994	0.994	0.995	0.995	0.998	0.999		0.995	0.994
CNCM I-1631	0.994	0.991	0.999	0.999	0.991	0.999	0.999	0.999	0.994	0.996	0.995	0.999	0.999	0.998	0.991	0.999	0.998	0.998	0.988	0.999	0.999	0.995	0.995	0.995		0.999
KF147	0.994	0.991	0.999	0.998	0.991	0.998	0.998	0.999	0.992	0.994	0.994	0.998	0.998	0.997	0.990	0.998	1.000	0.999	0.985	0.999	0.999	0.993	0.994	0.994	0.999	

[a] Subspecies *cremoris* genotype strains are identified by *.

Table 7. Minimum inhibitory concentrations (MICs) for 16 antibiotics of non-dairy strains and three dairy lactococcal strains examined using VetMIC plates[a].

Origin/ strain	MIC (µg/ ml)															
	Gm	Km	Sm	Nm	Tc	Em	Cl	Cm	Am	PG	Va	Vi	Lz	Tm	Ci	Ri
Non-dairy																
DPC6857	2	16	32	8	0.5	0.06	0.25	4	0.25	0.25	<0.25	2	2	>64	4	8
DPC6858	2	8	16	8	0.25	0.06	0.25	4	0.25	0.25	<0.25	2	1	>64	2	8
DPC6855	1	8	16	4	0.25	0.06	0.06	2	<0.03	0.06	<0.25	0.5	1	>64	<0.25	8
DPC6859	2	16	32	8	0.5	0.06	0.12	4	0.25	025	<0.25	2	2	>64	4	16
DPC6860	1	8	16	32	0.25	0.06	0.25	2	0.12	0.25	<0.25	2	1	>64	8	8
DPC6856	2	16	32	16	1	0.06	0.06	2	0.12	0.12	0.5	1	2	>64	1	2
DPC6854	1	8	16	4	0.25	0.06	<0.03	2	0.06	0.06	0.5	1	2	>64	<0.25	2
DPC6853	4	16	32	32	64	0.06	0.25	4	0.5	0.25	0.5	2	1	>64	8	8
Dairy																
303	2	16	16	16	0.25	0.06	0.12	2	0.25	0.25	0.5	1	1	>64	8	8
SK1G	2	8	16	4	1	0.25	0.25	8	0.5	0.5	0.5	2	2	>64	8	64
DRC3	2	16	16	16	0.25	0.06	0.06	2	0.25	0.25	0.5	1	1	>64	2	4

[a] MIC assays were carried out in duplicate.

Antibiotic abbreviations: Gm, gentamicin; Km, kanamycin; Sm, streptomycin; Nm, neomycin; Tc, tetracycline; Em, erythromycin; Cl, clindamycin; Cm, chloramphenicol; Am, ampicillin; PG, penicillin G; Va, vancomycin; Vi, virginiamycin; Lz, linezolid; Tm, trimethoprim; Ci, ciprofloxacin; Ri, rifampicin

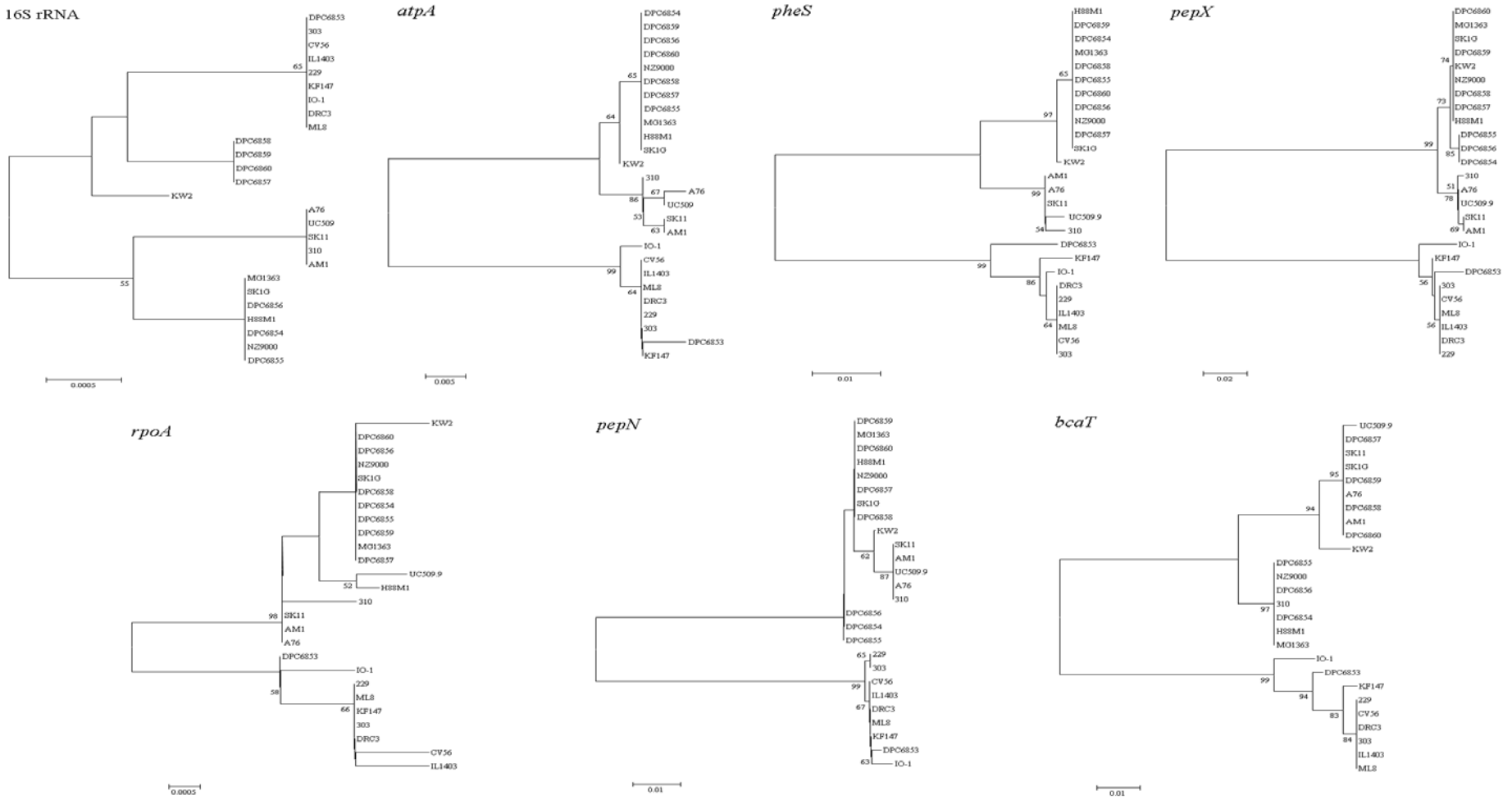


Fig. 1. Neighbour-joining cluster analysis of DNA segments of individual loci for dairy and non-dairy *Lactococcus* strains after 1,000 Bootstrap

replications with values ≥50 shown.

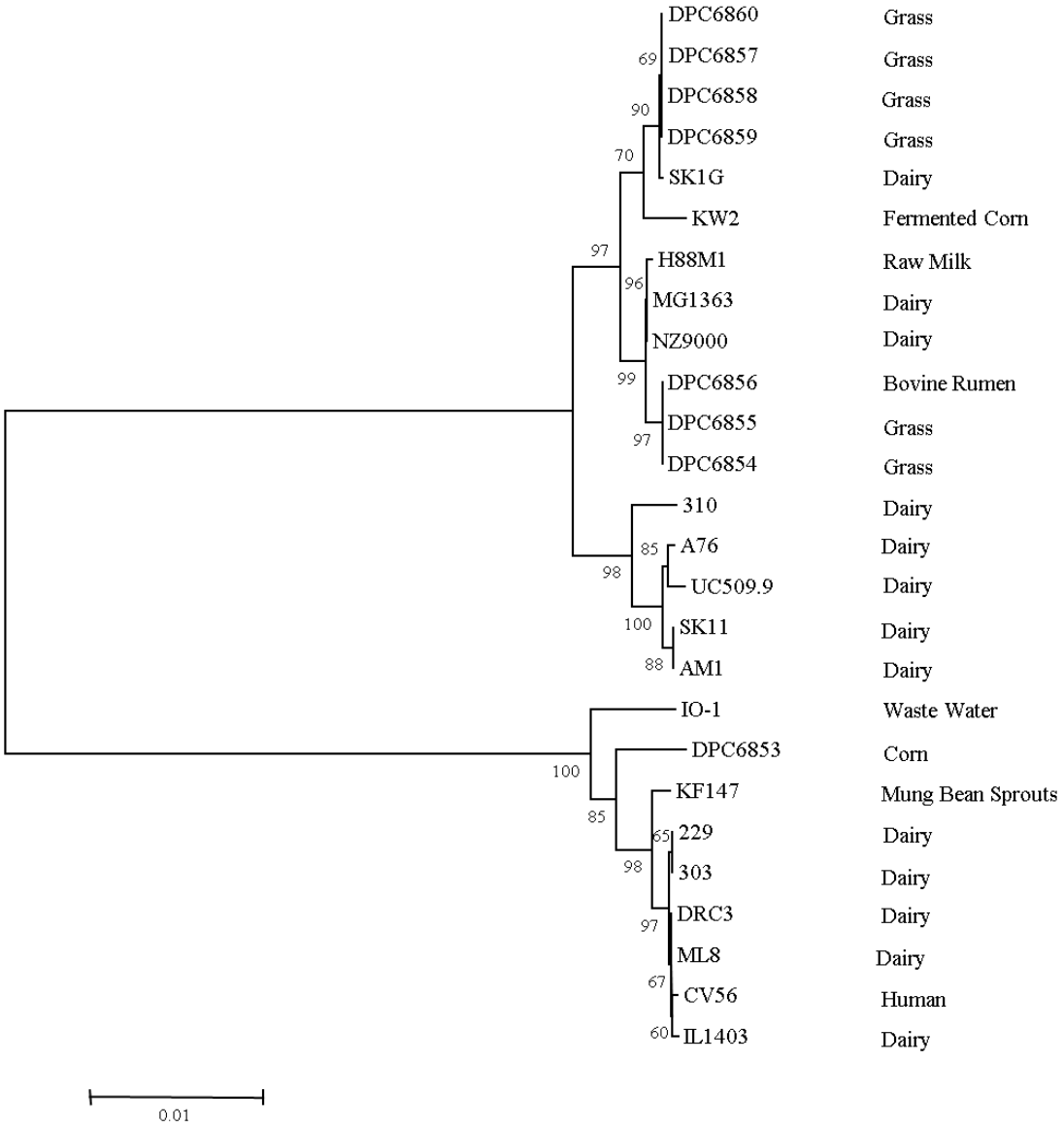


Fig. 2. Neighbour-joining cluster analysis of a composite data set for seven loci of dairy and non-dairy *Lactococcus* strains. Phylogenetic analysis was carried out with 1,000 bootstrap replications with bootstrap percentages ≥50 shown. The origin of each strain is indicated on the right hand side.

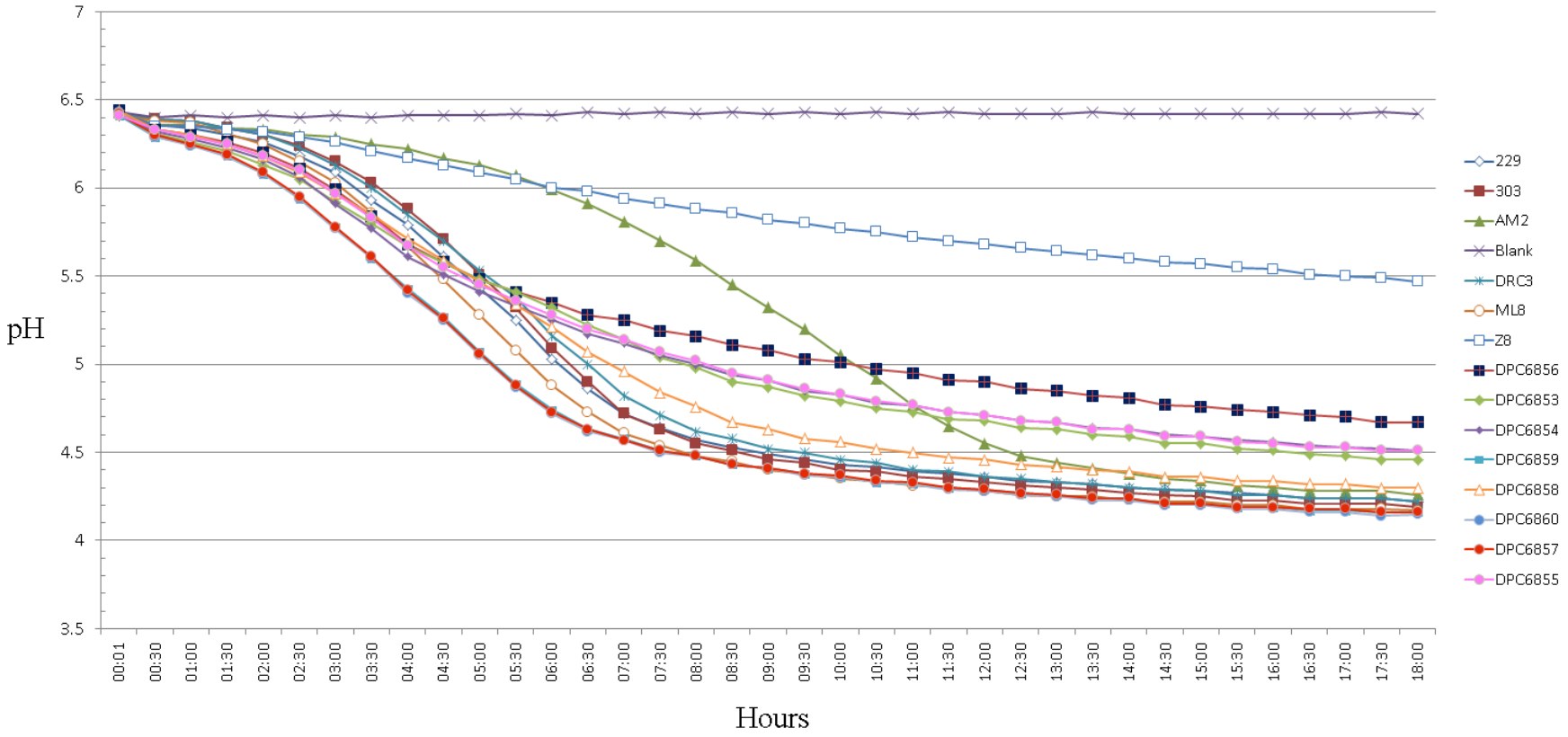


Fig. 3. Milk acidification profile of dairy and non-dairy strains in 10% reconstituted semi-skimmed milk (RSM).

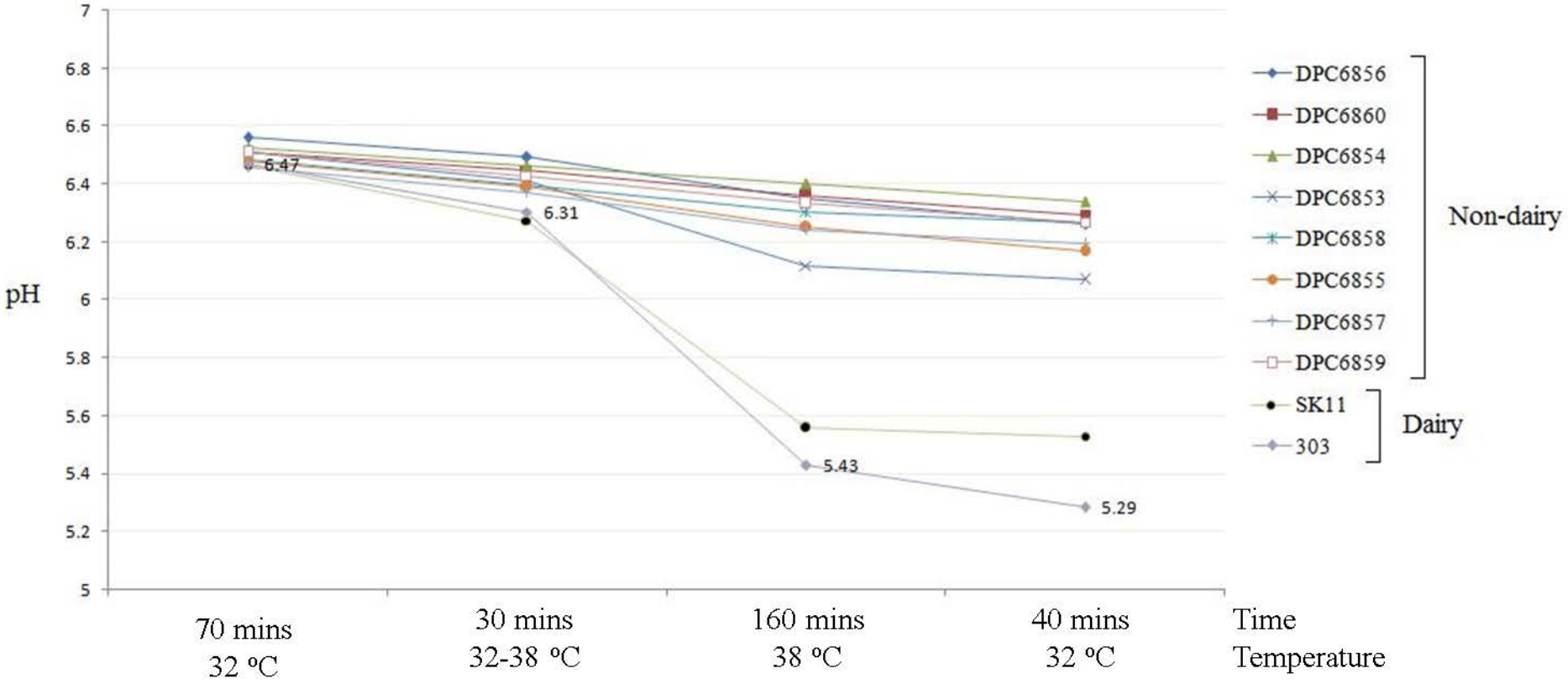


Fig. 4. Pearce test analysis of non-dairy strains. *L. lactis* 303 and SK11 were used as representative industrial dairy cultures. Results are the average of duplicate experiments.



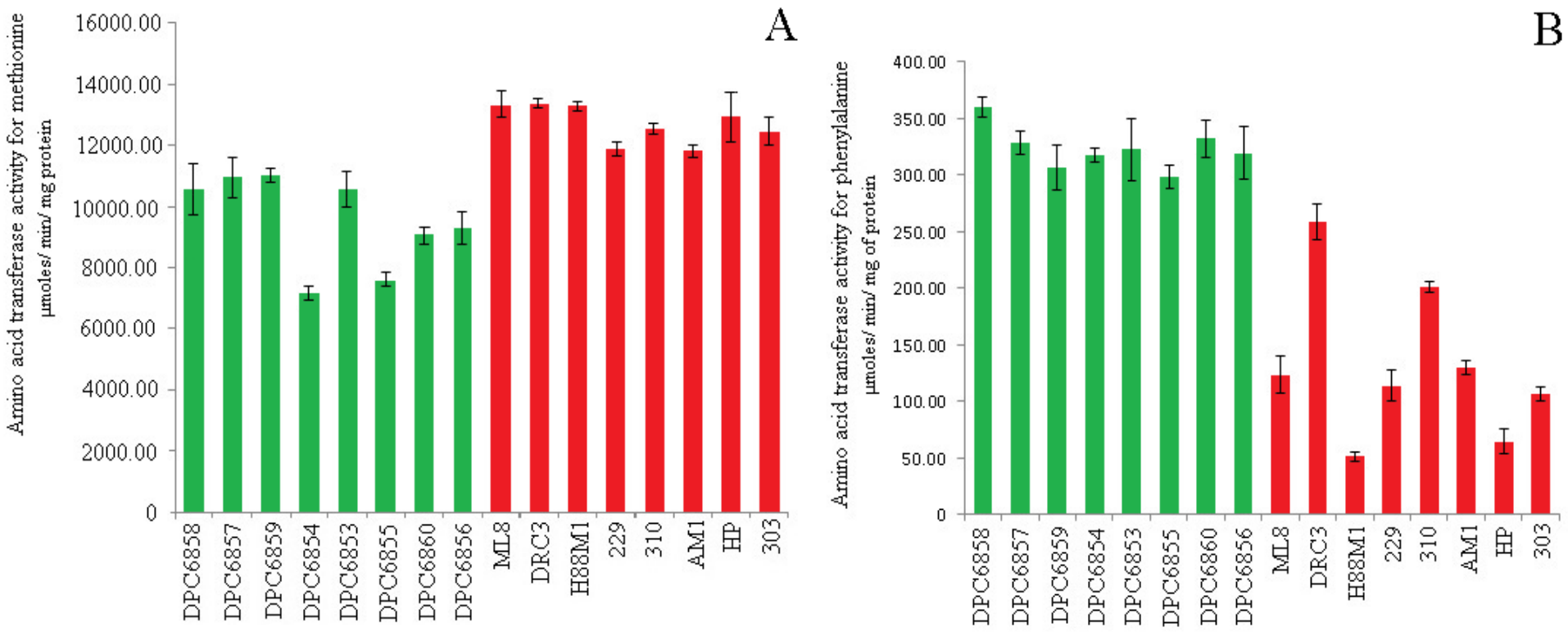


Figure 5. Amino acid transferase activities for methionine (A) and phenylalanine (B) with non-dairy strains in green and dairy strains in red. Results are the average of triplicate experiments.

Chapter 3

Assessment of wild non-dairy lactococcal strains for flavour diversification in a mini Gouda-type cheese model

All of this chapter has been published in Food Research International (2014), 62, 432-440. DOI:10.1016/j.foodres.2014.03.043

3.1. Abstract

Wild lactococci possess enhanced metabolic capabilities in comparison to industrial dairy strains, including increased amino acid-converting enzyme activities. A bank of *Lactococcus lactis* strains isolated from different non-dairy environments exhibited wider carbohydrate fermentation profiles in comparison to dairy lactococcal strains. In addition, these non-dairy lactococci had the ability to ferment lactose and produce diverse aroma profiles when grown in milk. Based on volatile analysis, five of these non-dairy strains were selected and investigated as adjuncts to diversify cheese flavour using a model mini Gouda cheese making process and were compared to a cheese manufactured with a commercial adjunct. In total, eight different cheeses were evaluated in duplicate and ripened for 14 days at 12 ^{o}C, followed by 8 ^{o}C for 84 days thereafter. Physicochemical analysis of cheeses was performed at day 14 and sensory evaluation at day 84. Viable counts, intracellular enzyme activity and indices of proteolysis were monitored over ripening. The ability of the non-dairy strains to survive, lyse, release intracellular enzymes and alter proteolysis was strain dependent. Some strains performed as well as the commercial adjunct in terms of secondary proteolysis although others were associated with bitterness and development of off-flavours and off-aromas. Attenuation of DPC6853 positively reduced its association with bitterness during ripening. It is evident that non-dairy strains have potential as adjuncts in semi-hard type cheeses, and could be harnessed to diversify flavour profiles in semi-hard cheese varieties.

3.2. Introduction

Flavour is a prime factor for consumer selection of cheeses, such as Gouda and Cheddar (Ross *et al*., 2000). The formation of cheese flavour arises from a series of chemical, biochemical and microbiological processes involving the break-down of lactose (glycolysis), fat (lipolysis) and protein (proteolysis) (McSweeney & Sousa, 2000). For the generation of cheese aroma in semi-hard cheeses, it is widely agreed that the proteolytic system and amino acid-converting enzymes possessed by the starter cultures are of pivotal importance (Smit *et al.*, 2005). *Lactococcus lactis* strains are the most widely used starter cultures in the production of cheese. A few select strains are predominantly used as primary starters due to the technological attributes that they possess (Marshall, 1991). Although this has reduced inconsistencies in quality, it has resulted in a reduction in the diversity of flavour amongst commercial cheeses. A secondary starter or adjunct culture is often added to impart specific properties and generate characteristic flavour profiles associated with specific cheeses (Johnson, 2001).

While *L. lactis* is normally associated with dairy fermentations, so-called 'wild strains' of this species can be isolated from raw milk and non-dairy environments (Centeno, *et al.*, 2002; Nomura *et al.*, 2006). The use of milk and cheese models has highlighted the potential of 'wild' *L. lactis* strains to produce unusual flavours and diverse volatile profiles, that could be applied to cheese making (Alemayehu *et al.*, 2014; Ayad *et al.*, 2000; Ayad *et al.*, 1999b; Morales *et al.*, 2003a). Genome sequence analysis has revealed non-dairy lactococcal strains to be more metabolically diverse than dairy cultures (van Kranenburg *et al.*, 2002). For example, *Lactococcus* strains from the wider environment have been shown to possess more active amino acid-converting enzymes than industrial dairy cultures and are auxotrophic for fewer amino

acids than their dairy counterparts (Ayad *et al.*, 1999a). In lactococcal starter cultures, the formation of aroma compounds from amino acids begins with transamination and the conversion of an amino acid to its corresponding α-keto acid (Gao & Steele, 1998; Rijnen *et al.*, 1999). This reaction is limited by the amount of α-ketogluturate produced by the microorganism which acts as an amino group acceptor (Tanous *et al*., 2002). Some studies have investigated the potential of lactic acid bacteria (LAB) strains from raw milk, raw milk cheeses and non-dairy niches to create cheeses of more diverse or intense flavour (Ayad *et al.*, 1999a; Centeno *et al.*, 2002). Sensory analysis of cheeses made using wild strains has resulted in the use of atypical sensory descriptors (Morales *et al.*, 2003b). Furthermore co-cultures of wild strains, leading to the completion of particular flavour formation pathways, may further diversify flavour profiles (Ayad *et al.*, 2001a).

The aim of this work was to evaluate the flavour-forming ability of selected non-dairy lactococci as adjunct cultures in cheese. A bank of non-dairy strains was assessed for important technological characteristics including growth in milk, tolerance to salt and temperature, proteolytic activity and the ability to ferment different carbohydrates. In addition, the generation of volatile compounds in milk was also assessed for each strain. Non-dairy strains which formed diverse aroma profiles in milk were selected for use as adjunct cultures in the production of mini Gouda-type cheese. The resultant cheeses were assessed for physicochemical, microbial, biochemical, and sensory characteristics, and compared to control cheeses that were produced using a commercially available adjunct derived from *Lactobacillus helveticus.*

3.3. Materials and methods

3.3.1. Bacterial strains and growth conditions

Lactococcus adjuncts used in this study were isolated from non-dairy sources. *L. lactis* ssp. *cremoris* strains DPC6854, DPC6857, DPC6858, DPC6859, DPC6860 and DPC6855 were isolated from grass samples sourced at the Animal and Grassland Research and Innovation Centre (Teagasc, Fermoy, Ireland). Strain DPC6853 was isolated from corn sourced at a local supermarket, while strain DPC6856 was isolated from a sample of bovine rumen content. Direct vat inoculation (DVI) cultures used in this study CHOOZIT™ Classic 111 (a mix of seven strains of *L. lactis* ssp. *lactis* and *L. lactis* ssp. *cremoris*) and CHOOZIT™ Flav54 (a single *Lactobacillus helveticus* strain) were supplied by Du Pont-Danisco (Dangé St. Romain, France). Non-commercial lactococcal strains were cultured in M17 media (Oxoid, Hampshire, England) containing 5 g/ L lactose monohydrate (LM17) (VWR, Leuven, Belgium). *Lb. helveticus* cultures were enumerated on MRS (Oxoid, Hampshire, England) media adjusted to pH 5.3 (mMRS). Prior to the production of mini Gouda-type cheeses, whole cultures were grown in sterile (121 ^{o}C for 5 mins) 10% reconstituted semi-skimmed milk (RSM) (Kerry Foods, Ireland) to 1 x 10^{8} CFU/ ml, pelletized in liquid nitrogen and stored at -80 ^{o}C. Non-dairy cultures underwent minimal sub-culturing in milk to maintain their natural metabolic characteristics.

L. lactis ssp. *lactis* DPC6853 was attenuated using a high pressure microfluidiser to increase the percentage of permeabilised or lysed cells within the population (Yarlagadda *et al.*, 2014a). This was achieved by growing the strain overnight in M17 media containing 0.5 g/ L lactose to 1 x 10^{8} CFU/ ml and passing 500 ml of culture through a Z chamber at 25,000 psi using a ME110EH Microfluidiser

(Microfluidics, Newton, MA, USA). The attenuated cells were pelletized in liquid nitrogen and stored at -80 ^{o}C until required.

3.3.2. Technological characterisation of strains

Lactococcus cultures were tested for their ability to grow at 8-10 ^{o}C, 30 ^{o}C and 42 ^{o}C for five days in M17 broth (Oxoid) containing 0.5% lactose (VWR, Belgium). The ability to grow at pH 9.2 and pH 9.6 and in the presence of 4 g/ 100 ml and 6.5 g/ 100 ml NaCl (Sigma Aldrich, Dublin, Ireland) was examined for five days at 30 ^{o}C in the same media. Casein breakdown was assessed using skim-milk agar plates containing 10% RSM (Kerry Foods, Ireland) and 1.5 g/ 100 ml bacteriological agar. The ability to ferment different carbohydrate substrates was assessed using the API CH50 system (bioMérieux, Montalieu-Vercieu, France) according to the manufacturer's instructions. Carbohydrate fermentation was examined at 24 and 48 hrs.

Following overnight growth of individual strains in 10% RSM, 0.5 ml cultures were added to 10 ml of RSM and incubated for 24 hrs at 30 ^{o}C. The production of volatile compounds in milk was assessed using a headspace (HS) solid phase micro-extraction (SPME) method as described by Yarlagadda *et al.* (2014b). For this analysis, the detector was set in electron ionisation mode at a mass/charge (m/z) range between 35 to 350 Daltons. Mass spectra for compounds were compared to the NIST 2005 spectra library to identify individual compounds and form an internally generated compound list. Compounds were each assigned a distinct quantification ion to ensure accuracy in identification and quantification. Compounds were semi-quantified by calculating the area under each individual peak. An auto-tune was performed at the start of each sample run to ensure that conditions remained constant

for each run. All analyses were carried out in duplicate and data was analysed using Unscrambler Software, version 9.7 (CAMO ASA, Trondheim, Norway).

3.3.3. Mini cheese production

Sixteen mini Gouda-type cheeses were manufactured with pasteurised milk using the vat in place (VIP) mini cheese system at Du Pont-Danisco (Dangé St. Romain, France) with and without the addition of adjunct cultures. Cheeses were produced under conditions designed to minimise the growth of non-starter lactic acid bacteria (NSLAB) which may impact the flavour profile of the cheese. All cheeses were made using Classic 111 as a starter culture. A control cheese containing the starter only was produced and each non-dairy adjunct strain was also compared to a commercial adjunct, Flav 54. Experimental and control cheeses were made from the same batch of milk in duplicate over two days. Milk was sourced from a local co-operative and pasteurised before production. For cheese making, 2 L milk was pre-heated at 32 ^{o}C and inoculated with the starter Classic 111 (Du Pont-Danisco, France) at 7 direct culture units (DCU) and non–dairy adjunct cultures were added at 35 g of pelleted frozen culture/ L. Flav 54 adjunct cultures were added at 10 g/ L. Carlina™ Calf chymosin (Du Pont-Danisco, France) was added (4 ml of 200 g/ L) 1 h after starter/adjunct addition at pH ~6.55. The coagulum was allowed to form for ~ 40 mins. The curd was cut and stirred for 5 mins followed by whey removal (700 ml) and addition of 600 ml of distilled water at 32 ^{o}C. Each vat was maintained at 36 ^{o}C for 20 mins with constant stirring. Cheeses were pre-pressed for 30 mins under a 1 kg weight in a circular mould. Samples were transferred for final pressing at 36 ^{o}C under 50 kPa. Cheese pH was monitored at constant intervals until a pH of 5.3 was reached when cheeses were placed in 150 ml of brine solution 300 g/ L NaCl

containing 0.5 g/ L Natamax™ anti-mould agent (Du Pont-Danisco, France) for 1 hr. Finally, cheeses (200 g) were allowed to dry at room temperature for 10 mins before vacuum packing. Cheeses were ripened at 12 ^{o}C for 14 days and samples taken with cheeses subsequently ripened at 8 ^{o}C thereafter until 84 days. Cheese sampling and analysis were performed at days 14, 42, and 84.

3.3.4. Cheese physicochemical and microbiological analysis

Cheese physicochemical analysis was performed at day 14. The moisture, NaCl, fat, pH and protein were determined as described elsewhere (Hickey, *et al.*, 2007). Analysis of the microbiological content of cheeses was carried out in duplicate at each sampling time. Cheese samples were mixed with tri-sodium citrate (2% wt./ vol.) at a 1:10 (wt./ vol.) dilution and homogenised for 5 mins using an IUL stomacher (Lennox Lab. supplies Ltd., Dublin, Ireland). Starter culture cell counts were enumerated on M17 agar (Oxoid, Hampshire, England) containing 0.5% wt./ vol. lactose monohydrate (VWR, Leuven, Belgium) after three days incubation at 30 ^{o}C. NSLAB were enumerated on LBS agar (Difco, Detroit, Michigan, USA) after 5 days incubation at 37 ^{o}C under anaerobic conditions.

3.3.5. Cheese extract preparation

A homogenised cheese extract was prepared at each sampling point to determine post proline aminopeptidase (PEPX), aminopeptidase N (PEPN) and lactate dehydrogenase (LDH) activities. To generate an extract, 2-3 g of cheese was added to a sterile stomacher bag and mixed with twice the volume of sterile 0.05 M potassium phosphate buffer (pH 7). The sample was stomached for 5 mins using an IUL stomacher (Lennox Lab. Supplies Ltd., Dublin, Ireland) followed by

centrifugation at 3,800 x *g* for 20 mins at 4 ^{o}C using a Sorvall 5U centrifuge (Unitech, Dublin, Ireland). The fat layer formed above the aqueous phase was removed and 2 ml of the solution was centrifuged at 13,000 x *g* for 5 mins using an Eppendorf 5417C centrifuge (VWR International Ltd., Dublin, Ireland). The resultant supernatant was diluted 1:2 with sterile distilled water. Samples were stored on ice and analysed rapidly to avoid loss of enzyme activity.

3.3.6. Enzyme activity assays

PEPX, PEPN and LDH activities were determined in cheese extracts over ripening at days 14, 42 and 84. Peptidase activities were measured as described in Kilcawley *et al.* (2012). Starter culture autolysis was determined in triplicate by measuring the release of LDH in cheese extract as described in Hickey *et al.* (2007).

3.3.7. Proteolysis

Proteolysis for each sample was analysed in duplicate at days 14, 42 and 84 using the methodology described in Hickey *et al.* (2007).

3.3.8. Sensory evaluation

Cheese samples (n=five) were analysed by sensory flash profiling (Yarlagadda *et al.*, 2014a) with an experienced sensory panel (n=10, five female, five male, ages 21–48 years) in booths conforming to international standards (ISO 8589: 2007). 6853 att and 6853 w cheeses were selected for sensory analysis to determine if attenuation had any effect on the flavour profile of the resultant cheese. 6855 and Flav54 cheeses were selected according to the relatively higher levels of secondary proteolysis observed in these cheeses throughout ripening. Selection criteria for panellists were

availability and motivation to participate on all days of the experiments and that assessors were regular Gouda cheese consumers. Assessors used the 23 sensory descriptors only (11 flavour; two texture; eight aroma) in appendix 3-1 which were compiled to represent the main sensory attributes found in the Gouda-type cheeses to be profiled (Yarlagadda *et al.*, 2014b). Samples were held at refrigeration temperatures (4 ºC), but were presented to the panel at ambient temperature (21 °C) and coded with a randomly selected three digit code. Each assessor was provided with deionised water and instructed to cleanse their palates between tastings. In addition each assessor was presented with five samples which were presented in duplicate over two sessions to overcome the effects of panel fatigue and asked to assess the attributes, according to a ten-point scale. The order of the presentation of all test samples was randomized to prevent first order and carryover effects.

3.3.9. Statistical analysis

Analysis of data generated from all assays excluding sensory analysis was carried out using SAS version 9.1.3 (SAS institute, Cary, NC, USA). To determine the significance of difference between mean values, the Tukey's significance difference test was carried out. The volatile and sensory data were analysed using ANOVA-partial least squares regression (APLSR). Both Partial Least Squares Regression (PLSR) and Principal Component Analysis (PCA) are bilinear modelling methods. Whereas PCA gives a visual summary of the main information in a multidimensional data platform, PLSR combines portions of principle component analysis and multiple regression techniques and is generally more suited to describing complex relationships and avoids problems encountered with other models such as inadmissible solutions and factor indeterminacy. The X-matrix was designed as 0/ 1

design variables for the Gouda cheese variants and the sensory data, for the Y-matrix design. The optimal number of components in the APLSR models presented was determined to be three principal components. Regression coefficients were analysed by Jack-knifing to derive significance indicators for the relationships determined in the quantitative APLSR (plot not shown), which is based on cross-validation and stability plots (Martens & Martens, 2000). The volatile and sensory data were analysed using Unscrambler Software, version 9.7 (CAMO ASA, Trondheim, Norway).

3.4. Results and discussion

3.4.1.1. Technological characterisation of strains

Eight lactococcal isolates from various non-dairy sources, including grass, vegetables and the bovine rumen, were tested for technological traits which may be important in cheese production and compared to eight dairy lactococci, commonly used in the dairy industry. In contrast to the dairy strains tested, non-dairy strains DPC6853, DPC6857, DPC6858, DPC6859 and DPC6860 were capable of growth at 45 ^{o}C (Table 1). DPC6853 and DPC6856 were capable of growth in the presence of 6.5% NaCl along with the dairy strains 303 and DRC3. Strains DPC6854, DPC6857 and DPC6858 were the only non-dairy strains capable of growth at 8-10 ^{o}C, while all dairy strains, except AM1, were able to grow at this temperature. Previously, wild strains were found to be more tolerant of higher temperatures and salt concentrations than industrial starters but this did not extend beyond 40 ^{o}C and 4% NaCl, respectively (Ayad *et al.*, 2000). The authors suggest that such strains may have applications for cheeses that contain relatively high salt concentrations and are cooked at relatively high temperatures. Growth of non-dairy cultures on skim-milk agar plates revealed that all isolates were capable of hydrolysing casein, as shown by zones of clearing around colonies. The ability of the strains to ferment different carbohydrates is shown in Table 1, which presents the total number of sugars that dairy and non-dairy strains were capable of fermenting. All non-dairy strains were capable of fermenting lactose. In comparison to the non-dairy isolates, known dairy cultures displayed a reduced capacity to ferment sugars, metabolising on average 12 out of 49 sugars, compared to 17 out of 49 for the non-dairy strains. Of these strains HP, 310 and AM1 were only able to utilise between six-seven sugars. The capacity of *L. lactis* to ferment different sugars is correlated to how domesticated a particular strain is to the dairy environment

(Kelly *et al.*, 2010). Genome sequence analysis of two non-dairy *Lactococcus* isolates established that these strains possessed a much higher number of genes involved in carbohydrate utilisation in comparison to dairy lactococcal genomes (Siezen *et al.*, 2008).

3.4.1.2. Production of volatile compounds

The volatile profiles generated by the eight non-dairy lactococci grown in milk were compared to that of a number of *Lactococcus* strains used extensively in cheese making by HS SPME GC-MS. *L. lactis* H88M1, isolated from raw milk, was also included to determine whether this strain formed volatile profiles similar to non-dairy isolates. Fig. 1 is a PCA bi-plot of 39 compounds (PC1: 34%; PC2 16%) associated with milk fermentations using individual cultures of dairy and non-dairy strains. Milk fermentations using strains DPC6853, DPC6854, DPC6855, and DPC6856 formed volatile profiles which were clearly separated from dairy strains and the other non-dairy strains. Samples inoculated with DPC6857, DPC6858, DPC6859, DPC6860 and ML8, AM1 and 303 were more associated with 2-butanone (etheric), octanoic acid (sweaty, body odour), benzaldehyde (almond) and 2-pentanone (orange peel, sweet, fruity) (Curioni & Bosset, 2002; Singh *et al.*, 2003). Samples containing DPC6854 and DPC6855 were strongly associated with the compounds 3-methylbutanal and 2-methylpropanal. Aldehydes can be formed in cheese via the transamination of amino acids or by Strecker degradation (Curioni & Bosset, 2002) and compounds such as 3-methylbutanal and 2-methylpropanal are associated with dark chocolate, malt aromas (Singh *et al.*, 2003). 3-methylbutanal in particular, is implicated as being an important compound in many cheese varieties such as Proosdij, a Gouda-type cheese (Smit *et al.*, 2005). Numerous metabolic pathways are

involved in the formation of alcohols in cheese, such as the degradation of amino acids and methylketone reduction (Molimard & Spinnler, 1996). 2-heptanol and 2-nononal were most strongly associated with DPC6856 and have green, earthy, herbaceous aromas (Curioni & Bosset, 2002; Singh *et al.*, 2003). Due to the similar volatile profiles produced by DPC6857, DPC6858, DPC6859 and DPC6860, only DPC6860 was selected for use as an adjunct in mini cheese trials as a representative of these non-dairy isolates. Both DPC6855 and DPC6854 were selected for use in cheese models due to their strong association with 2, 3 butandione (buttery), 2-methyl propanal (pungent) and pentanal (almond) which dairy strains were not strongly associated with the dairy strains examined (Singh *et al*., 2006). Pearce test analysis (O'Donovan *et al*., 1996) showed that non-dairy isolates were incapable of acid production during cheese processing temperatures in milk (Chapter 2) and as a result, these strains (DPC6853, DPC6854, DPC6855, DPC6856 and DPC6860) were used as adjunct cultures in combination with an industrial dairy starter.

3.4.2. Cheese physicochemical analysis

Based on the results of volatile analysis (Fig. 1), which indicated their potential ability to produce varied and diverse flavour profiles, strains DPC6853, DPC6854, DPC6855, DPC6856 and DPC6860 were chosen for cheese trials. The performance of these five non-dairy strains as flavour adjuncts were evaluated using a mini Gouda-type model cheese. The adjunct CHOOZIT™ Flav54 *Lb. helveticus* was included as a comparison strain due to its high aminopeptidase activity and wide use as an adjunct in commercial semi-hard cheese production. Use of the wild adjuncts *L. lactis* DPC6854 and *L. lactis* DPC6856 appeared to hinder the acidification of the starter and additional processing time was required for these cheeses (data not shown).

Although no antimicrobial activity was identified previously (see Chapter 2), DPC6854 and DPC6856 may produce a compound that may inhibit the growth of the starter. The physicochemical and quality parameters for cheeses are shown in appendix 3-2. No differences ($P < 0.05$) were observed for the main parameters of moisture, protein and fat, or moisture in not fat substances (MNFS), salt in moisture (S/ M) and fat in dry matter (FDM), highlighting the consistency of the make procedure on the VIP system. The attenuated *L. lactis* DPC6853 cheeses had a lower pH of 5.07 ($P < 0.05$) in comparison to all the other cheeses. This was attributed to the increased moisture content of 6853 att cheeses, ~2% higher, resulting in a higher retention of lactose in the curd and the increased production of lactic acid. The pH vales of the remaining cheeses were higher at 5.30 (control), 5.24 (Flav54), 5.28 (6853 w), 5.21 (6860), 5.29 (6854), 5.28 (6855) and 5.25 (6856).

3.4.3. Stability of non-dairy cultures over ripening

Bacteria were enumerated on LM17, mMRS and LBS agar at days 14, 42 and 84. Lactococcal viable cell counts on LM17 (Table 2) at day 14 ranged from 10.61 log CFU/ g in the control cheeses to 8.91 log CFU/ g in 6855 cheeses. With the exception of 6860 cheeses, viable cell numbers declined in all cheeses up to day 84. Viable cell counts for 6860 cheeses increased from 9.75 log CFU/ g at day 14 to 11.05 log CFU/ g at day 42 and decreased sharply until day 84. Following cheese production, starter counts are generally between 10^8-10^9 CFU/ g of cheese which decline over ripening, in contrast to NSLAB counts which increase significantly from low numbers at the start of ripening (Settanni & Moschetti, 2010). Ayad *et al*. (2001b) investigated the population dynamics of Gouda-type cheeses made using a non-dairy strain as an adjunct culture. They observed that cell numbers increased by more than

three orders of magnitude for *L. lactis* B1156, isolated from grass, from an initial inoculum of 0.25% over ripening. In a previous study, comparative analysis between plant and dairy *Lactococcus* strains revealed non-dairy strains capable of utilising a broader array of complex and simple carbohydrates (Siezen *et al.*, 2008). It is conceivable that the unusual increased growth in 6860 cheeses from days 14 to 42 may be due to the utilisation of carbohydrates released by cell autolysis as observed by Adamberg *et al.* (2005) in addition to a relatively high inoculum. Although a strain-to-strain variation in terms of autolysis of non-dairy strains was observed, our findings agree largely with previous work, that wild lactococci possess poor autolytic activity and remained relatively stable over ripening (Ayad *et al.*, 2001b). In the Flav54 cheeses, viable cell counts on mMRS agar showed that the *Lb. helveticus* adjunct underwent the largest decline in viable cell numbers over ripening ($P < 0.05$), indicating a high degree of autolysis. No NSLAB were detected at any sampling point in any cheeses.

Numerous investigations have reported a positive correlation between starter autolysis and the release of flavour precursors in cheese (O'Donovan *et al.*, 1996; Wilkinson *et al.*, 1994). Cell lysis was monitored by measuring LDH activity in cheese extracts (Table 3). LDH activity was lower in 6853 w and 6860 cheeses($P < 0.05$), but higher in 6855 ($P < 0.05$) cheeses compared to controls at day 14. LDH activity in 6853 w cheeses remained lower at day 42 ($P < 0.05$), but not at day 84. At day 84, 6856 cheeses had lower LDH activity ($P < 0.05$) compared to control cheeses. A higher level of mean LDH activity was observed in 6855 cheeses in comparison to all other cheeses ($P < 0.05$). This result corresponds well with viable cell count data, where the 6855 strain possessed the lowest counts over ripening (Table 2). In 6853 att cheeses, LDH levels increased from day 14 to 42 ($P < 0.05$) as the intracellular

contents were released into the cheese matrix. From day 42 to 84, as the number of cells declined, LDH levels returned to levels in 6853 w cheeses (those of cheeses containing whole DPC6853 cultures) as there was no statistical difference at this stage of ripening ($P < 0.05$).

3.4.4. Peptidase activity

PEPN and PEPX activities in cheese extracts over ripening are shown in Table 4. PEPX activity increased ($P < 0.05$) in 6853 w cheeses over ripening and from day 42 to 84 for the control, 6854 and 6855 cheeses. Cheeses 6856 and 6853 att showed no significant difference in PEPX activity over ripening. Cheeses 6855 had a higher level of PEPX activity at day 84 in comparison to all other cheeses, again corresponding well with viable count and LDH data. A higher PEPX ($P < 0.05$) level was observed in control and 6860 cheeses at 42 days compared to 6854, 6856 and 6853 att cheeses. 6855 cheeses possessed higher ($P < 0.05$) levels of PEPN activity at day 84 in comparison to Flav54, 6853 att, 6853 w, 6860 and 6856 cheeses.

No difference ($P > 0.05$) in PEPX and PEPN activities were detected in cheeses made with whole (6853 w) or attenuated DPC6853 cultures (6853 att) over days 14, 42 and 84. A reduction in PEPX and PEPN activities over time was observed in 6853 att cheeses whereas an increase ($P < 0.05$) in PEPX activity for 6853 w cheeses after day 42. At day 84, PEPX and PEPN activities were higher in 6853 w cheeses compared to 6853 att cheeses. This may be explained by higher cell counts in the 6853 w cheeses over this period as evident from Table 2. It is likely that more intracellular peptidases were lost to the whey in the production of 6853 att cheeses as has been previously shown (Kilcawley *et al.*, 2012). This may have been exacerbated in Gouda cheese by the additional washing step during production. It should also be

noted that levels of released enzymes in cheese extract do not necessarily correspond to results determined directly from a cell free extract of the strain due to strain-dependent autolysis and external factors within the cheese environment during production and ripening, and the poor stability of enzymes at differing salt concentrations and pH levels (Sheehan *et al.*, 2006).

3.4.5. Proteolysis

3.4.5.1. pH 4.6 Soluble Nitrogen

The amount of pH 4.6 soluble nitrogen (SN) as % of total N (Fig. 2.), indicative of primary proteolysis, increased ($P < 0.001$) over ripening for all cheeses. At day 14, 6853 att cheeses possessed higher ($P < 0.05$) levels of pH 4.6 SN in comparison to the control cheeses. At day 42, pH 4.6 SN levels in Flav54, 6853 w and 6853 att cheeses were higher ($P < 0.05$) than in 6854 cheeses, but not in the control cheeses. No difference ($P > 0.05$) was observed in pH 4.6 SN at day 84. *L. lactis* possesses an extra-cellular bound proteinase, lactocepin, which aids in the breakdown of caseins to facilitate the up-take of smaller peptides by the cell. However, the added chymosin and indigenous plasmin present in the milk, are primarily responsible for the initial breakdown of casein and the release of small to medium sized peptides (McSweeney, 2004). Higher moisture levels in 6853 att cheeses most likely account for the observed difference in pH 4.6 SN at day 14 due to the higher chymosin retention (increased activity) and greater water activity, all of which favour greater enzymatic action on casein (Choisy *et al.*, 2000).

3.4.5.2. Free Amino Acid Analysis

The concentration of total free amino acids (TFAA) increased ($P < 0.001$) over ripening for all cheeses as shown in Fig. 3. Flav54 cheeses possessed higher ($P < 0.05$) mean levels of TFAA compared with 6856, 6853 att and 6860 cheeses. At all sampling points over ripening, levels of TFAA were highest in cheeses made with Flav54 compared to those made with added lactococci. Flav54 is composed of a *Lb. helveticus* strain, which as a species are known to contain greater levels of intracellular peptidase activity than *Lactococcus* strains (Kenny *et al*., 2006), which likely accounts for the high levels of TFAA found in this study and in found in previous studies (Hannon *et al.*, 2007). Mean TFAA levels for control and 6855 cheeses were comparable ($P < 0.05$) with those of Flav54 cheeses. It was observed that 6855 cheeses contained the highest concentration of leucine at the end of ripening, but not isoleucine or valine (data not shown). Casein hydrolysis accounts for the largest proportion of the TFAA concentration in cheese (Fox, 1989). However the composition of TFAA is also influenced by the cheese microflora; by the utilisation of specific amino acids or the secretion of amino acids into the cheese matrix, or both (Sousa, Ardö, & McSweeney, 2001).

3.4.6. Sensory evaluation

A number of cheeses were selected for sensory analysis, based on data from primary and secondary proteolysis analysis. 6853 att and 6853 w cheeses were selected to determine the effect of attenuation by microfluidisation of the culture on the flavour of the cheese. Fig. 4 shows the PCA that best discriminated the sensory attributes of the selected cheeses (PC1: 46%; PC2 23%). The results of sensory analysis are shown in appendix 3-3. Control cheeses were clearly separated from test cheeses and were associated with butterscotch aroma and sweaty/ sour and pungent

aroma. Flav54 cheeses were clearly distinguishable from all other cheeses and were associated with sweet, buttery, creamy, fruity and cheesey attributes. The lactococcal adjuncts were associated away from the control and Flav54 cheeses and more associated with differences in sensorial textural attributes, nutty flavours, nutty aromas, bitterness, astringency and off flavours and off-aromas. These results correspond with previous findings whereby wild lactococcal cultures formed off-flavours in semi-hard cheeses, but may be important flavour notes in other hard cheese varieties (Ayad *et al.*, 2000; Barbieri *et al.*, 1994; Molimard & Spinnler, 1996).

The sensory PCA results compare with the initial work of volatile profiles generated in milk. Non-dairy cultures generated aroma profiles distinct from industrial dairy cultures while a similar trend was observed in the PCA bi-plot of sensory analysis; grouping DPC6853 and DPC6855 cheeses away from control and Flav54 cheeses. DPC6853 cheeses were identified as being strongly associated with off-aromas and off-flavours, and in milk this strain was found to produce increased concentrations of butanoic acid (sweaty, cheesy), propanoic acid (pungent) and hexanoic acid (goat-like) in milk (Singh *et al.*, 2003). Milk samples cultured with DPC6855, were associated with higher levels of pentanal, 3-methylbutanal and 2-methylpropanal which have been shown to contribute to a nutty flavour in aged cheddar cheeses (Avsar *et al.*, 2004). However the aroma of a particular compound in food may be different to that detected by analytical instruments due to associations with other compounds and within the matrix itself (Drake & Civille, 2003). Similarly, at higher concentrations, compounds which are positively associated with flavour may be associated with off-flavours. Therefore although growth in milk offers an insight into the ability of the strains to produce particular aromas, without the use of

techniques such as gas chromatography olfactory (GC-O), we may only speculate on the relationship between aroma profiles in milk and cheese flavour in this study.

Free amino acids (FAAs) contribute to the basic taste of cheese but are also converted into potent flavour compounds via microbial conversion. Thus, the relative concentration of specific amino acids can impact on the flavour profile of a cheese. 6853 att and 6860 cheeses possessed the highest relative concentration of branched-chain amino acids (BCAA) (leucine, isoleucine and valine) and aromatic amino acids (AAA) respectively as a percentage of total FAA at day 84 (Table 5). The ratio of BCAA to AAA in cheeses differs according to the starter system used. The use of attenuated 6853 culture altered the relative amounts of BCAA and AAA in comparison to 6853 w cheeses. Control and Flav54 cheeses were not correlated with off flavours or off aromas and possessed a ratio of 1: 0.64 (BCAA: AAA). 6853 w cheeses possessed the highest concentration of AAA to BCAA of cheeses that underwent sensory analysis (Table 5). Both 6853 w and 6853 att cheeses were closely associated with off flavours and off aromas. However, 6853 att cheeses possessed a similar BCAA: AAA ratio as observed in Flav54 and control cheeses. This may be explained by the increased conversion of available FAAs to flavour compounds by attenuated cells whereby increased autolysis leads to an increase in the formation of flavour compounds from phenylalanine as observed by Bourdat-Deschamps *et al.* (2004). The adjunct used also affected the concentration of methionine at day 84, the conversion products of which are considered very important for cheese flavour (Smit *et al.*, 2000). The PCA bi-plot (Fig. 4) clearly separated 6853 att and 6853 w cheeses. Attenuation led to 6853 att cheeses becoming more associated with nutty flavours and aromas, altered textural sensory attributes, but also strengthened its association with off flavours and off aromas and astringency, all of which have been associated with

the use of wild strains in cheese production (Morales *et al.*, 2003b). However, the attenuation of 6853 decreased its association with bitterness determined by sensory analysis. Based on differences in moisture content between cheeses, we may only speculate that microfluidisation may be a useful method to maximise the metabolic and enzymatic complement of wild lactococcal cultures for cheese flavour development. To further examine the potential role of microfluidisation in contributing to bitterness reduction, model cheese production should be performed using cultures examined in this chapter where composition can be consistent between cheeses.

3.5. Conclusions

In general, the inclusion of non-dairy lactococci in the starter system did not adversely affect cheese composition. The ability of non-dairy strains to survive, lyse, release intracellular peptidases and enhance proteolysis during Gouda cheese ripening was strain dependent. Strain 6855 showed the greatest capacity to increase secondary proteolysis relative to other non-dairy cultures. The perception of bitterness was greater in cheeses produced using whole 6853 cultures in comparison to cheeses made with attenuated DPC6853 (6853 att). Of the non-dairy strains evaluated for sensory profiling, all had greater association with nutty flavours and aromas, and bitterness, which corresponded with previous research in semi-hard cheese production. We can conclude that particular non-dairy lactococci survive in Gouda-type cheese over ripening and impact on the proteolytic and sensory characteristics of the resultant cheese without adversely impacting on cheese composition. However further work in terms of dosage, synergy between strains, or attenuation is required in order to maximise their potential to positively alter cheese flavour.

3.6. Acknowledgements

Special thanks to Du Pont Danisco and Jonathon Goodwins for facilitating the use of the VIP mini-cheese system and the production of cheeses.

Thanks to Paula O'Connor (Teagasc) for conducting free amino acid analysis.

Thanks to Anil Yarlagadda (Teagasc) for assistance in cheese analysis.

Thanks to Maurice O'Sullivan for conducting sensory analysis.

3.7. References

Adamberg, K., Antonsson, M., Vogensen, F. K., Nielsen, E. W., Kask, S., Møller, P. L., & Ardö, Y. (2005). Fermentation of carbohydrates from cheese sources by non-starter lactic acid bacteria isolated from semi-hard Danish cheese. *Int. Dairy J.* 15, 873-882.

Alemayehu, D., Hannon, J. A., McAuliffe, O., & Ross, R. P. (2014). Characterization of plant-derived lactococci on the basis of their volatile compounds profile when grown in milk. *Int. J. Food Microbiol.* 172, 57-61.

Avsar, Y., Karagul-Yuceer, Y., Drake, M., Singh, T., Yoon, Y., & Cadwallader, K. (2004). Characterization of nutty flavor in Cheddar cheese. *J. Dairy Sci. 87*, 1999-2010.

Ayad, E. H. E., Verheul, A., de Jong, C., Wouters, J. T. M., & Smit, G. (1999a). Flavour forming abilities and amino acid requirements of *Lactococcus lactis* strains isolated from artisanal and non-dairy origin. *Int. Dairy J.* 9, 725-735.

Ayad, E. H. E., Verheul, A., Engels, W., Wouters, J., & Smit, G. (2001a). Enhanced flavour formation by combination of selected lactococci from industrial and artisanal origin with focus on completion of a metabolic pathway. *J. Appl. Microbiol.* 90, 59-67.

Ayad, E. H. E., Verheul, A., Wouters, J., & Smit, G. (2000). Application of wild starter cultures for flavour development in pilot plant cheese making. *Int. Dairy J.* 10, 169-179.

Ayad, E. H. E., Verheul, A., de Jong, C., Wouters, J. T. M., & Smit, G. (1999b). Flavour forming abilities and amino acid requirements of *Lactococcus lactis* strains isolated from artisanal and non-dairy origin. *Int. Dairy J.* 9, 725-735.

Ayad, E. H. E., Verheul, A., Wouters, J. T. M., & Smit, G. (2001b). Population dynamics of lactococci from industrial, artisanal and non-dairy origins in defined strain starters for Gouda-type cheese. *Int. Dairy J.* 11, 51-61.

Barbieri, G., Bolzoni, L., Careri, M., Mangia, A., Parolari, G., Spagnoli, S., & Virgili, R. (1994). Study of the volatile fraction of Parmesan cheese. *J. Agr. Food Chem.* 42, 1170-1176.

Bourdat-Deschamps, M., Le Bars, D., Yvon, M., & Chapot-Chartier, M. P. (2004). Autolysis of *Lactococcus lactis* AM2 stimulates the formation of certain aroma compounds from amino acids in a cheese model. *Int. Dairy J.* 14, 791-800.

Centeno, J., Tomillo, F., Fernández-García, E., Gaya, P., & Nuñez, M. (2002). Effect of wild strains of *Lactococcus lactis* on the volatile profile and the sensory characteristics of ewes' raw milk cheese. *J. Dairy Sci.* 85, 3164-3172.

Choisy, C., Desmazeaud, M., Gripon, J., Lamberet, G., Lenoir, J., Eck, A., & Gillis, J. (2000). The biochemistry of ripening. In *Cheesemaking: from science to quality assurance* (pp. 82-151).

Curioni, P. M. G., & Bosset, J. O. (2002). Key odorants in various cheese types as determined by gas chromatography-olfactometry. *Int. Dairy J.* 12, 959-984.

Drake, M. A., & Civille, G. (2003). Flavor lexicons. *Comprehensive Reviews in Food Science and Food Safety.* 2, 33-40.

Fox, P. (1989). Proteolysis during cheese manufacture and ripening. *J. Dairy Sci.* 72, 1379-1400.

Gao, S., & Steele, J. L. (1998). Purification and characterization of oligomeric species of an aromatic amino acid aminotransferase from *Lactococcus lactis* subsp. *lactis* S3. *J. Food Biochem.* 22, 197-211.

Hannon, J., Kilcawley, K., Wilkinson, M., Delahunty, C., & Beresford, T. (2007). Flavour precursor development in Cheddar cheese due to lactococcal starters and the presence and lysis of *Lactobacillus helveticus*. *Int. Dairy J.* 17, 316-327.

Hickey, D. K., Kilcawley, K. N., Beresford, T. P., Sheehan, E. M., & Wilkinson, M. G. (2007). Starter strain related effects on the biochemical and sensory properties of Cheddar cheese. *J. Dairy Res.* 74, 9-17.

Johnson, M. E. (2001). *Cheese products*. New York: Marcel Dekker.

Kelly, Ward, L. J., & Leahy, S. C. (2010). Chromosomal diversity in *Lactococcus lactis* and the origin of dairy starter cultures. *Genome Biol. Evol.* 2, 729-744.

Kenny, O., Fitzgerald, R. J., O'Cuinn, G., Beresford, T., & Jordan, K. (2006). Autolysis of selected *Lactobacillus helveticus* adjunct strains during Cheddar cheese ripening. *Int. Dairy J.* 16, 797-804.

Kilcawley, K., Nongonierma, A., Hannon, J., Doolan, I., & Wilkinson, M. (2012). Evaluation of commercial enzyme systems to accelerate Cheddar cheese ripening. *Int. Dairy J.* 26, 50-57.

Marshall, V. M. (1991). Inoculated ecosystem in a milk environment. *J. Appl. Bacteriol.* 73 (S21), 127S-135S.

Martens, H., & Martens, M. (2000). Modified Jack-knife estimation of parameter uncertainty in bilinear modelling by partial least squares regression (PLSR). *Food Qual. Prefer.* 11, 5-16.

McSweeney. (2004). Biochemistry of cheese ripening: Introduction and overview. In *Cheese: Chemistry, Physics and Microbiology*. eds P. Fox, P. McSweeney, T. Cogan & a. P. Guinee, (Vol. 1, pp. 347-360). London, UK: Academic Press.

McSweeney, & Sousa, M. J. (2000). Biochemical pathways for the production of flavour compounds in cheeses during ripening: A review. *Le Lait.* 80, 293-324.

Molimard, P., & Spinnler, H. (1996). Review: compounds involved in the flavor of surface mold-ripened cheeses: origins and properties. *J. Dairy Sci.* 79, 169-184.

Morales, P., Fernández-García, E., Gaya, P., & Nuñez, M. (2003a). Formation of volatile compounds by wild *Lactococcus lactis* strains isolated from raw ewes' milk cheese. *Int. Dairy J.* 13, 201-209.

Morales, P., Fernández-García, E., Gaya, P., & Nuñez, M. (2003b). Formation of volatile compounds by wild *Lactococcus lactis* strains isolated from raw ewes' milk cheese. *Int. Dairy J.* 13, 201-209.

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.

O'Donovan, C. M., Wilkinson, M. G., Guinee, T. P., & Fox, P. F. (1996). An investigation of the autolytic properties of three lactococcal strains during cheese ripening. *Int. Dairy J.* 6, 1149-1165.

Rijnen, L., Delacroix-Buchet, A., Demaizieres, D., Le Quéré, J., Gripon, J., & Yvon, M. (1999). Inactivation of lactococcal aromatic aminotransferase prevents the formation of floral aroma compounds from aromatic amino acids in semi-hard cheese. *Int. Dairy J.* , 877-885.

Ross, R. P., Stanton, C., Hill, C., Fitzgerald, G. F., & Coffey, A. (2000). Novel cultures for cheese improvement. *Trends Food Sci. Tech.* 11, 96-104.

Settanni, L., & Moschetti, G. (2010). Non-starter lactic acid bacteria used to improve cheese quality and provide health benefits. *Food Microbiol.* 27, 691-697.

Sheehan, A., Cuinn, G., Fitzgerald, R., & Wilkinson, M. (2006). Proteolytic enzyme activities in Cheddar cheese juice made using lactococcal starters of differing autolytic properties. *J. Appl. Microbiol.* 100, 893-901.

Siezen, R. J., Starrenburg, M. J., Boekhorst, J., Renckens, B., Molenaar, D., & van Hylckama Vlieg, J. E. (2008). Genome-scale genotype-phenotype matching of two

Lactococcus lactis isolates from plants identifies mechanisms of adaptation to the plant niche. *Appl. Environ. Microbiol.* 74, 424-436.

Singh, T., Drake, M., & Cadwallader, K. (2003). Flavor of Cheddar cheese: a chemical and sensory perspective. *Comprehensive Reviews in Food Science and Food Safety.* 2, 166-189.

Smit, G., Smit, B. A., & Engels, W. J. M. (2005). Flavour formation by lactic acid bacteria and biochemical flavour profiling of cheese products. *FEMS Microbiol. Rev.* 29, 591-610.

Smit, G., Verheul, A., van Kranenburg, R., Ayad, E., Siezen, R., & Engels, W. (2000). Cheese flavour development by enzymatic conversions of peptides and amino acids. *Food Res. Int.* 33, 153-160.

Sousa, M., Ardö, Y., & McSweeney, P. (2001). Advances in the study of proteolysis during cheese ripening. *Int. Dairy J.* 11 , 327-345.

Tanous, C., Kieronczyk, A., Helinck, S., Chambellon, E., & Yvon, M. (2002). Glutamate dehydrogenase activity: a major criterion for the selection of flavour-producing lactic acid bacteria strains. *Antonie van Leeuwenhoek.* 82, 271-278.

van Kranenburg, R., Kleerebezem, M., van Hylckama Vlieg, J., Ursing, B. M., Boekhorst, J., Smit, B. A., Ayad, E. H. E., Smit, G., & Siezen, R. J. (2002). Flavour formation from amino acids by lactic acid bacteria: predictions from genome sequence analysis. *Int. Dairy J.* 12, 111-121.

Wilkinson, M. G., Guinee, T. P., O'Callaghan, D. M., & Fox, P. F. (1994). Autolysis and proteolysis in different strains of starter bacteria during Cheddar cheese ripening. *J. Dairy Res.* 61, 249-262.

Yarlagadda, A., Wilkinson, M., O'Sullivan, M., & Kilcawley, K. (2014a). Utilisation of microfluidisation to enhance enzymatic and metabolic potential of lactococcal strains as adjuncts in Gouda type cheese. *Int. Dairy J.* 66, 1-10.

Yarlagadda, A. B., Wilkinson, M. G., Ryan, S. P., Doolan, I. A., O'Sullivan, M. G., & Kilcawley, K. N. (2014b). Utilisation of a cell-free extract of lactic acid bacteria entrapped in yeast to enhance flavour development in Cheddar cheese. *Int. J. Dairy Technol.* 67, 21-30.

Table 1. Technological characterisation of dairy and non-dairy lactococcal strains.

Strain	Origin	Species/ subspecies (ssp)	Proteolytic activity[a]	Carbohydrate fermentation[b]	Growth characteristics				
					4% NaCl	6.5% NaCl	8 -10 ºC	42 ºC	45 ºC
Non-dairy strains									
DPC6853	Corn	*L. lactis* ssp. *lactis*	+	22	+	+	–	+	+
DPC6854	Grass	*L. lactis* ssp. *cremoris*	+	16	+	–	+	–	–
DPC6855	Grass	*L. lactis* ssp. *cremoris*	+	17	+	–	–	–	–
DPC6856	Bovine rumen	*L. lactis* ssp. *cremoris*	+/–	11	+	+	–	–	–
DPC6857	Grass	*L. lactis* ssp. *cremoris*	+	18	+	–	+	+	+
DPC6858	Grass	*L. lactis* ssp. *cremoris*	+	16	+	–	+	+	+
DPC6859	Grass	*L. lactis* ssp. *cremoris*	+	18	+	–	–	+	+
DPC6860	Grass	*L. lactis* ssp. *cremoris*	+	21	+	–	–	+	+
Dairy Strains									
DRC3	Cheese	*L. lactis* ssp. *lactis biovar. diacetylactis*	+	17	+	+	+	+	–
303	Cheese starter	*L. lactis* ssp. *lactis*	+	18	+	+	+	+	–
ML8	Cheese starter	*L. lactis* ssp. *lactis*	+	14	+	–	+	–	–
229	Cheese starter	*L. lactis* ssp. *lactis*	+	19	+	–	+	+	–
H88M1	Raw milk	*L. lactis* ssp. *cremoris*	+	14	+	–	+	+	–
AM1	Cheese	*L. lactis* ssp. *cremoris*	+	7	–	–	–	–	–
310	Cheese	*L. lactis* ssp. *cremoris*	+	6	+	–	+	–	–
HP	Cheese	*L. lactis* ssp. *cremoris*	+/–	6	–	–	+	–	–

[a] Proteolysis was determined by clearing on skim milk agar plates.

[b] Total number of sugars fermented by strains using the API CH50 test strips (bioMérieux, Montalieu-Vercieu, France).

Table 2. Viable cell counts[1] (Log CFU/ g cheese) on LM17 and mMRS media in cheeses at days (d)14, 42 and 84.

Cheese	Log CFU/ g cheese		
	d 14	d 42	d 84
control[a]	10.61	9.11	7.34
Flav54[b]	9.09	7.80	7.54
Flav54[c]	6.82	5.18	4.19
6853 att	9.42	8.01	6.90
6853 w	9.72	8.17	7.45
6860	9.75	11.05	7.78
6854	9.26	7.50	6.92
6855	8.61	6.76	6.56
6856	9.48	7.45	7.04

[a] Mean values for two batches of each cheese analysed in duplicate.

[b] Viable cell counts for *Lactococcus* cultures on LM17 agar.

[c] Viable cell counts for *Lb. helveticus* adjunct on mMRS agar.

Table 3. Lactate dehydrogenase (LDH units /ml of cheese extract) activity in cheeses at days (d) 14, 42 and 84.

Cheese	LDH		
	d 14	d 42	d 84
control	15.78^{cd}	23.85^{bc}	19.30^{bc}
Flav54[1]	19.22^{bc}	21.20^{bc}	12.52^{bcd}
Flav54[2]	31.12^{a}	55.18^{a}	59.24^{a}
6853 att	13.70^{cd}	20.53^{bc}	10.29^{bcd}
6853 w	8.53^{d}	7.71^{d}	8.78^{cd}
6860	9.23^{d}	13.17^{cd}	9.46^{bcd}
6854	13.97^{cd}	16.38^{bc}	12.30^{bcd}
6855	23.71^{b}	26.71^{b}	20.28^{b}
6856	13.21^{cd}	16.81^{bcd}	7.51^{c}

$^{a-d}$Means within a column with different superscripts differ significantly ($P < 0.05$).

[1] LDH activity for starter *Lactococcus* cultures.

[2] LDH activity for *Lb. helveticus* adjunct only.

Table 4. Post proline dipeptidyl aminopeptidase activity[1] (PEPX) and aminopeptidase N activity[1] (PEPN) activities in cheese extracts at days (d) 14, 42 and 84.

Cheese	PEPX			PEPN		
	d 14	d 42	d 84	d 14	d 42	d 84
control	27.92^{a}	60.10^{a}	119.79^{ab}	1.03^{ab}	1.66^{bc}	7.04^{a}
Flav54	35.42^{a}	55.76^{abc}	78.34^{b}	2.03^{a}	3.56^{a}	1.96^{abc}
6853 att	18.28^{a}	32.06^{cd}	21.23^{b}	0.96^{ab}	1.20^{bc}	0.38^{c}
6853 w	12.46^{a}	49.95^{abc}	89.64^{b}	0.58^{b}	1.03^{bc}	3.65^{abc}
6860	9.21^{a}	67.21^{a}	67.87^{b}	0.64^{b}	0.60^{c}	3.31^{abc}
6854	28.69^{a}	35.50^{bcd}	139.40^{ab}	1.57^{ab}	1.67^{bc}	5.68^{ab}
6855	36.43^{a}	56.55^{ab}	241.01^{a}	2.04^{a}	2.10^{b}	6.18^{ab}
6856	19.77^{a}	18.02^{d}	32.64^{b}	1.63^{ab}	0.92^{c}	1.65^{bc}

$^{a-d}$Means within a column with different superscripts differ significantly ($P < 0.05$).

[1]Expressed as relative fluorescent units (RFU)/ mg protein.

Table 5. Relative concentrations[1] of branch chained amino acids (BCAA), aromatic amino acids (AAA), methionine (*Met*) and total free amino acids (TFAA) in cheeses at day 84.

Cheese	BCAA	AAA	*Met*	BCAA: AAA[2]	TFAA[1]
control	646.41^{ab}	433.41^{b}	34.73^{bc}	1: 0.64	5468.1^{abc}
Flav54	985.53^{a}	640.59^{a}	69.87^{a}	1: 0.64	8317.3^{a}
6853 att	544.14^{b}	349.45^{b}	20.11^{c}	1: 0.64	3881.6^{c}
6853 w	525.64^{b}	403.85^{ab}	26.40^{bc}	1: 0.76	4358.4b^{c}
6860	423.06^{b}	393.47^{ab}	24.11^{bc}	1: 0.93	3774.5^{c}
6854	709.02^{ab}	381.03^{ab}	25.09^{bc}	1: 0.53	5229.0^{bc}
6855	964.31^{a}	485.59^{ab}	46.47^{ab}	1: 0.50	7238.4^{ab}
6856	622.57^{b}	328.14^{b}	20.94^{c}	1: 0.52	4482.5^{bc}

[1]TFAA are expressed as µg/ g cheese.

[2]The ratio of branch chained amino acids to aromatic amino acids.

[a-d] Means within a column with different superscripts differ significantly ($P < 0.05$)

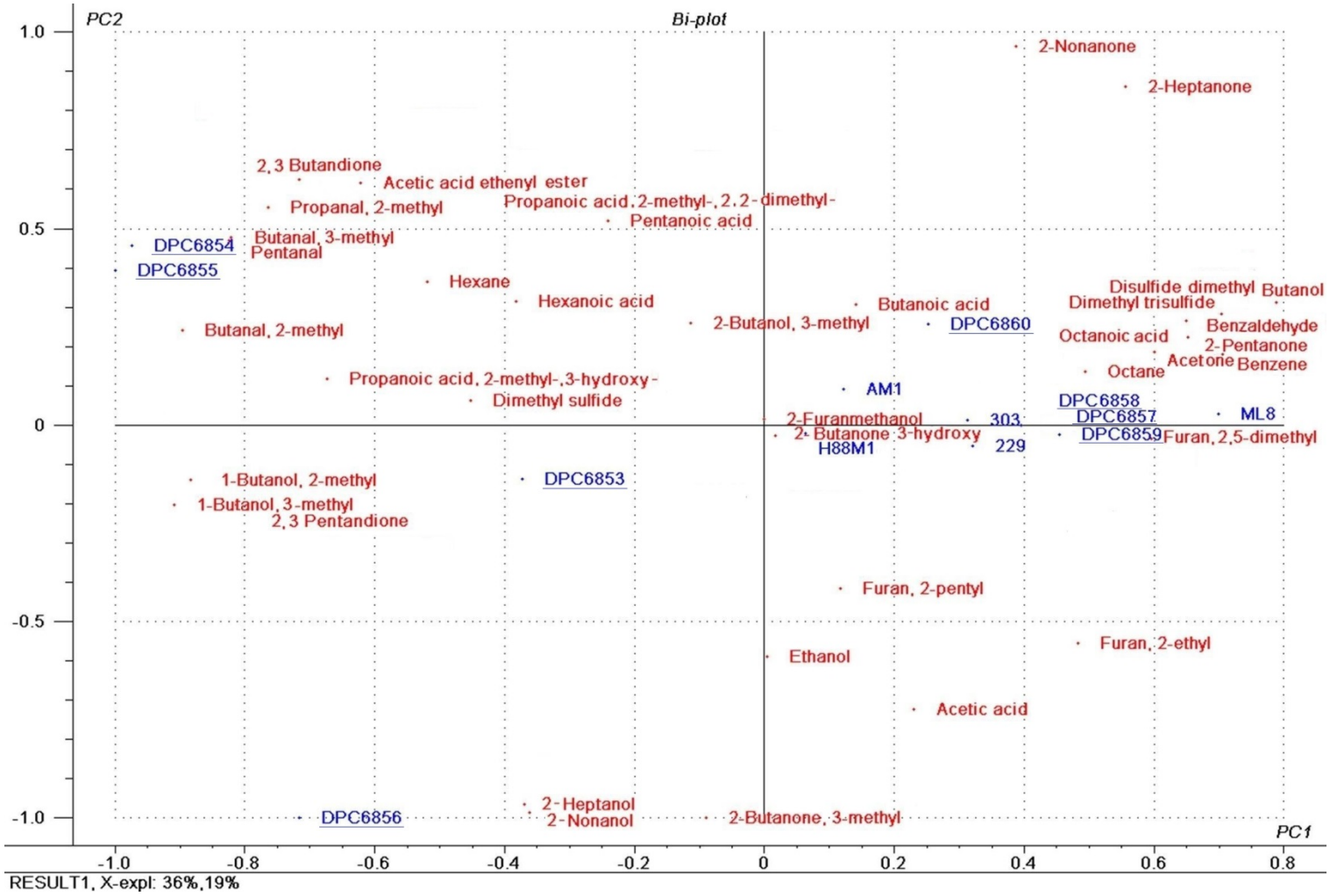


Fig. 1. Principal component bi-plot of volatile compounds identified in cultured milk samples. Non-dairy strains are underlined.

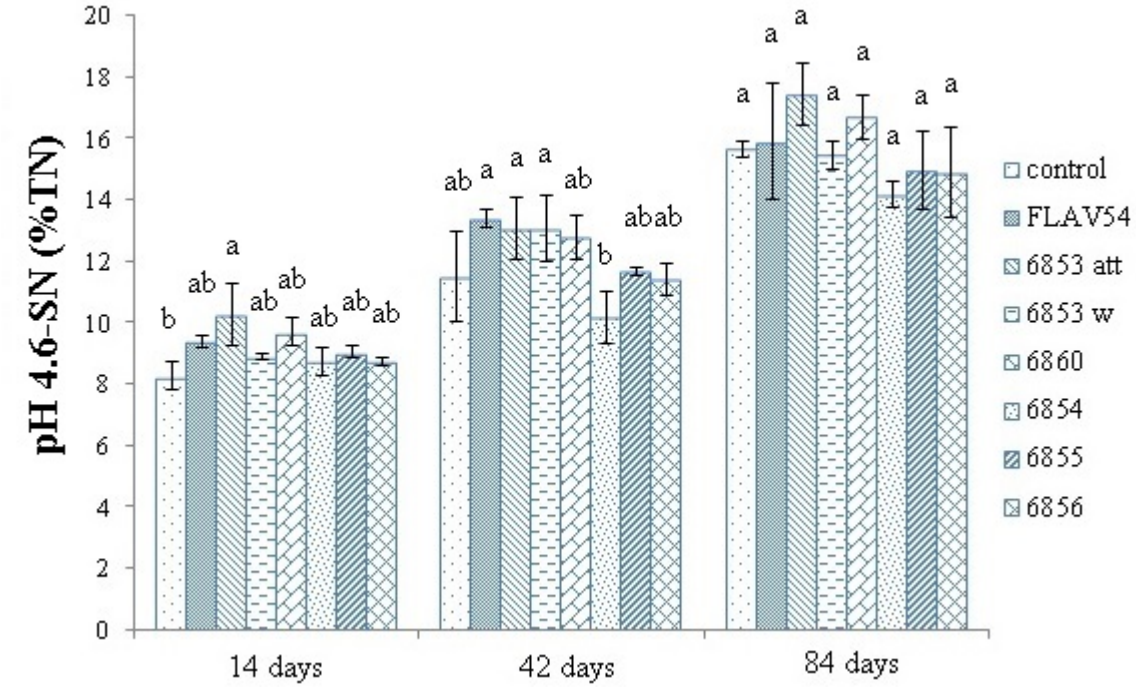


Fig. 2. Average values for pH 4.6 soluble-nitrogen (SN) as % of total nitrogen for cheeses at days 14, 42 and 84. Means within a sampling point with different superscripts (a-d) differ significantly ($P < 0.05$).

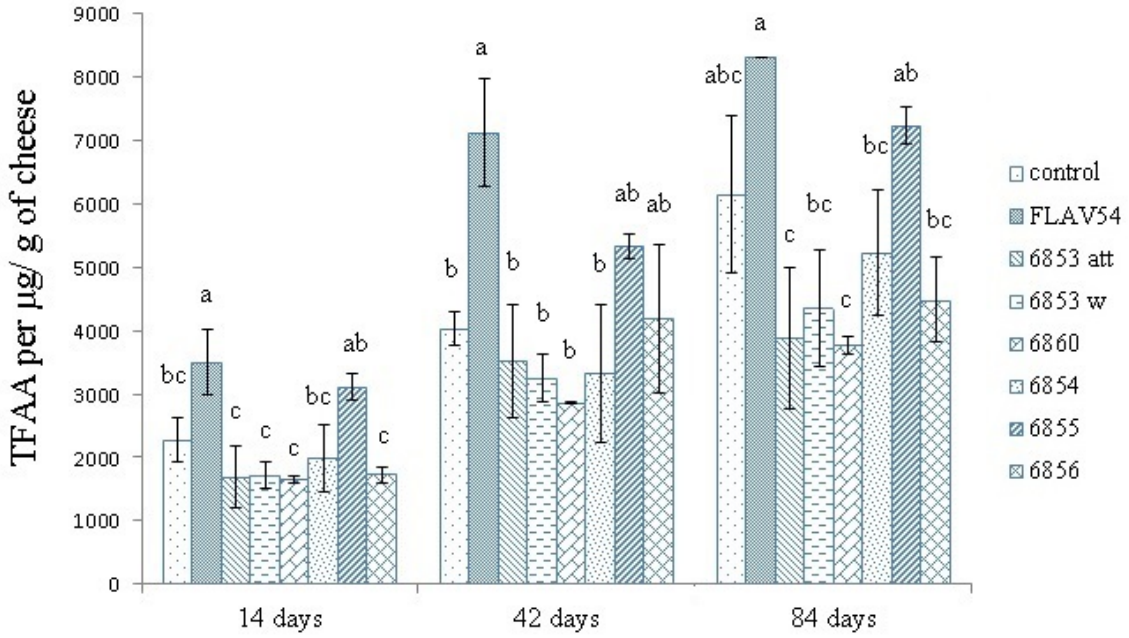


Fig. 3. Total free amino acids (TFAA) concentrations at days 14, 42 and 84 in mini Gouda-type cheeses. The average of two experiments is presented. Means within a sampling point with different superscripts (a-d) differ significantly ($P < 0.05$).

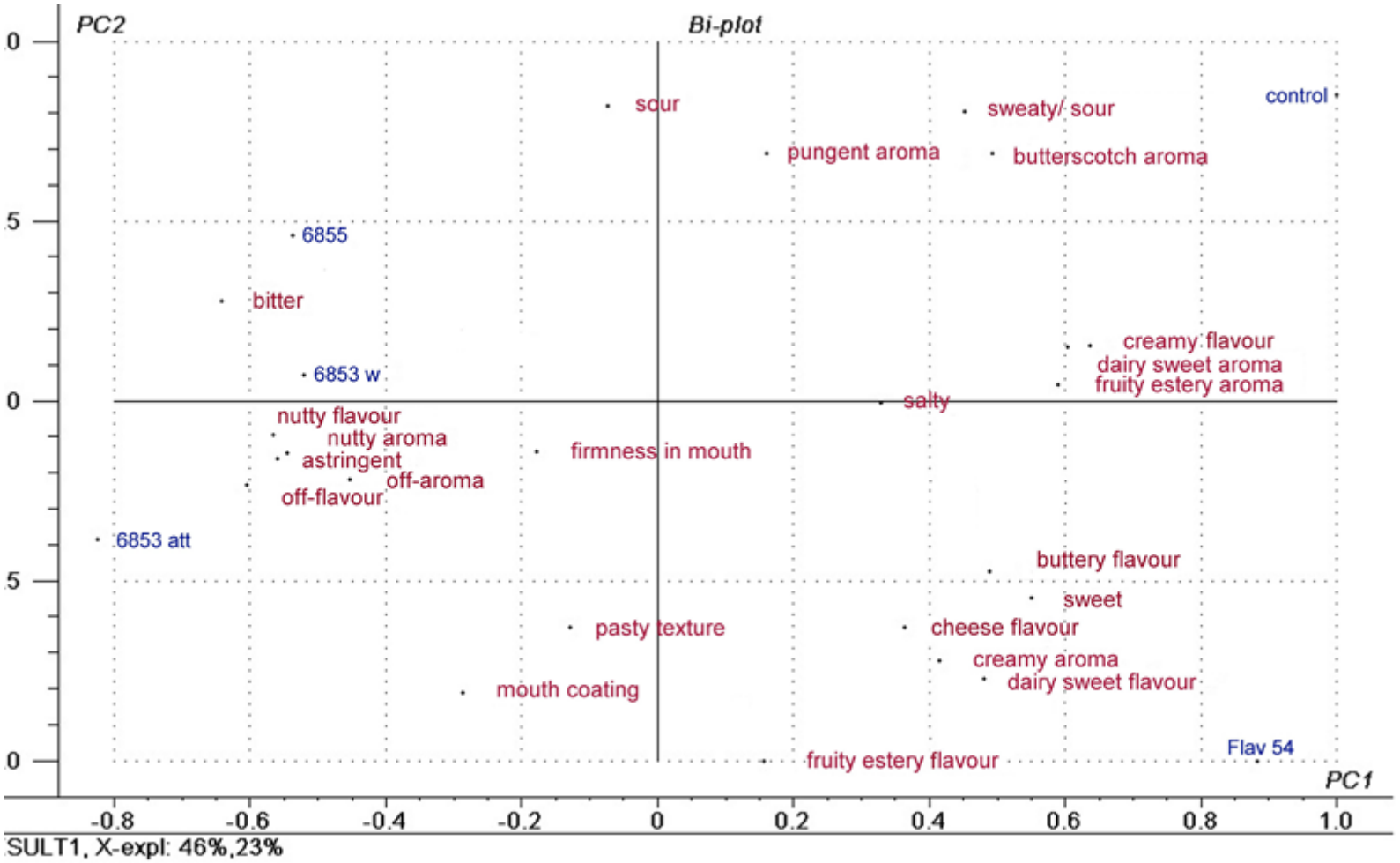


Fig. 4. Principal component bi-plot of sensory evaluation of mini Gouda-type cheeses.

3.8. Appendices

Appendix 3-1

Sensory terms used for the assessor evaluation of mini Gouda-type cheeses.

Attribute	Definition	Scale
Sweaty/ sour aroma	The aromatics reminiscent of perspiration, foot odour. Sour, stale, slightly cheesy, moist, stained or odorous with sweat	0 = none, 10 = extreme
Dairy sweet aroma	Dairy sweet aroma-intensity. The aromatics associated with sweetened cultured dairy products such as fruit yoghurt	0 = none, 10 = extreme
Pungent aroma	Physically penetrating sensation in the nasal cavity. Sharp smelling or tasting, irritant	0 = none, 10 = extreme
Off aroma	Off aroma (Rancid)	0 = none, 10 = extreme
Buttery aroma	Diacetyl aroma-buttery, butterscotch, popcorn	0 = none, 10 = extreme
Fruity estery aroma	Fruity estery aroma-short chain fatty acid esters	0 = none, 10 = extreme
Creamy aroma	The smell associated with creamy/milky products	0 = none, 10 = extreme
Nutty aroma	The non-specific nut flavour that is characteristic of several different nuts, e.g. peanuts, hazelnuts, pecans	0 = none, 10 = extreme
Sweet taste	Fundamental taste sensation of which sucrose is typical	0 = none, 10 = extreme
Salt taste	Fundamental taste sensation of which sodium chloride is typical	0 = none, 10 = extreme
Sour taste	Fundamental taste sensation of which lactic acid is typical	0 = none, 10 = extreme
Bitter taste	Fundamental taste sensation of which caffeine or quinine in soda water is typical	0 = none, 10 = extreme
Astringent	Fundamental taste sensation of which aluminium sulphate is typical	0 = none, 10 = extreme
Firmness in the mouth	Firm texture in the mouth	0 = none, 10 = extreme
Pasty texture	Pasty texture	0 = none, 10 = extreme
Cream flavour	The flavour associated with creamy/milky products	0 = none, 10 = extreme
Off-flavour	Off-flavour (Rancid)	0 = none, 10 = extreme
Dairy sweet flavour	The flavours associated with sweetened cultured dairy products such as fruit yoghurt	0 = none, 10 = extreme
Buttery flavour	Of the nature of, or containing butter	0 = none, 10 = extreme
Mouth coating	Mouth coating	0 = none, 10 = extreme
Butterscotch aroma	The smell associated with toffee products	0 = none, 10 = extreme
Fruity estery flavour	Fruity estery flavour- short chain fatty acid esters	0 = none, 10 = extreme
Nutty Flavour	The non-specific nut flavour that is characteristic of several different nuts, e.g. peanuts, hazelnuts, pecans	0 = none, 10 = extreme

Appendix 3-2

Cheese composition analysis carried out at day 14.

Cheese trial	pH	Moisture (%)	Protein (%)	Fat (%)	S/M (%)	FDM (%)	MNFS (%)
control	5.30a (0.03)	43.83a (0.80)	22.70a (0.59)	28.10a (0.91)	2.81a (0.15)	50.03a (2.21)	61.24a (1.90)
Flav54	5.24ab (0.00)	43.57a (0.19)	23.05a (0.12)	28.29a (0.36)	2.44a (0.12)	50.14a (0.47)	60.76a (0.04)
6853 att	5.07c (0.03)	45.81a (0.33)	22.90a (0.12)	27.39a (0.33)	2.83a (0.38)	50.55a (0.90)	63.09a (0.73)
6853 w	5.28ab (0.01)	43.46a (0.01)	23.44a (0.10)	29.33a (0.42)	2.54a (0.28)	51.88a (0.72)	61.50a (0.34)
6860	5.21b (0.02)	44.50a (1.08)	22.29a (0.68)	28.51a (0.46)	2.66a (0.17)	51.39a (1.82)	62.25a (1.92)
6854	5.29ab (0.04)	42.81a (1.48)	23.16a (0.87)	29.08a (1.14)	2.37a (0.38)	50.83a (0.69)	60.35a (1.11)
6855	5.28ab (0.00)	43.26a (0.43)	23.23a (0.37)	28.66a (1.22)	2.92a (0.08)	50.49a (1.75)	60.63a (0.42)
6856	5.25ab (0.00)	43.34a (0.46)	21.49a (0.88)	30.11a (1.84)	2.67a (0.21)	53.13a (2.81)	62.03a (0.97)

The following abbreviations are used: S/M, salt in moisture; FDM, fat in dry matter; MNFS, moisture in non- fat substances.

[a-d] Means within columns not sharing a common superscript differ statistically ($P < 0.05$). Standard deviation is presented in parenthesis.

Appendix 3-3

P-Values[a] of the estimated regression coefficients (ANOVA values) for the relationships of sensory terms as derived by Jack-knife uncertainty testing for Gouda-type cheese samples.

	Liking of appearance	Dairy sweet aroma	Sweaty/sour	Pungent aroma	Off-aroma	Butterscotch aroma	Fruity/ estery aroma	Creamy aroma	Nutty Aroma	Salt Taste	Sour Taste
control	-0.610	0.241	0.593	0.838	-0.441	0.451	1.000	0.763	-0.816	0.609	-0.814
6853 w	0.677	-0.257	-0.723	-0.961	0.645	-0.821	-0.949	-0.864	0.843	-0.259	0.885
6855	0.712	-0.628	-0.871	-0.926	0.812	-0.947	-0.985	-0.991	0.870	-0.255	0.774
6853 att	0.736	-0.175	-0.606	-0.804	0.820	-0.278	-0.914	-0.537	0.721	-0.953	0.763
Flav54	-0.463	0.493	0.706	0.888	-0.961	0.852	0.888	0.420	-0.859	0.790	-0.174

	Bitter taste	Sweet taste	Astringent taste	Firmness in mouth	Pasty texture	Mouth coating	Cream flavour	Nutty flavour	Off-flavour	Dairy sweet flavour	Cheese flavour	Buttery flavour	Fruity/ estery flavour
control	-0.297	0.832	-0.455	-0.555	-0.757	-0.376	0.210	-0.040	-0.054	0.887	0.790	0.846	-0.296
6853 w	0.190	-0.389	0.658	0.344	0.412	0.901	-0.022	0.296	0.166	-0.218	-0.475	-0.138	0.948
6855	0.464	-0.634	0.868	0.690	0.326	0.520	-0.348	0.628	0.493	-0.322	-0.688	-0.180	0.986
6853 att	0.427	-0.302	0.551	0.694	0.124	0.156	-0.250	0.578	0.091	-0.592	-0.930	-0.711	0.850
Flav54	-0.415	0.023	-0.928	-0.738	-0.846	-0.756	0.658	-0.902	-0.374	0.044	0.357	0.517	-0.502

[a]*P*-Values from the estimated regression coefficients from ANOVA-Partial Least Squares Regression (PLSR). The sign dictates whether the correlation is positively or negatively correlated.

Chapter 4

Phages of non-dairy lactococci: isolation and characterisation of ϕL47, a phage infecting the grass isolate *Lactococcus lactis* ssp. *cremoris* DPC6860

All of this chapter has been published in Frontiers in Microbiology (2013). 4: 417, 1-15. DOI: 10.3389/fmicb.2013.00417

4.1. Abstract

Lactococci isolated from non-dairy sources have been found to possess enhanced metabolic activity when compared to dairy strains. These capabilities may be harnessed through the use of these strains as starter or adjunct cultures to produce more diverse flavour profiles in cheese and other dairy products. To understand the interactions between these organisms and the phages that infect them, a number of phages were isolated against lactococcal strains of non-dairy origin. One such phage, ϕL47, was isolated from a sewage sample using the grass isolate *L. lactis* ssp. *cremoris* DPC6860 as a host. Visualisation of phage virions by scanning electron microscopy established that this phage belongs to the family *Siphoviridae* and possesses a long tail fiber, previously unseen in dairy lactococcal phages. Determination of the lytic spectrum revealed a broader than expected host range, with ϕL47 capable of infecting four industrial dairy strains, including ML8, HP and 310, as well as three additional non-dairy isolates. Whole genome sequencing of ϕL47 revealed a dsDNA genome of 128,546 bp, making it the largest sequenced lactococcal phage to date. In total, 190 open reading frames (ORFs) were identified, and comparative analysis revealed that the predicted products of 117 of these ORFs shared greater than 50% amino acid identity with those of *L. lactis* phage ϕ949, a phage isolated from cheese whey. Despite their different ecological niches, the genomic content and organisation of ϕL47 and ϕ949 are quite similar, with both containing four gene clusters oriented in different transcriptional directions. Other features that distinguish ϕL47 from ϕ949 and other lactococcal phages, in addition to the presence of the tail fiber and the genome length, include a low GC content (32.5%) and a high number of predicted tRNA genes (eight). Comparative genome analysis supports the conclusion that ϕL47

is a new member of the 949 lactococcal phage group which currently includes the dairy ϕ949.

4.2. Introduction

Cultures of lactic acid bacteria (LAB) used in cheese manufacture play a pivotal role in the formation of cheese flavour (Urbach, 1995). While the limited numbers of established dairy cultures in use, have greatly reduced inconsistencies in cheese quality, it can be at the expense of cheese flavour. LAB isolated from non-dairy environments, such as plant material, often exhibit enhanced metabolic capabilities when compared to those of dairy origin and have been shown to contribute to a more diverse flavour profile in the cheese (Ayad *et al.*, 1999; Ayad *et al.*, 2000; Morales *et al.*, 2003). In addition to their role in flavour enhancement, cultures of non-dairy origin have also been shown to be naturally insensitive to bacteriophages which infect industrial strains (Ayad *et al.*, 2000). Phage infection is the single largest cause of industrial milk fermentation problems, negatively impacting on the consistency of cheese quality and resulting in large economic losses (Coffey and Ross, 2002). Non-dairy lactococcal strains could potentially be exploited for use in cheese culture rotations as phage insensitive strains and may well reduce the negative consequences of phage infection within a production plant. However, given the high evolution rate of phages, it is highly likely that over time, phages would emerge to threaten these bacterial strains.

All lactococcal phages isolated to date are members of the *Caudovirales* order which comprises of three families: *Siphoviridae* (long, non-contractile tails), *Myoviridae* (long, contractile tails) and *Podoviridae* (short tails) (for review see Veesler and Cambillau, 2011). Phage classification systems have been utilised in the design of anti-phage strategies to prevent phages amplifying to high titres in manufacturing plants (Jarvis *et al.*, 1991). DNA hybridisation assays and comparative genomic analysis coupled with visualisation of virions has led to the division of

lactococcal phages into ten groups (Deveau *et al.*, 2006). Of these phage groups, the c2, 936 and P335 groups are predominantly associated with failed dairy fermentations and are frequently isolated from cheese manufacturing plants (Moineau *et al.*, 1992; Murphy *et al.*, 2013). Currently within the NCBI database, there are complete genome sequences for over 30 lactococcal bacteriophages. Phages classified within the same group can share a large degree of genome similarity as observed within the c2 and 936 groups (Lubbers *et al.*, 1995; Chopin *et al.*, 2001; Crutz-Le Coq *et al.*, 2002). However, within the P335 group, there exists a high proportion of genetic diversity attributed to the presence of both temperate and virulent bacteriophages (Labrie and Moineau, 2007; Garneau *et al.*, 2008). Collectively, these groups account for the majority of sequenced lactococcal phage genomes, ranging in size from 22-40 kb. Larger lactococcal phages do also exist, with ϕ949 at 114.8 kb being the largest sequenced lactococcal phage prior to this study (Samson and Moineau, 2010). With the elucidation of the ϕ949 genome, a genomic sequence is available for at least one member of each phage group.

The use of a wide range of different bacterial strains possessing an industrially important phenotype has presumably contributed to the diversity of lactococcal phages. Genome analysis of the rarer, uncommon lactococcal phages has shed further light on this diversity. Comparative analysis of the structural proteins of ϕ1358, show a high degree of similarity to 2 *Listeria monocytogenes* phages. Moreover this phage was also found to possess an uncommonly high G-C content of 51 (Dupuis and Moineau, 2010). Recent analysis of the ϕ1706 genome, revealed that 22 ORFs shared similarities with proteins of *Firmicutes*, typically found in the human gut (Garneau *et al.*, 2008). The ability of this phage to infect lactococci was attributed to the acquisition of a four gene module, allowing for host recognition of lactococcal cells

(Garneau *et al.*, 2008). Genomic analysis of the ϕKSY1 genome has suggested the exchange of genetic material between bacteria and phages from different environments (Chopin *et al.*, 2007). ϕQ54 was found to possess a different modular configuration, thought to be derived from recombination events with 936 and c2 type phages (Fortier *et al.*, 2006). It is probable that these rarer phages arose from recombination events with other lactococcal phages and phages infecting other Gram-positive bacteria (Dupuis and Moineau, 2010). Moreover, these phages appear to be less suited to thrive in milk fermentations in contrast to phages of the more common lactococcal phage groups, which possess rapid reproduction rates believed to be driven by evolutionary pressure (Ferguson and Coombs, 2000; Dupuis and Moineau, 2010).

In this study we describe the isolation and characterisation of ϕL47, a large lytic phage which infects the non-dairy isolate *L. lactis* ssp. cremoris DPC6860. We also report the complete genome sequence of ϕL47 which, to our knowledge, is the largest lactococcal phage reported to date. Due to the emergence of non-dairy lactococci as dairy cultures with enhanced flavour-forming activity, sufficient data needs to be generated with regards to these cultures and their phage resistance if they are to be successfully utilised in dairy processing. Therefore, the objective of this study was to provide a better understanding of phages of lactococci from non-dairy origins at a phenotypic and genomic level, thus offering further insight into phage-host interactions.

4.3. Materials and methods

4.3.1. Bacterial strains, bacteriophage and culture conditions

Dairy and non-dairy *Lactococcus* strains were supplied by the TFRC-Moorepark culture collection (Teagasc Food Research Centre Moorepark, Cork, Ireland). *L. lactis* ssp. *cremoris* DPC6860 was previously isolated from grass (Chapter 2) and characterised by 16S rRNA analysis as described by Alander *et. al*. (1999). All lactococcal strains used in this study were cultured in M17 (Oxoid, Hampshire, England) media containing 0.5% lactose (wt./ vol.) (VWR, Leuven, Belgium) (LM17) at 30 ^{o}C for 18-24 hrs under aerobic conditions. Double-strength M17 broth, used for phage enrichment of sewage samples, was prepared by doubling the amount of dry M17 media and reconstituting it in the same amount of distilled water as the 1x media with the addition of 1% lactose (wt./ vol.). Soft agar overlays and solid agar medium contained 0.75% and 1.5% agar respectively. Lactococcal phages ϕKSY1, ϕc2, ϕbIL170 and ϕP008 (hosts IE-16, IL1403, IL1403 and IL1403 respectively) were originally obtained from the Felix d'Hérelle Reference Centre for Bacterial Viruses (GREB, Pavillon de Médecine Dentaire Université Laval, QC, Canada**)** while phages ϕ712, ϕHP, ϕebI and ϕML3 (hosts MG1363, HP, M31363 and MG1363 respectively) were obtained from the UCC culture collection (University College Cork, Cork, Ireland). Phages were propagated using their respective hosts at 30 ^{o}C in M17 media containing 0.5% glucose (wt./ vol.) (Sigma-Aldrich, Dublin, Ireland) and 10 mM $CaCl_2$.

4.3.2. Sensitivity of non-dairy isolates to dairy phages

Bacterial infection by phages was assessed using spot plate assays with phage titres of 10^8 PFU/ ml. Briefly, 10 µl of phage lysate was spotted onto an LM17 soft

agar overlay containing 10 mM $CaCl_2$ and inoculated with $1x10^8$ CFU/ ml of the host organism. Spot plates were allowed to dry before incubation aerobically, at 30 ^oC for 24 hrs.

4.3.3. Phage isolation and propagation

Raw sewage samples were collected from the sewage treatment facility in Mitchelstown, Co. Cork, Ireland. Phage isolation was conducted as described previously (Alemayehu *et al.*, 2009). Individual plaques isolated following this method underwent three successive rounds of plaque purification, to ensure that a pure phage was isolated. Briefly, a single plaque was aseptically removed from an overlay plate using a sterile 1 ml pipette tip and added to 5 ml of mid-log phase host, containing 10 mM $CaCl_2$. Following overnight incubation at 30 ^oC, the mixture was centrifuged at 4,500 rpm for 15 mins and filtered through a 0.45 μm pore filter (Sarstedt, Wexford, Ireland). The filtrate was diluted to 10^{-8} and plaqued on the appropriate media for 24 hrs at 30 ^oC. These steps were repeated twice on the resulting plaques until a single pure phage was obtained.

4.3.4. Lytic spectrum and adsorption to lactococcal cells

The lytic spectrum of ϕL47 was determined, as for dairy lactococcal phages, by spotting 10 μl of phage lysate containing 10^7 PFU/ ml onto soft agar seeded with a *L. lactis* strain. Adsorption of phage particles to lactococcal cells was determined as follows: 10 μl of phage lysate (~10^7 PFU/ ml) was mixed with 2 ml of late exponential cells. Samples were incubated at 30 ^oC for 10 mins with shaking, to allow the phage particles to attach. Each sample was centrifuged at 14,000 rpm for 5 mins and filtered through a 0.2 μm filter (Sarstedt, Wexford, Ireland). The number of phages in the

supernatant was determined by plaque assay and % adsorption was calculated using the formula [(initial titre – titre in supernatant)/ initial titre x 100%].

4.3.5. Electron microscopy

High titre phage suspensions for visualisation of phage particles were prepared using CsCl gradient purification and ammonium acetate concentration. Pure phage samples were obtained using a CsCl gradient of polyethylene glycol (Sigma-Aldrich, Dublin, Ireland) precipitates as described by Sambrook and Russell (2001). For ammonium acetate precipitation, 1 L of fresh lysate was centrifuged at 10,000 rpm at 4 ^{o}C for 10 mins and filter sterilized using a 0.45 μm pore filter (Sarstedt). Phage particles were subsequently precipitated by centrifugation at 20,000 rpm for 1 hr at 4 ^{o}C. The supernatant was removed and the phage pellet re-suspended in 10 ml of ice cold 0.1 M ammonium acetate (Sigma-Aldrich). After pooling of phage samples, a further 10 ml of ammonium acetate was added and the sample was centrifuged at 20,000 rpm for 1 hr. The supernatant was again removed and the pellet re-suspended in a final volume of 1 ml ammonium acetate. Negative staining was performed on both phage samples using 2% uranyl acetate on carbon films. Each grid was examined at an 80 kV acceleration voltage using a Tecnai 10 transmission electron microscope (FEI Compan, Eindhoven, Holland). Micrograph images were captured using a MegaView 2 CCD-camera (SIS, Münster, Germany). Phage structure dimensions were determined based on the average of 10-15 measurements.

4.3.6. Structural analysis of phage proteins

Analysis of structural proteins was performed as described previously (Kelly *et al.*, 2012) using high titre phage suspensions obtained from ammonium acetate

concentration. Samples were mixed with 4X sample loading buffer and heated at 95 ^{o}C prior to loading in a 12% SDS polyacrylamide gel. Protein bands were stained using Coomassie blue staining and excess dye removed using a de-staining solution (40% ethanol, 53% distilled water and 7% acetic acid). Protein size was estimated using a broad range protein ladder (New England Biolabs, Hertfordshire, UK) as a relative molecular weight marker (MWM).

4.3.7. Restriction Digest Profiles

DNA from ϕL47 was extracted as previously described (Alemayehu et al., 2009). Phage DNA was digested using the endonuclease enzymes EcoRI and HindIII (New England Biolabs) in 20 µl reactions. Each reaction contained 10 µl of PCR grade H_2O, 1 µg of phage DNA, 2 µl of buffer and 1 µl of enzyme, with BSA was added as required. The entire endonuclease reaction was run on a 0.8 % agarose gel containing 200 ng/ ml ethidium bromide at 50 V for 18 hrs at 4 ^{o}C using Hyperladder 1 kb (Bioline, London, England) as a relative molecular marker. The size of the ϕL47 genome was estimated by the addition of restriction digest fragments. Digestion profiles were compared to those of sequenced lactococcal phage using the rebase software tool (New England Biolabs).

4.3.8. DNA sequencing, annotation and comparative analysis

Bacteriophage DNA sequencing was performed using the Roche GS FLX+ system to >20x coverage (MWG, Ebersberg, Germany). The quality of the raw assembly reads were visualised and verified using the programme Hawkeye (Amos) (Schatz *et al.*, 2013) and Consed (Gordon, 2003). To verify the genome structure, PCR amplicons were generated with the Platinum Hi-fidelity PCR Supermix

(Invitrogen, Life Technologies, Dublin, Ireland) for various regions of the genome and at contig ends, followed by direct sequencing. The final phage genome was assembled using the Phred-Phrap-Consed package (Ewing and Green, 1998; Gordon, 2003). ORFs were predicted using the programs prodigal (Hyatt *et al.*, 2010) and Glimmer (Delcher *et al.*, 1999). Annotation was provided by the RAST annotation software (Aziz *et al.*, 2008) and GAMOLA (Altermann and Klaenhammer, 2003). Genome annotation was verified manually using Artemis (http://www.sanger.ac.uk/resources/software/artemis/) (Rutherford *et al.*, 2000) and detected open reading frames (ORFs) were functionally annotated using BLASTp (Altschul *et al.*, 1990). Conserved domains were detected using InterProScan (http://www.ebi.ac.uk/InterProScan/) and DELTA-BLAST (Domain Enhanced Lookup Time Accelerated BLAST) (Marchler-Bauer *et al.*, 2011). Identification of tRNAs was achieved using the software packages tRNA scan-SE (Lowe and Eddy, 1997) and ARAGORN v1.2.36 (Lowe and Eddy, 1997; Laslett and Canback, 2004). Comparative genome analysis of the *Lactococcus* ϕL47 with its most similar relative, *Lactococcus* ϕ949, was performed using the Artemis Comparison Tool (ACT) programme (Carver *et al.*, 2005). The genome sequence of ϕL47 is available from GenBank/EMBL under the accession number KF926093.

4.4. Results and discussion

4.4.1. Isolation of ϕL47 from sewage

Previous work led to the isolation of a bank of lactococcal strains from non-dairy origins, including grass, vegetable matter and bovine rumen samples (Chapter 2). These strains, which included both *L. lactis* ssp. *lactis* and *L. lactis* ssp. *cremoris*, were evaluated for their flavour-forming ability in a mini Gouda-type cheese model and were shown to have potential to diversify flavour in this system (Cavanagh, *et al.*, 2014). Future use of these strains in a commercial setting will depend on a number of technological characteristics inherent in the strains, including their sensitivity to phages. The phage sensitivity of the non-dairy isolate bank was assessed using eight common dairy lactococcal phages. Phage titres in excess of 10^7 PFU/ ml were propagated and tested against the target strain using a spot plate assay performed in triplicate (Table 1). ϕPOO8, ϕbIL170, ϕHP, ϕC2 and ϕKSY1 were incapable of infecting all of the non-dairy isolates tested. However, ϕML3 was capable of infecting strains DPC6855 and DPC6856, isolated from grass and bovine rumen, respectively, while ϕebI infected the strains DPC6854 and DPC6855 but not DPC6856 (Table 1).

Overall, the non-dairy isolates displayed higher levels of insensitivity to the dairy phages. Given the high evolution rate of phages to overcome bacterial defence systems (Sturino and Klaenhammer, 2006); however, it is highly likely that over time, phages would emerge to threaten these bacterial strains. To understand the interaction of these isolates with the phages that infect them, a screening programme was initiated to identify phages specific for some of the non-dairy lactococci. Following three enrichment cycles of raw sewage samples with the grass isolate *L. lactis* DPC6860 as a host, phage ϕL47 was isolated. ϕL47 was found to form clear plaques of ~0.6 mm diameter when plaqued on *L. lactis* DPC6860. The formation of clear plaques would

suggest that ϕL47 is virulent. Phage titres of 10^7-10^8 PFU/ ml were recovered after a single propagation.

4.4.2. Limited host range of ϕL47

In order to determine the host range of ϕL47, a range of lactococcal strains were chosen for analysis including industrial dairy starters, strains of non-dairy origins and a strain from raw milk (Table 2). It was found that ϕL47 is capable of infecting three out of seven non-dairy *Lactococcus* isolates but only four out of the 19 dairy strains, corresponding to hosts of two phage groups including the 949 group (Table 2). ϕL47 was unable to infect *L. lactis* IL1403, another host strain of ϕ949. Contrastingly, although ϕ949 was only able to infect seven of 59 *Lactococcus* strains, these corresponded to hosts belonging to five phage groups (Samson and Moineau, 2010). Levels of adsorption of ϕL47 to *L. lactis* DPC6860 were >85% compared to <40% for other strains tested (Fig. 1). However, ϕL47 was capable of lysing strain ML8 but not 303 even though a higher level of phage adsorption was observed in the latter. This would suggest that *L. lactis* 303 possesses a form of anti-phage defence which prevents phage DNA entry or cleaves phage DNA.

4.4.3. Extended tail fiber of ϕL47

Visualisation of phage particles was achieved by transmission electron microscopy performed on samples obtained from both CsCl gradient purification of PEG precipitations and ammonium acetate concentration. Images generated from samples obtained by CsCl gradient showed the majority of phage virions to have disintegrated, with broken tails and empty capsids. Images generated from ammonium acetate preparations displayed intact phage particles which allowed for the estimation

of capsid diameter and tail length (Fig. 2). It was determined that ϕL47 possesses an icosahedral capsid (diameter 75 nm) and a non-contractile tail (length 480 nm; width 15 nm) indicating that this phage is a member of the family *Siphoviridae* of the class *Caudovirales*. Morphologically, ϕL47 is very similar to the other lactococcal phage, ϕ949 which possesses a non-contractile tail of 500 nm in length with an icosahedral capsid of 79 nm in diameter (Jarvis, 1984; Deveau *et al.*, 2006; Samson and Moineau, 2010). However, ϕL47 appears to possess a distinctive tail fiber of 280 nm in length and tail fiber width of <10 nm (Fig. 2B). *Siphoviridae* phages frequently possess a tail fiber involved in phage infection; however, a tail fiber of this length has not previously been reported in lactococcal phages (Vegge *et al.*, 2005; Boulanger *et al.*, 2008; Jakutytė *et al.*, 2012). *Enterococcus faecalis* bacteriophage ϕEF24C-P2 was found to possess a long tail fiber which was associated with higher infectivity against bacterial strains than the un-mutated phage which does not have a tail fiber (Uchiyama *et al.*, 2011). Similar to ϕEF24C-P2, an extended tail fiber was found in *Bacteroides fragilis* phage ATCC 51477-BI, which was isolated from wastewater (Hawkins *et al.*, 2008). No bouquet-like arrangement was identified in ϕL47, as demonstrated by phage ϕγ of *Bacillus anthracis* which possessed a 63 nm long tail fiber (Schuch and Fischetti, 2006).

4.4.4. Structural protein analysis

SDS-PAGE was employed to isolate the structural proteins of ϕL47 and this led to the identification of nine individual protein bands (Fig. 3). Five proteins of low molecular weight (<30 kDa) identified from genome sequence analysis, were not identified on the gel. These proteins may not have been present at a sufficient concentration to allow for visualisation. Molecular masses of proteins estimated by

SDS-PAGE corresponded to the estimated molecular weight of structural proteins predicted from genome sequencing. Three major structural proteins were identified from SDS-PAGE, with the largest having a molecular weight of 31.4 kDa and may correspond to a base plate protein (ORF 176). One large minor protein of ~100 kDa was visualised on the gel but did not correspond to any identified ORF. Multiple bands were observed at ~50 kDa and ~30 kDa which could not be assigned to identified ORFs. It is possible that these proteins are proteolytic products of some other phage structural proteins, or they are of bacterial origin. Techniques such as N-terminal sequencing of protein bands would allow for a more definite relationship to be established between genome analyses and SDS-PAGE.

4.4.5. Restriction digestion profiling

DNA was extracted from phage samples and subjected to endonuclease digestion using the enzymes ClaI, SacI, HindIII, EcoRV and EcoRI. HindIII and EcoRV were found to be most suitable for forming comparable digestion patterns (Fig. 2). To verify that ϕL47 were unlike previously sequenced phages, hypothetical digestion profiles were generated for currently sequenced lactococcal phages available in the NCBI database (http://www.ncbi.nlm.nih.gov; accessed March 2012) and compared to those created in this study using Rebase software tool (Roberts *et al.*, 2007) (http://rebase.neb.com/rebase/rebase.html). The patterns observed for HindIII and EcoRV for sequenced lactococcal phages were unlike those of ϕL47 indicating the isolation of a novel phage.

4.5. Analysis of the L47 genome

4.5.1. General characteristics

To date, all 59 sequenced *L. lactis* phages available in the public domain possess a dsDNA genome and are predominately 15 to 35 kb in size with the exception of the larger genomes of ϕKSY1 (79.2 kb), ϕP087 (60.1 kb) and ϕ949 (114.8 kb). Analysis of the ϕL47 genome revealed a dsDNA molecule of 128, 932 bp, making it the largest sequenced lactococcal phage to date. ϕL47 was found to possess a molecular GC content of 32.5% which corresponds to that observed in the similarly sized *Lactococcus* phage, ϕ949 (32.7%). The GC content of the host organism, *L. lactis* DPC6860, was calculated as 35.6%. A total of 190 ORFs were identified in the ϕL47 genome, 117 of which shared significant homology with *L. lactis* ϕ949 (>50% amino acid identity). The majority of these genes were involved in capsid and tail morphogenesis as well as various hypothetical proteins. ϕ949 was isolated from cheese whey in New Zealand over 35 years ago, while ϕL47 was isolated from raw sewage. Five ORFs were identified that shared no significant identity with proteins in the databases. Similar to the unusual genomic arrangement of ϕ949, four gene clusters were identified in the ϕL47 genome, with two clusters transcribed in the opposite direction. Large inter-genic regions were identified between ORFs 37-38 and ORFs 54-55 which were preceded by a shift in the direction of gene transcription. The 190 ORFs identified in the ϕL47 genome (Table 3) account for 88% of the genome which is the same as that identified in ϕ949. No putative function could be assigned to 132 ORFs (69.47%). The predominant start codon (89.71%) was ATG with only eight ORFs starting with an uncommon start codon (TTG, GTG).

4.5.2. Tail morphology

As stated previously, ϕL47 possessed a long non-contractile tail of 480 nm in addition to a long tail fiber of 280 nm. Both capsid and tail structural genes were encoded from ORFs 157 to 182, and organised in a modular arrangement located downstream from the holin gene (Fig. 4). The products of 14 of these ORFs encoded putative structural proteins sharing >70% amino acid identity with ϕ949 (Table 3). This is not unusual as the major structural proteins are generally somewhat conserved in related phages (Suyama and Bork, 2001; Ceyssens *et al.*, 2011). In contrast, proteins involved in adsorption to host cells, such as tail fibers, are expected to differ to a larger degree as they are modified to complement with surface receptors of bacterial hosts (Sandmeier *et al.*, 1992; Hatfull, 2002; Desplats and Krisch, 2003; Silhavy *et al.*, 2010). ORFs 160-163, located within the structural module, encode proteins sharing a high percentage identity (>60%) to proteins of ϕbIL286. ORF 159 encodes a protein with a high percentage identity to a putative tail protein of *Lactococcus* ϕ949 but also the putative base plate protein of *Lactococcus* ϕP335, and the phage tail assembly protein of *L. lactis* ssp. *cremoris* A76. A conserved domain search of the putative tail protein of ϕ949 reveals the presence of a prophage tail super family cl12123 domain which may possibly act as an endopeptidase.

ORF 160 encodes a putative host-specificity protein sharing a 60% identity at the amino acid level with that of the prophage ϕbIL286. This ORF also shares identity to the phage receptor binding protein (35%) and tail host specificity protein (36%) of the plant strain *L. lactis* KF147. ORFs 162 and 163 also indicate homology with ORF 55 of ϕbIL286, both with 89% and 94% identity respectively. No conserved domains were identified within these two proteins; however, using BLASTp, both proteins were found to possess some identity to a putative tail fiber of *Bacillus* ϕNIt1. In order to assign a putative function to these ORFs of ϕL47, ORF 55 of ϕbIL286 was

analysed by InterProScan. The C-terminal region of this gene product was predicted to have an immunoglobulin-like domain IPR007110 as found in the major tail subunit of Enterobacteria ϕHK97 (Juhala *et al.*, 2000) and other phages. Immunoglobulin-like domains participate in various functions, including cell-cell recognition and cell-surface receptors. Therefore it is likely that these genes also form part of the ϕL47 tail.

4.5.3. Host cell lysis

In dsDNA phages, the combined action of the holin-lysin genes function in the release of new phage particles from an infected cell (Daniel *et al.*, 2007). ORF 69 was found to encode a putative endolysin sharing 89% identity with that of ϕ949, which possesses an amidase_2 domain pfam01510. This ORF also shows a high degree of similarity with gp073, sequence ID YP_001469072.1, of the temperate lactococcal phage ϕKSY1 (83% identity). No holin gene was identified in the same locus as the endolysin, however, a putative holin was located remotely from the endolysin (Fig. 4). ORF 154 exhibits an 87% identity to the putative phage holin from ϕ949, containing a holin LLH superfamily cl09890 domain, a conserved domain of approximately 100 amino acids found in prophage and phage of Gram-positive bacteria. The arrangement of the holin and endolysin genes in the ϕL47 genome is not typical of lactococcal phages but a similar organisation has been reported in ϕ949, ϕP087 and ϕ1708 (Garneau *et al.*, 2008; Villion *et al.*, 2009).

4.5.4. DNA replication and nucleotide biosynthesis

The DNA replication module was located downstream from the endolysin gene and upstream from the holin gene. The predicted gene products for ORFs 75

(helicase), 76 (primase), 66 (nucleotidase), 110 and 111 (DNA gyrase) and 114 (deoxyribosyltransferase) displayed a large degree of homology with ϕ949. ORF 89 was identified as a putative replication protein which shares 84% identity with the putative replication initiator protein of *L. lactis* ssp. *lactis* IO-1 (Kato *et al.*, 2012). A DHH domain pfam01368 was identified in ORF 77 which includes the single-stranded DNA exonuclease RecJ. The protein encoded by this ORF exhibits a 95% identity with the single-stranded DNA exonuclease of *Lactococcus* ϕC13, *Lactococcus* ϕCaesusJM1 and *Lactococcus* ϕ949. An exo-nuclease functions to catalyse the cleavage of a single nucleotide from the end of a polynucleotide chain and is involved in DNA repair, recombination and replication (Ceska and Sayers, 1998). No putative endonucleases were identified in the L47 genome, in contrast to two distinct HNH endonucleases in ϕ949. Structure specific endonucleases cut at particular DNA structures and in some cases can give a competitive advantage to a given bacteriophage in a mixed infection (Goodrich-Blair and Shub, 1996). It is possible that ϕL47 possesses no HNH endonucleases as it has acquired other characteristics, such as possessing a reduced number of endonuclease recognition sites, that offer an advantage over other bacteriophages. Group I introns are frequently associated with endonucleases (Chevalier and Stoddard, 2001), however, ORFs 118 and 126 were both found to encode this type of intervening sequence (IVS) between ribonucleotide reductase proteins. Previously, two introns were found within the ribonucleotide reductase large subunit of *Staphylococcus aureus* ϕTwort (Landthaler *et al.*, 2002). In *Escherichia coli* ϕT4, the aerobic Ribonucleotide reductase small subunit and the anaerobic ribonucleotide reductase were both found to possess group I introns (Sjöberg *et al.*, 1986; Young *et al.*, 1994). One explanation offered for the retention of introns in genes, such as those involved in ribonucleotide reduction, is that they encode

functions replicated in the host organism that can be utilised for phage survival until such time as the insertion element and host environment adapt to one another (Derbyshire and Belfort, 1998).

The acquisition of ribonucleotide reductases (RNRs) is thought to arise from the host genome, to enable the adaptation to certain environmental conditions (Dwivedi *et al.*, 2013). These enzymes function in DNA replication and repair via the conversion of ribonucleotides to deoxyribonucleotides (Dwivedi *et al.*, 2013). The gene products of seven ORFs were found to encode ribonucleotide di- and tri-phosphate reductases. ORF 116 encodes a putative anaerobic ribonucleoside-triphosphate reductase with a 25% identity to a putative anaerobic ribonucleoside-triphosphate reductase of *Clostridium* ϕ8074-B1. A class III ribonucleotide reductase domain cd01675 was identified within this gene product which uses a FeS cluster and S-adenosylmethionine to generate a glycyl radical. ORF 124 shares a 65% identity with the ribonucleotide-diphosphate reductase subunit alpha of *Bacillus alcalophilus* and contains a class I RNR domain cd01679. In contrast to class III, a class I RNR domain uses a di-iron-tyrosyl radical. The presence of class I and class III RNR genes in ϕL47 conforms with the RNR complement generally found in phages isolated from sewage (Dwivedi *et al.*, 2013). Other ribonucleotide reductases shared a large degree of homology with ϕ949 as well as other genes involved in nucleotide transport, modification and degradation.

4.5.5. tRNA encoding genes in L47 genome

The role of tRNAs in phage genomes is thought to be the encoding of codons that are less frequent in the host genome, thus allowing for the increased expression of phage proteins (Bailly-Bechet *et al.*, 2007). In all, eight tRNA genes (Table 4) were

identified over a small region of the genome (48,524 to 53,674 bp) in ϕL47, using the tRNA-scan SE and Aragorn software (Lowe and Eddy, 1997; Laslett and Canback, 2004). tRNAs had a %GC content ranging from 39.4 and 50.7. Between ORFs 106 and 107, were positioned $tRNA^{Arg}$, $tRNA^{Asp}$ and $tRNA^{Ala}$ while between ORFs 107 and 108 were $tRNA^{Asn}$, $tRNA^{Ser}$ and $tRNA^{Trp}$. A 58 bp intron, from position 34-35, was identified in $tRNA^{Ser}$. Two $tRNA^{Met}$ genes were identified, the first situated between ORFs 101 and 102, and the second between ORFs 107 and 108. The possession of more than one tRNA for a specific amino acid is a phenomenon that has been observed in *Lactobacillus plantarum* bacteriophage ϕLP65 (Chibani-Chennoufi *et al.*, 2004b). The acquisition of tRNA genes from the host is proposed as being a random event and the genes are either retained, via a set of selection mechanisms, or they are lost (Bailly-Bechet *et al.*, 2007). It is conceivable that ϕL47 has acquired two $tRNA^{Met}$ in order to gain a fitness advantage, due to the relatively harsh environment that it inhabits. *In-situ* burst sizes for organisms in a nutrient-poor environment, such as that from which the host was isolated, are generally smaller than when the host is infected in a nutrient rich, chemically defined media (Chibani-Chennoufi *et al.*, 2004a). The retention of particular tRNA genes in phage genomes are thought to correspond to codons that are less abundant in the host genome (Bailly-Bechet *et al.*, 2007) although this was not observed for ϕ949 and its host *L. lactis* IL1403 (Samson and Moineau, 2010). Therefore we may only hypothesise that these tRNA genes are associated with controlling phage protein production and are possibly involved in increasing reproduction rate and reducing latency time (Bailly-Bechet *et al.*, 2007).

The arrangement of the tRNAs into distinct blocks suggests that they were obtained through separate recombination events with either host DNA, other phages or a combination of both (Weinbauer, 2004). The presence of tRNAs has been found

to be particularly common in phages with a larger genome size (Bailly-Bechet *et al.*, 2007). Six tRNA genes were identified in ϕ949, the most identified in a lactococcal bacteriophage (Samson and Moineau, 2010). Similarly, in ϕLb338-1 and ϕK, both with a genome larger than 100 kb, 2 and 4 tRNAs were identified respectively (O' Flaherty *et al.*, 2004; Alemayehu *et al.*, 2009). No tRNA encoding regions were found in 2 recently elucidated lactococcal phage genomes, belonging to a new P335 group, with sizes of approximately 31 kb (Mahony *et al.*, 2013). However, 5 tRNA genes have been identified in ϕP087, a P087 species, of a smaller genome size (60,074 bp) than ϕ949 and ϕL47 (Villion *et al.*, 2009).

4.5.6. Comparative genome analysis

Numerous ORFs in the L47 genome were found to show identity to putative proteins of other lactococcal phages. ORFs 27-30 encode hypothetical proteins belonging to different lactococcal phages (Table 3). No conserved domains were detected within these ORFs. ORF 119, showing 39% identity with *Lactococcus* ϕKSY1, was found to encode a pfam01139 domain, a conserved domain of the uncharacterized protein family UPF0027. Putative proteins expressed by *Enterococcus* ϕEF24C and *Enterococcus faecium* are also found to be similar to those identified in the ϕL47 genome (3 ORFs; >40% identity). Previously, lactococcal ϕP087 was shown to possess genes with a high degree of similarity to structural genes of a prophage of *Enterococcus faecalis* V538 (Villion *et al.*, 2009). Of note, ORF 23 was found to be highly similar to a phage protein of the plant *Lactococcus* strain, KF147. Upstream from the structural genes, ORF 130 encodes a M23 peptidase domain containing protein similar to *Clostridium perfringens* F262 (Nowell *et al.*, 2012). The position of this gene with genes involved in DNA replication, suggests

that it is not tail associated as in *Lactococcus* phages Tuc2009 and TP901-1 (Seegers *et al.*, 2004; Stockdale *et al.*, 2013). This peptidase would possibly hydrolyse peptidoglycan via D-Ala–D-Asp endopeptidase activity enabling the penetration of stationary phase lactococcal cells (Samson *et al.*, 2013).

Comparative genomic analysis shows the high similarity at a nucleotide level between the ϕL47 and ϕ949 phages (Fig. 5). In addition, given the close amino acid identity at protein level across the genome (Table 3), suggesting a number of shared functions, it is likely that these phages have shared a common ancestor at some point. Studies of the more common lactococcal phage groups have indicated a large degree of homology between members of the c2 and 936 species while the P335 species displays a polythetic species theory (Deveau *et al.*, 2006). Although much discussion surrounds phage evolution, in relation to virulent phages vertical lines of evolution are believed to be crucial in the development of certain phage families (Brussow and Kutter, 2005) and are particularly evident in dairy phages (Brüssow and Desiere, 2001). However, horizontal genetic exchange also plays a key role in the evolution of phages as exemplified by genetic mosaicism formed from a high occurrence of transfer events (Canchaya *et al.*, 2003). Genetic exchange between bacteria and other phages from distinct niches is hypothesised to be pivotal in the evolution of rare lactococcal phages (Fortier *et al.*, 2006; Chopin *et al.*, 2007; Garneau *et al.*, 2008; Villion *et al.*, 2009). Due to the high degree of sequence homology between these two phages, it is expected that ϕL47 is another member of the 949 group. Analysis of ϕ111 revealed this lactococcal phage to possess a genome size of approximately 134 kb and a long tail of 470 nm (Prevots *et al.*, 1990). However, without comparative genome analysis it can only be speculated that ϕ111 is a representative of the 949 group.

4.6. Conclusions

To date, numerous studies have investigated dairy lactococcal phages and how they interact with their respective hosts. Industrial *Lactococcus* strains, used in modern cheese production, are thought to have evolved from plant strains (Kelly *et al.*, 2010). ϕL47, isolated from a non-dairy environment, possessed significant similarity to the rare, dairy phage ϕ949, with both phages isolated from different environments almost 40 years apart. ϕL47 was found to possess a number of features which differentiate it from ϕ949, most notably a long tail fiber, not previously reported in phages of *Lactococcus*. This tail fiber may play an important role in enabling ϕL47 to infect *L. lactis* DPC6860, which was largely resistant to dairy phages, and may account for the persistence of successful virulent phages in the wider environment as observed for ϕQ33 and ϕBM13 (Mahony *et al.*, 2013). Further studies are required to establish the diversity of lactococcal bacteriophages from non-dairy origins and the similarity they may possess with dairy phages of other groups. This information could shed further light on the mechanisms they possess that allow them to thrive in harsher environments and may advance our understanding of host recognition and infection by lactococcal phages.

4.7. Acknowledgements

Thanks to Horst Neve for phage imaging.

Thanks to Caitriona Guinane for phage genome assembly and annotation.

4.8. References

Alander, M., Satokari, R., Korpela, R., Saxelin, M., Vilpponen-Salmela, T., Mattila-Sandholm, T., and Von Wright, A. (1999). Persistence of colonization of human colonic mucosa by a probiotic strain, *Lactobacillus rhamnosus* GG, after oral consumption. *Appl. Environ. Microbiol.* 65**,** 351-354.

Alemayehu, D., Ross, R. P., O'Sullivan, O., Coffey, A., Stanton, C., Fitzgerald, G. F., and McAuliffe, O. (2009). Genome of a virulent bacteriophage Lb338-1 that lyses the probiotic *Lactobacillus paracasei* cheese strain. *Gene.* 448**,** 29-39. DOI: 10.1016/j.gene.2009.08.008.

Altermann, E., and Klaenhammer, T. R. (2003). GAMOLA: a new local solution for sequence annotation and analyzing draft and finished prokaryotic genomes. *OMICS.* 7**,** 161-169. DOI:10.1089/153623103322246557.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215**,** 403-410.

Ayad, E. H. E., Verheul, A., De Jong, C., Wouters, J. T. M., and Smit, G. (1999). Flavour forming abilities and amino acid requirements of *Lactococcus lactis* strains isolated from artisanal and non-dairy origin. *Int. Dairy J.* 9**,** 725-735. DOI: 10.1016/s0958-6946(99)00140-5.

Ayad, E. H. E., Verheul, A., Wouters, J. T. M., and Smit, G. (2000). Application of wild starter cultures for flavour development in pilot plant cheese making. *Int. Dairy J.* 10**,** 169-179. DOI: 10.1016/s0958-6946(00)00041-8.

Aziz, R. K., Bartels, D., Best, A. A., Dejongh, M., Disz, T., Edwards, R. A., Formsma, K., Gerdes, S., Glass, E. M., and Kubal, M. (2008). The RAST Server: rapid annotations using subsystems technology. *BMC Genomics.* 9:75. DOI: 10.1186/1471-2164-9-75.

Bailly-Bechet, M., Vergassola, M., and Rocha, E. (2007). Causes for the intriguing presence of tRNAs in phages. *Genome Res.* 17**,** 1486-1495. DOI: 10.1101/gr.6649807.

Boulanger, P., Jacquot, P., Plançon, L., Chami, M., Engel, A., Parquet, C., Herbeuval, C., and Letellier, L. (2008). Phage T5 straight tail fiber is a multifunctional protein acting as a tape measure and carrying fusogenic and muralytic activities. *J. Biol. Chem.* 283**,** 13556-13564. DOI: 10.1074/jbc.M800052200.

Brüssow, H., and Desiere, F. (2001). Comparative phage genomics and the evolution of *Siphoviridae*: insights from dairy phages. *Mol. Microbiol.* 39**,** 213-223. DOI: 10.1046/j.1365-2958.2001.02228.x.

Brussow, H., and Kutter, E. (2005). "Genomics and evolution of tailed phages," in *Bacteriophages: biology and applications*, eds E. Kutter and A. Sulakvelidze (Washington, DC: CRC Press), 91-128.

Canchaya, C., Fournous, G., Chibani-Chennoufi, S., Dillmann, M.-L., and Brüssow, H. (2003). Phage as agents of lateral gene transfer. *Curr. Opin. Microbiol.* 6**,** 417-424. DOI: http://dx.doi.org/10.1016/S1369-5274(03)00086-9.

Carver, T. J., Rutherford, K. M., Berriman, M., Rajandream, M. A., Barrell, B. G., and Parkhill, J. (2005). ACT: the Artemis Comparison Tool. *Bioinformatics.* 21**,** 3422-3423.

Cavanagh, D., Kilcawley, K. N., O'Sullivan, M. G., Fitzgerald, G. F., and McAuliffe, O. (2014). Assessment of wild non-dairy lactococcal strains for flavour diversification in a mini-Gouda type cheese model. *Food Res. Int.* 62, 432-440. DOI: 10.1016/j.foodres.2014.03.043.

Ceska, T. A., and Sayers, J. R. (1998). Structure-specific DNA cleavage by 5′ nucleases. *Trends Biochem. Sci.* 23**,** 331-336.

Ceyssens, P.-J., Glonti, T., Kropinski, A. M., Lavigne, R., Chanishvili, N., Kulakov, L., Lashkhi, N., Tediashvili, M., and Merabishvili, M. (2011). Phenotypic and

genotypic variations within a single bacteriophage species. *Virology. J.* 8, 1-5. DOI: 10.1186/1743-422X-8-134.

Chevalier, B. S., and Stoddard, B. L. (2001). Homing endonucleases: structural and functional insight into the catalysts of intron/intein mobility. *Nucleic Acids Res.* 29**,** 3757-3774.

Chibani-Chennoufi, S., Bruttin, A., Dillmann, M.-L., and Brüssow, H. (2004a). Phage-host interaction: an ecological perspective. *J. Bacteriol.* 186**,** 3677-3686. DOI: 10.1128/JB.186.12.3677-3686.2004.

Chibani-Chennoufi, S., Dillmann, M.-L., Marvin-Guy, L., Rami-Shojaei, S., and Brüssow, H. (2004b). *Lactobacillus plantarum* bacteriophage LP65: a new member of the SPO1-like genus of the family Myoviridae. *J. Bacteriol.* 186**,** 7069-7083. DOI:10.1128/JB.186.21.7069-7083.2004.

Chopin, A., Bolotin, A., Sorokin, A., Ehrlich, S. D., and Chopin, M. C. (2001). Analysis of six prophages in *Lactococcus lactis* IL1403: different genetic structure of temperate and virulent phage populations. *Nucleic Acids Res.* 29**,** 644-651.

Chopin, A., Deveau, H., Ehrlich, S. D., Moineau, S., and Chopin, M.-C. (2007). KSY1, a lactococcal phage with a T7-like transcription. *Virology.* 365**,** 1-9. DOI: 10.1016/j.virol.2007.03.044.

Coffey, A., and Ross, R. P. (2002). Bacteriophage-resistance systems in dairy starter strains: molecular analysis to application. *Antonie van Leeuwenhoek.* 82**,** 303-321.

Crutz-Le Coq, A.-M., Cesselin, B., Commissaire, J., and Anba, J. (2002). Sequence analysis of the lactococcal bacteriophage bIL170: insights into structural proteins and HNH endonucleases in dairy phages. *Microbiol.* 148**,** 985-1001.

Daniel, A., Bonnen, P. E., and Fischetti, V. A. (2007). First complete genome sequence of two *Staphylococcus epidermidis* bacteriophages. *J. Bacteriol.* 189**,** 2086-2100. DOI: 10.1128/JB.01637-06.

Delcher, A. L., Harmon, D., Kasif, S., White, O., and Salzberg, S. L. (1999). Improved microbial gene identification with GLIMMER. *Nucleic Acids Res.* 27**,** 4636-4641.

Derbyshire, V., and Belfort, M. (1998). Lightning strikes twice: intron–intein coincidence. *Proc. Nat. Acad. Sci.* 95**,** 1356-1357.

Desplats, C., and Krisch, H. M. (2003). The diversity and evolution of the T4-type bacteriophages. *Res. Microbiol.* 154**,** 259-267.

Deveau, H., Labrie, S. J., Chopin, M.-C., and Moineau, S. (2006). Biodiversity and classification of lactococcal phages. *Appl. Environ. Microbiol.* 72**,** 4338-4346. DOI: 10.1128/AEM.02517-05

Dupuis, M.-È., and Moineau, S. (2010). Genome organization and characterization of the virulent lactococcal phage 1358 and its similarities to listeria phages. *Appl. Environ. Microbiol.* 76**,** 1623-1632. DOI: 10.1128/aem.02173-09.

Dwivedi, B., Xue, B., Lundin, D., Edwards, R. A., and Breitbart, M. 2013. A bioinformatic analysis of ribonucleotide reductase genes in phage genomes and metagenomes. *BMC Evol. Biol.* 13:1. DOI: 10.1186/1471-2148-13-33.

Ewing, B., and Green, P. (1998). Base-calling of automated sequencer traces usingPhred. II. error probabilities. *Genome Res.* 8**,** 186-194.

Ferguson, P., and Coombs, D. (2000). Pulse-chase analysis of the *in vivo* assembly of the bacteriophage T4 tail. *J. Mol. Biol.* 297, 99-117.

Fortier, L.-C., Bransi, A., and 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*. *J. Bacteriol.* 188**,** 6101-6114. DOI: 10.1128/JB.00581-06.

Garneau, J. E., Tremblay, D. M., and Moineau, S. (2008). Characterization of 1706, a virulent phage from *Lactococcus lactis* with similarities to prophages from other Firmicutes. *Virology.* 373**,** 298-309. DOI: 10.1016/j.virol.2007.12.002.

Gasteiger, E., Hoogland, C., Gattiker, A., Wilkins, M. R., Appel, R. D., and Bairoch, A. (2005). "Protein identification and analysis tools on the ExPASy server", in *The proteomics protocols handbook,* ed J. M. Walker (New Jersey, USA: Springer), 571-607.

Goodrich-Blair, H., and Shub, D. A. (1996). Beyond homing: competition between intron endonucleases confers a selective advantage on flanking genetic markers. *Cell.* 84**,** 211-221.

Gordon, D. (2003). Viewing and editing assembled sequences using Consed. *Curr. Protoc. Bioinformatics***.** 2: 11.12.11–11.12.43. DOI: 10.1002/0471250953.bi1102s02.

Hatfull, G. F. (2002). A tail of two specifi-cities. *Science.* 295**,** 2031-2032.

Hawkins, S. A., Layton, A. C., Ripp, S., Williams, D., and Sayler, G. S. 2008. Genome sequence of the *Bacteroides fragilis* phage ATCC 51477-B1. *Virology. J.* 5:97. DOI: 10.1186/1743-422X-5-97.

Hyatt, D., Chen, G.-L., Locascio, P., Land, M., Larimer, F., and Hauser, L. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics.* 11:1. DOI: 10.1186/1471-2105-11-119.

Jakutytė, L., Lurz, R., Baptista, C., Carballido-Lopez, R., São-José, C., Tavares, P., and Daugelavičius, R. (2012). First steps of bacteriophage SPP1 entry into *Bacillus subtilis*. *Virology.* 422**,** 425-434. DOI: 10.1074/jbc.M301805200.

Jarvis, A. W. (1984). Differentiation of lactic streptococcal phages into phage species by DNA-DNA homology. *Appl. Environ. Microbiol.* 47**,** 343-349.

Jarvis, A. W., Fitzgerald, G. F., Mata, M., Mercenier, A., Neve, H., Powell, I. B., Ronda, C., Saxelin, M., and Teuber, M. (1991). Species and type phages of lactococcal bacteriophages. *Intervirology.* 32**,** 2-9.

Juhala, R. J., Ford, M. E., Duda, R. L., Youlton, A., Hatfull, G. F., and Hendrix, R. W. (2000). Genomic sequences of bacteriophages HK97 and HK022: pervasive genetic mosaicism in the lambdoid bacteriophages. *J. Mol. Biol.* 299**,** 27-51. DOI: 10.1006/jmbi.2000.3729

Kato, H., Shiwa, Y., Oshima, K., Machii, M., Araya-Kojima, T., Zendo, T., Shimizu-Kadota, M., Hattori, M., Sonomoto, K., and 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. DOI: 10.1128/JB.00074-12.

Kelly, D., O'Sullivan, O., Mills, S., McAuliffe, O., Ross, P., Neve, H., and Coffey, A. (2012). Genome sequence of the phage clP1, which infects the beer spoilage bacterium *Pediococcus damnosus*. *Gene.* 504**,** 53-63. DOI: 10.1016/j.gene.2012.04.085.

Kelly, W. J., Ward, L. J. H., and Leahy, S. C. (2010). Chromosomal diversity in *Lactococcus lactis* and the origin of dairy starter cultures. *Genome Biol. Evol.* 2**,** 729-744. DOI: 10.1093/gbe/evq056.

Labrie, S. J., and Moineau, S. (2007). Abortive infection mechanisms and prophage sequences significantly influence the genetic makeup of emerging lytic lactococcal phages. *J. Bacteriol.* 189**,** 1482-1487.

Landthaler, M., Begley, U., Lau, N. C., and Shub, D. A. (2002). Two self-splicing group I introns in the ribonucleotide reductase large subunit gene of *Staphylococcus aureus* phage Twort. *Nucleic Acids Res.* 30**,** 1935-1943. DOI:10.1093/nar/30.9.1935.

Laslett, D., and Canback, B. (2004). ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Res.* 32**,** 11-16. DOI: 10.1093/nar/gkh152.

Lowe, T. M., and Eddy, S. R. (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* 25**,** 0955-0964.

Lubbers, M. W., Waterfield, N. R., Beresford, T., Le Page, R., and Jarvis, A. W. (1995). Sequencing and analysis of the prolate-headed lactococcal bacteriophage c2 genome and identification of the structural genes. *Appl. Environ. Microbiol.* 61**,** 4348-4356.

Mahony, J., Martel, B., Tremblay, D. M., Neve, H., Heller, K. J., Moineau, S., and Van Sinderen, D. (2013). Identification of a new P335 subgroup through molecular analysis of lactococcal phages Q33 and BM13. *Appl. Environ. Microbiol.* 79**,** 4401-4409. DOI: 10.1128/AEM.00832-13.

Marchler-Bauer, A., Lu, S., Anderson, J. B., Chitsaz, F., Derbyshire, M. K., Deweese-Scott, C., Fong, J. H., Geer, L. Y., Geer, R. C., and Gonzales, N. R. (2011). CDD: a Conserved Domain Database for the functional annotation of proteins. *Nucleic Acids Res.* 39**,** D225-D229. DOI: 10.1093/nar/gkq1189.

Moineau, S., Fortier, J., Ackermann, H.-W., and Pandian, S. (1992). Characterization of lactococcal bacteriophages from Quebec cheese plants. *Can. J. Microbiol.* 38**,** 875-882.

Morales, P., Fernández-GarcíA, E., Gaya, P., and Nuñez, M. (2003). Formation of volatile compounds by wild *Lactococcus lactis* strains isolated from raw ewes' milk cheese. *Int. Dairy J.* 13**,** 201-209. DOI: 10.1016/s0958-6946(02)00151-6.

Murphy, J., Royer, B., Mahony, J., Hoyles, L., Heller, K., Neve, H., Bonestroo, M., Nauta, A., and Van Sinderen, D. (2013). Biodiversity of lactococcal bacteriophages isolated from 3 Gouda-type cheese producing plants. *J. Dairy Sci.* 96**,** 4945-4957.

Nowell, V. J., Kropinski, A. M., Songer, J. G., Macinnes, J. I., Parreira, V. R., and Prescott, J. F. 2012. Genome sequencing and analysis of a Type A *Clostridium*

perfringens isolate from a case of bovine clostridial abomasitis. *PloS One.* 7:3. DOI: 10.1371/journal.pone.0032271.

O'Flaherty, S., Coffey, A., Edwards, R., Meaney, W., Fitzgerald, G. F., and Ross, R. P. (2004). Genome of staphylococcal phage K: a new lineage of Myoviridae infecting gram-positive bacteria with a low G+C content. *J. Bacteriol.* 186**,** 2862-2871. DOI: 10.1128/JB.186.9.2862-2871.2004.

Prevots, F., Mata, M., and Ritzenthaler, P. (1990). Taxonomic differentiation of 101 lactococcal bacteriophages and characterization of bacteriophages with unusually large genomes. *Appl. Environ. Microbiol.* 56**,** 2180-2185.

Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M. A., and Barrell, B. (2000). Artemis: sequence visualization and annotation. *Bioinformatics.* 16**,** 944-945. DOI:10.1093/bioinformatics/16.10.944.

Sambrook, J., and Russell, D. W. (2001). *Molecular cloning: a laboratory manual. Volume 1–3.* Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press.

Samson, J. E., Magadán, A. H., Sabri, M., and Moineau, S. (2013). Revenge of the phages: defeating bacterial defences. *Nature Rev. Microbiol.* 11**,** 675-687. DOI: 10.1038/nrmicro3096.

Samson, J. E., and Moineau, S. (2010). Characterization of *Lactococcus lactis* phage 949 and comparison with other lactococcal phages. *Appl. Environ. Microbiol.* 76**,** 6843-6852. DOI: 10.1128/aem.00796-10.

Sandmeier, H., Iida, S., and Arber, W. (1992). DNA inversion regions Min of plasmid p15B and Cin of bacteriophage P1: evolution of bacteriophage tail fiber genes. *J. Bacteriol.* 174**,** 3936-3944.

Schatz, M. C., Phillippy, A. M., Sommer, D. D., Delcher, A. L., Puiu, D., Narzisi, G., Salzberg, S. L., and Pop, M. (2013). Hawkeye and AMOS: visualizing and assessing

the quality of genome assemblies. *Brief. Bioinform.* 14**,** 213-224. DOI: 10.1093/bib/bbr074.

Schuch, R., and Fischetti, V. A. (2006). Detailed genomic analysis of the Wβ and γ phages infecting *Bacillus anthracis*: implications for evolution of environmental fitness and antibiotic resistance. *J. Bacteriol.* 188**,** 3037-3051. DOI: 10.1128/JB.188.8.3037-3051.2006.

Seegers, J. F., McGrath, S., O'Connell-Motherway, M., Arendt, E. K., Van De Guchte, M., Creaven, M., Fitzgerald, G. F., and Van Sinderen, D. (2004). Molecular and transcriptional analysis of the temperate lactococcal bacteriophage Tuc2009. *Virology.* 329**,** 40-52.

Silhavy, T. J., Kahne, D., and Walker, S. (2010). The bacterial cell envelope. *Cold Spring Harb. Perspect. Biol.* 2**,** 1-16. DOI: 10.1101/cshperspect.a000414.

Sjöberg, B. M., Hahne, S., Mathews, C. Z., Mathews, C. K., Rand, K. N., and Gait, M. J. (1986). The bacteriophage T4 gene for the small subunit of ribonucleotide reductase contains an intron. *EMBO J.* 5**,** 2031-2036.

Stockdale, S. R., Mahony, J., Courtin, P., Chapot-Chartier, M.-P., Van Pijkeren, J.-P., Britton, R. A., Neve, H., Heller, K. J., Aideh, B., and Vogensen, F. K. (2013). The lactococcal phages Tuc2009 and TP901-1 incorporate two alternate forms of their tail fiber into their virions for infection specialization. *J. Biol. Chem.* 288**,** 5581-5590. DOI: 10.1074/jbc.M112.444901.

Sturino, J. M., and Klaenhammer, T. R. (2006). Engineered bacteriophage-defence systems in bioprocessing. *Nature Rev. Microbiol.* 4**,** 395-404.

Suyama, M., and Bork, P. (2001). Evolution of prokaryotic gene order: genome rearrangements in closely related species. *Trends Genet.* 17**,** 10-13. DOI: 10.1016/S0168-9525(00)02159-4.

Uchiyama, J., Takemura, I., Satoh, M., Kato, S.-I., Ujihara, T., Akechi, K., Matsuzaki, S., and Daibata, M. (2011). Improved Adsorption of an *Enterococcus faecalis* bacteriophage ϕEF24C with a spontaneous point mutation. *PloS One.* 6:10. DOI: 10.1371/journal.pone.0026648.

Urbach, G. (1995). Contribution of lactic acid bacteria to flavour compound formation in dairy products. *Int. Dairy J.* 5**,** 877-903.

Veesler, D., and Cambillau, C. (2011). A common evolutionary origin for tailed-bacteriophage functional modules and bacterial machineries. *Microbiol. Mol. Biol. Rev.* 75**,** 423-433. DOI: 10.1128/MMBR.00014-11.

Vegge, C. S., Brøndsted, L., Neve, H., Mc Grath, S., Van Sinderen, D., and Vogensen, F. K. (2005). Structural characterization and assembly of the distal tail structure of the temperate lactococcal bacteriophage TP901-1. *J. Bacteriol.* 187**,** 4187-4197. DOI: 10.1128/JB.187.12.4187-4197.2005.

Villion, M., Chopin, M. C., Deveau, H., Ehrlich, S. D., Moineau, S., and Chopin, A. (2009). P087, a lactococcal phage with a morphogenesis module similar to an *Enterococcus faecalis* prophage. *Virology.* 388**,** 49-56. DOI: 10.1016/j.virol.2009.03.011.

Weinbauer, M. G. (2004). Ecology of prokaryotic viruses. *FEMS Microbiol. Rev.* 28**,** 127-181. DOI: 10.1016/j.femsre.2003.08.001.

Young, P., Öhman, M., Xu, M. Q., Shub, D. A., and Sjöberg, B. M. (1994). Intron-containing T4 bacteriophage gene *sunY* encodes an anaerobic ribonucleotide reductase. *J. Biol. Chem.* 269**,** 20229-20232.

Table 1. Phage resistance profile of dairy lactococcal phages to non-dairy *Lactococcus lactis* strains.

		Phage							
	Phage group	936				c2			KSY1
Strain	Origin	ϕ712	ϕPOO8	ϕbIL170	ϕHP	ϕC2	ϕebI	ϕML3	ϕKSY1
L. lactis ssp. *lactis* DPC6853	Corn	–	–	–	–	–	–	–	–
L. lactis ssp. *cremoris* DPC6854	Grass	–	–	–	–	–	+	–	–
L. lactis ssp. *cremoris* DPC6855	Grass	–	–	–	–	–	+	+	–
L. lactis ssp. cremoris DPC6856	Rumen	+	–	–	–	–	–	+	–
L. lactis ssp. *cremoris* DPC6857	Grass	–	–	–	–	–	–	–	–
L. lactis ssp. *cremoris* DPC6858	Grass	–	–	–	–	–	–	–	–
L. lactis ssp. *cremoris* DPC6859	Grass	–	–	–	–	–	–	–	–
L. lactis ssp. *cremoris* DPC6860	Grass	–	–	–	–	–	–	–	–

Table 2. Host range of *Lactococcus* phage ϕL47.

Species	Origin/ Use	Host cell lysis
L. lactis ssp. *cremoris* DPC6855	Grass	–
L. lactis ssp. *cremoris* DPC6859	Grass	+
L. lactis ssp. *cremoris* DPC6854	Grass	–
L. lactis ssp. *cremoris* DPC6858	Grass	+
L. lactis ssp. *cremoris* DPC6857	Grass	+
L. lactis ssp. *cremoris* DPC6856	Rumen	–
L. lactis ssp. *lactis* DPC6853	Corn	–
L. lactis ssp. *cremoris* HP	Dairy	+
L. lactis ssp. *cremoris* IE16	Dairy	–
L. lactis ssp. *cremoris* MG1363	Dairy	–
L. lactis ssp. *lactis* IL1403	Dairy	–
L. lactis SMQ562	Dairy	–
L. lactis ssp. *cremoris* UC509	Dairy	–
L. lactis ssp. *lactis* biovar diacetylactis F7/2	Dairy	–
L. lactis ssp. *lactis* ML8	Cheese	+
L. lactis ssp. *cremoris* H88M1	Raw milk	–
L. lactis ssp. *lactis* 303	Cheese	–
L. lactis HT-2	Dairy	–
L. lactis ssp. *cremoris* SK11	Cheese	–
L. lactis U53	Dairy	+
L. lactis ssp. *lactis* biovar diacetylactis DRC3	Dairy	–
L. lactis 83	Dairy	–
L. lactis BA1	Dairy	–
L. lactis ssp. *lactis* biovar *diacetylactis* 938	Dairy	–
L. lactis ssp. *cremoris* 310	Cheese	+
L. lactis ssp. *lactis* 229	Cheese	–
Lactobacillus plantarum	Grass	–
Lactobacillus brevis	Grass	–

Table 3. Predicted open reading frames (ORFs) of ϕL47[a, b] and predicted database matches.

ORF	Start	Stop	MW (kDa)	Best match	% ID	E value	Accession number[c]
1	146	1165	39.0	Putative integrase-recombinase [*Lactococcus* ϕ949]	324/339 (96)	0.0	YP_004306161.1
2	1174	3495	87.6	Putative exodeoxyribonuclease *[Lactococcus* ϕ949]	707/772 (92)	0.0	YP_004306162.1
4	4260	4949	24.0	Putative transglycosylase [*Lactococcus* ϕ949]	195/231 (84)	4e-132	YP_004306164.1
5	4963	5517	21.4	Putative uncharacterized protein [Firmicutes bacterium CAG:882]	29/89 (32)	1.8	CDD64083.1
12	7633	8229	22.7	Putative thymidine kinase [*Lactococcus* ϕ949]	184/196 (94)	3e-134	YP_004306171.1
14	8674	8817	5.55	Hypothetical protein, unlikely [*Trypanosoma congolense* IL3000]	16/33 (48)	5.8	CCD16863.1
15	8841	8951	5.46	Transglycosylase-like domain protein [*Streptococcus ictaluri*]	18/44 (41)	0.32	WP_008087686.1
23	10569	10796	8.43	Phage protein [*L. lactis* ssp. *lactis* KF147]	35/64 (55)	5e-18	YP_003353508.1
24	10814	11191	14.0	Hypothetical protein PelgB_21892 [*Paenibacillus elgii* B69]	61/122 (50)	1e-33	ZP_09077133.1
26	11447	11770	12.4	Hypothetical protein LACR_2119 [*L. lactis* ssp. *cremoris* SK11]	81/120 (68)	4e-46	YP_811707.1
27	11770	12045	10.1	Hypothetical protein ul36_25 [*Lactococcus* ϕul36]	50/84 (60)	3e-20	NP_663659.1
28	12038	12367	12.5	ORF23 [*Lactococcus* ϕTP901-1]	39/89 (44)	9e-8	NP_112686
29	12364	12609	88.9	Hypothetical protein Q33_0026 [*Lactococcus* ϕQ33]	77/81 (95)	3e-50	AFV51055
30	12644	12835	7.07	Hypothetical protein [*Lactococcus* ϕp2]	38/64 (59)	1e-14	ACI94897.1
31	12912	13415	19.8	Hypothetical protein [*Limnohabitans* sp. Rim47]	20/93 (22)	0.32	WP_019430841.1
32	13415	13630	8.27	Hypothetical protein [*Methyloversatilis universalis*]	15/48 (31)	2.0	WP_018228448.1
35	14561	14914	13.9	Orf28 [*Lactococcus* ϕbIL286]	35/108 (32)	0.37	NP_076662.1
37	15241	15360	4.67	Hypothetical protein [*L. lactis* ssp. *lactis* CV56]	24/35 (69)	1e-06	YP_005869010.1
39	17948	17772	6.99	Unknown			
40	18472	18083	15.0	Hypothetical protein [*Enterococcus* ϕSAP6]	46/118 (39)	6e-11	AEM24750.1
44	19772	19539	9.23	Hypothetical protein ACD_5C00356G0007 [uncultured bacterium]	18/61 (30)	1.3	EKE24920.1
45	20124	19852	10.9	Hypothetical protein ECED1_3531 [*Escherichia coli* ED1a]	24/69 (35)	0.002	YP_002399391.1
46	20281	20141	5.35	Unknown			
47	20856	20716	4.52	Unknown			
48	21121	20867	9.73	Hypothetical protein EFP_gp119 [*Enterococcus* ϕEF24C]	45/79 (57)	1e-19	YP_001504228.1
49	21414	21178	9.60	Hypothetical protein [*Enterococcus faecium*]	38/74 (51)	2e-14	WP_002350596.1
52	22318	22007	11.7	Hypothetical protein EFP_gp116 [*Enterococcus* ϕEF24C]	29/ 72 (40)	4e-12	YP_001504225.1

53	22680	22411	10.4	Hypothetical protein LmalK35_10926 [*Lactobacillus mali* KCTC 3596]	25/ 75 (33)	5e-04	ZP_09449152.1
54	23092	22751	13.3	LamB/YcsF family protein [*Mesoflavibacter zeaxanthinifaciens* S86]	28/ 122 (23)	1.6	ZP_09499326.1
55	25273	25410	5.59	Hypothetical protein ARALYDRAFT_900440 [*Arabidopsis lyrata*]	16/40 (40)	0.63	XP_002886300.1
57	25619	25768	5.61	Zinc finger protein 263 [*Monodelphis domestica*]	17/42 (40)	0.34	XP_001370637.2
58	25765	26346	22.6	Hypothetical protein [*Streptococcus henryi*]	90/187 (48)	2e-44	WP_018165624.1
60	26593	26979	15.8	Hypothetical protein CLJ_0158 [*C. botulinum* Ba4 str. 657]	39/130 (30)	5e-07	YP_002860465.1
61	27057	27260	7.92	Hypothetical protein CAEBREN_29888 [*Caenorhabditis brenneri*]	19/53 (36)	0.45	EGT53499.1
62	27257	27691	16.8	Hypothetical protein EfaCPT1_gp44 [*Enterococcus* ϕEfaCPT1]	41/139 (29)	6e-04	AFO10841.1
64	28027	28230	7.90	Hypothetical protein LPV712_gp041 [*Lactococcus* ϕ712]	34/66 (52)	4e-13	YP_764301.1
66	28820	29419	23.2	Putative nucleotidase [*Lactococcus* ϕ949]	190/199 (95)	5e-136	YP_004306212.1
69	30662	31708	37.8	Putative endolysin [*Lactococcus* ϕ949]	311/348 (89)	0.0	YP_001469072.1
72	33459	33629	6.35	Hypothetical protein xccb100_2160 [*Xanthomonas campestris* pv. *campestris* str. B100]	15/53 (28)	1.5	YP_001903564.1
75	35411	36955	59.2	Putative replicative DNA helicase [*Lactococcus* ϕ949]	481/514 (94)	0.0	YP_004306220.1
76	36966	38015	40.5	Putative DNA primase [*Lactococcus* ϕ949]	302/329 (92)	0.0	YP_004306221.1
77	38022	39806	68.0	Putative single-stranded DNA specific exonuclease [*Lactococcus* ϕCB13]	566/593 (95)	0.0	ACU46843.1
80	40336	40815	18.2	Membrane protein containing HD superfamily hydrolase domain, YQFF ortholog [*Caloramator australicus* RC3]	27/71 (38)	1.1	CCC58155.1
82	41311	41496	7.74	Unknown			
83	41493	41636	4.99	Unknown			
84	41639	41854	8.84	GJ15471 [*Drosophila virilis*]	20/54 (37)	0.34	XP_002060057.1
85	41848	42081	9.16	Conserved hypothetical protein [*Anaerococcus tetradius* ATCC 35098]	16/54 (30)	1.3	ZP_03930743.1
86	42081	42458	14.4	GA28089 [*Drosophila pseudoobscura pseudoobscura*]	21/54 (39)	0.44	XP_002133330.1
88	42897	43262	14.4	Putative phage protein [*L. lactis* ssp. *lactis* A12]	60/99 (61)	1e-24	CDG04128.1
89	43259	43648	14.6	Putative replication initiator protein [*L. lactis* ssp. *lactis* IO-1]	108/129 (84)	3-75	YP_007508915.1
90	43695	44000	12.2	Hypothetical protein [*L. lactis* ssp. *lactis* CV56]	36/105 (34)	6e-07	YP_005868381.1
92	44413	44592	7.14	Catechol 1,2-dioxygenase [*Pseudomonas stutzeri* DSM 10701]	23/53 (43)	3.3	AFN77848.1

100	46959	47195	8.87	Related to Beta-mannosidase precursor [*Sporisorium reilianum* SRZ2]	15/41 (37)	4.6	CBQ72430.1
105	49552	49986	16.8	Hypothetical protein ROI_04910 [*Roseburia intestinalis* M50/1]	21/65 (32)	0.058	CBL07781.1
110	55222	57189	73.4	Putative DNA gyrase subunit B-topoisomerase [*Lactococcus* ϕ949]	629/651 (97)	0.0	YP_004306254.1
111	57191	59305	79.22	Putative DNA gyrase subunit A-topoisomerase [*Lactococcus* ϕ949]	675/704 (96)	0.0	YP_004306255.1
114	59918	60436	19.0	Putative nucleoside-2-deoxyribosyltransferase [*Lactococcus* ϕ949]	166/172 (97)	1e-114	YP_004306259.1
115	60429	61088	24.9	Putative nicotinamide mononucleotide transporter [*Lactococcus* ϕ949]	205/219 (94)	3e-142	YP_004306260.1
116	61168	63669	95.7	Putative anaerobic ribonucleoside-triphosphate reductase [*Clostridium* ϕ8074-B1]	268/791 (34)	2e-134	YP_007237286.1
117	63669	64265	22.7	Putative ribonucleoside triphosphate reductase activator small subunit [*Lactococcus* ϕ949]	167/198 (84)	2e-126	YP_004306263.1
118	64356	65258	35.4	Hypothetical group I intron protein [*Lactococcus* ϕ949]	166/238 (70)	8e-111	YP_004306264.1
119	65327	66550	45.8	gp109 [*Lactococcus* ϕKSY1]	151/389 (39)	1e-69	YP_001469108.1
122	67083	67322	9.16	Putative ribonucleotide diphosphate reductase glutaredoxin subunit[*Lactococcus* ϕ949]	71/79 (90)	3e-44	YP_004306267.1
123	67315	67686	14.1	Putative ribonucleotide reductase stimulatory protein [*Lactococcus* ϕ949]	121/123 (98)	2e-79	YP_004306268.1
124	67700	69889	82.8	Ribonucleotide-diphosphate reductase subunit alpha [*Bacillus alcalophilus*]	465/715 (65)	0.0	WP_003323097.1
125	69902	70312	16.0	Hypothetical protein TTHERM_00313410 [*Tetrahymena thermophila*]	19/46 (41)	0.48	XP_001033522.1
126	70395	70991	23.0	Hypothetical group I intron protein [*Lactococcus* ϕ949]	89/126 (71)	3e-60	YP_004306264.1
127	71101	71751	25.7	Hypothetical protein [*L. lactis* ssp. *cremoris* A76]	89/208 (43)	2e-49	YP_005876546.1
128	71773	72513	29.3	ATPase [*Solibacillus silvestris* StLB046]	70/229 (31)	8e-18	YP_006461965.1
129	72726	72872	5.74	Unnamed protein product [*Tetraodon nigroviridis*]	16/40 (40)	1.2	CAG05661.1
130	72869	73153	10.7	M23 peptidase domain-containing protein [*C. perfringens* F262]	29/83 (35)	0.46	EIA16337.1
131	73169	73390	8.71	Hypothetical protein LP030nr2_047 [Listeria ϕLP-030-2]	19/37 (51)	3e-04	YP_008126743.1
132	73392	73787	15.5	Uncharacterized protein BN765_00487 [*Eubacterium eligens* CAG:72]	27/91(30)	3.1	CDA41513.1
133	73797	74150	13.8	Hypothetical protein CRE_21750 [*Caenorhabditis remanei*]	25/79 (32)	1.2	XP_003105387.1
134	74153	74335	7.15	Hypothetical protein GY4MC1_0607 [*Geobacillus* sp. Y4.1MC1]	15/48 (31)	1.5	YP_003988046.1

135	74404	75114	27.3	Hypothetical protein [*Bacillus amyloliquefaciens* TA208]	85/238 (36)	3e-36	YP_005541036.1
136	75475	76098	22.3	Diphosphate reductase [*Streptococcus ovis*]	131/184 (71)	5e-88	WP_018378302.1
137	76177	76746	17.0	Hypothetical protein [Lachnospiraceae bacterium A4]	35/142 (25)	0.92	WP_016282214.1
138	76752	77021	22.0	Hypothetical protein llmg_2104 [*L. lactis* ssp. *cremoris* MG1363]	27/ 90 (30)	0.008	YP_001033358.1
139	77047	77322	10.6	Hypothetical protein [Clostridium sp. 7_2_43FAA]	24/84 (29)	7.2	WP_008676880.1
140	77312	78226	10.6	Putative ribonucleotide diphosphate reductase alpha subunit [*Lactococcus* ϕ949]	278/ 298 (93)	0.0	YP_004306269.1
141	78305	78985	32.9	Putative Guanylate kinase [*Lactococcus* ϕ949]	198/ 226 (88)	6e-144	YP_004306270.1
142	78963	79706	27.8	Putative deoxyuridine 5'-triphosphate nucleotidohydrolase [*Lactococcus* ϕ949]	239/247 (97)	1e-176	YP_004306271.1
143	79803	80564	28.1	Putative protease [*Lactococcus* ϕ949]	227/ 253 (90)	1e-167	YP_004306272.1
147	82174	85336	122.3	Putative DNA polymerase III alpha subunit [*Lactococcus* ϕ949]-truncated	641/ 692 (93)	0.0	YP_004306276.1
148	85420	85596	65.1	Putative uncharacterized protein [Odoribacter sp. CAG:788]	14/32 (44)	9.4	WP_021989169.1
153	87639	86959	26.1	Putative Holliday junction resolvase [*Lactococcus* ϕ949]	208/ 226 (92)	2e-130	YP_004306283.1
154	88100	87693	14.7	Putative phage holin [*Lactococcus* ϕ949]	118/ 135 (87)	1e-77	YP_004306284.1
157	90418	88586	65.9	Putative tail protein [*Lactococcus* ϕ949]	468/ 616 (76)	0.0	YP_004306287.1
158	90891	90418	17.5	Putative tail protein [*Lactococcus* ϕ949]	150/157 (96)	5e-93	YP_004306288.1
159	91768	90932	31.4	Putative tail protein [*Lactococcus* ϕ949]	134/232 (58)	2e-75	YP_004306288.1
160	92858	91731	41.8	Host-specificity [*Lactococcus* ϕbIL286]	136/228 (60)	3e-67	NP_076691.1
161	93009	92836	6.4	Orf56 [*Lactococcus* ϕbIL286]	52/57 (91)	1e-29	NP_076690.1
162	93704	93009	27.4	Orf55 [*Lactococcus* ϕbIL286]	209/228 (92)	4e-141	NP_076689.1
163	94137	93751	13.2	Orf55 [*Lactococcus* ϕbIL286]	120/128 (94)	5e-62	NP_076689.1
164	98313	94153	156.2	Putative tail protein [*Lactococcus* ϕ949]	1088/ 1309 (83)	0.0	YP_004306288.1
165	99184	98420	28.6	Putative phage tail component [*Lactococcus* ϕ949]	236/ 254 (93)	3e-169	YP_004306289.1
166	109063	99236	357.7	Putative tail tape measure protein [*Lactococcus* ϕ949]	2780 /3276 (85)	0.0	YP_004306290.1
167	110073	109081	38.4	Putative site specific recombinase [*Lactococcus* ϕ949]	326/ 330 (99)	0.0	YP_004306291.1
168	110552	110070	18.7	Putative phage structural protein [*Lactococcus* ϕ949]	153/ 160 (96)	7e-111	YP_004306292.1
172	112890	112189	25.1	Putative phage structural protein [*Lactococcus* ϕ949]	197/ 234 (84)	4e-139	YP_004306296.1
174	114260	113655	22.9	Putative phage structural protein [*Lactococcus* ϕ949]	186/ 201 (93)	6e-136	YP_004306298.1
175	115012	114260	28.5	Putative phage structural protein [*Lactococcus* ϕ949]	240/ 250 (96)	2e172	YP_004306299.1
177	115749	115309	16.9	Putative phage structural protein [*Lactococcus* ϕ949]	143/146 (98)	3e-99	YP_004306301.1

179	117513	116494	38.3	Putative phage structural protein [*Lactococcus* ϕ949]	318/338 (94)	0.0	YP_004306303.1
180	118088	117552	19.3	Putative phage structural protein [*Lactococcus* ϕ949]	125/178 (70)	1e-82	YP_004306304.1
181	119358	118117	46.9	Putative phage structural protein [*Lactococcus* ϕ949]	401/417 (96)	0.0	YP_004306305.1
182	120911	119370	58.1	Putative phage structural protein [*Lactococcus* ϕ949]	490/513 (96)	0.0	YP_004306306.1
183	122632	120935	64.9	Putative terminase ATPase subunit [*Lactococcus* ϕ949]	558/565 (99)	0.0	YP_004306307.1
185	125008	123761	47.6	Putative DNA polymerase [*Lactococcus* ϕ949]	397/415 (96)	0.0	YP_004306309.1
186	126342	125017	50.9	Putative phosphate starvation-inducible protein PhoH/ATPase [*Lactococcus* ϕ949]	337/441 (76)	0.0	YP_004306310.1

[a] Hypothetical proteins with >60% identity to ϕ949 have been omitted

[b] ORFs shaded grey possess E-values >1

[c] Accession numbers corresponding to the NCBI database

Abbreviations: MW, Molecular weight; %ID, % Identity; E-value, Expect-value

Table 4. tRNA arrangement in ϕL47.

tRNA	Amino acid	Anti-codon	Size (bp)	% GC	Start	End
1	Met	CAT	73	47.9	48524	48596
2	Arg	TCT	73	39.7	50343	50415
3	Asp	GTC	75	50.7	51213	51287
4	Ala	CGC	71	50.7	51639	51709
5	Ser	ACT	92	43.5	51967	52116
6	Asn	ATT	76	39.5	52830	52905
7	Trp	CCA	71	39.4	53114	53184
8	Met	CAT	74	55.4	53400	53473

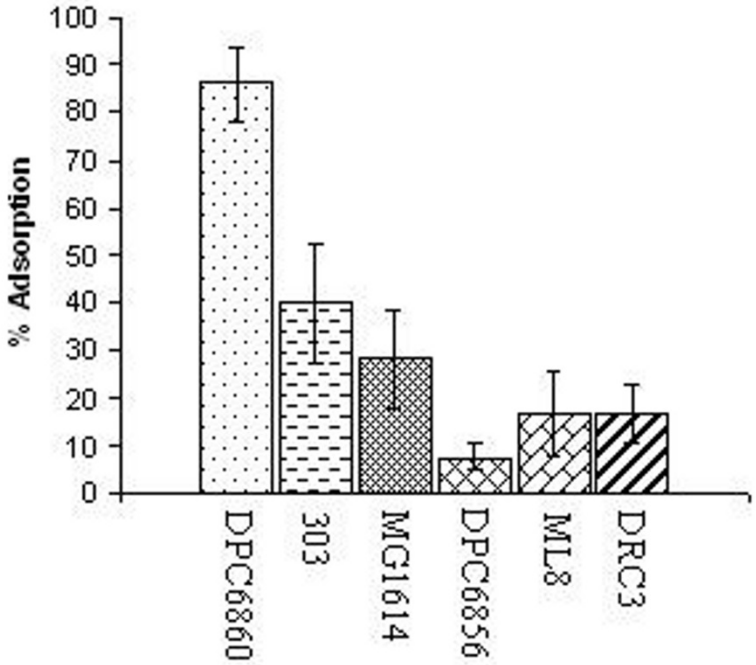


Fig. 1. Relative % adsorption of ϕL47 to *Lactococcus* strains of different environmental niches and subspecies designation.

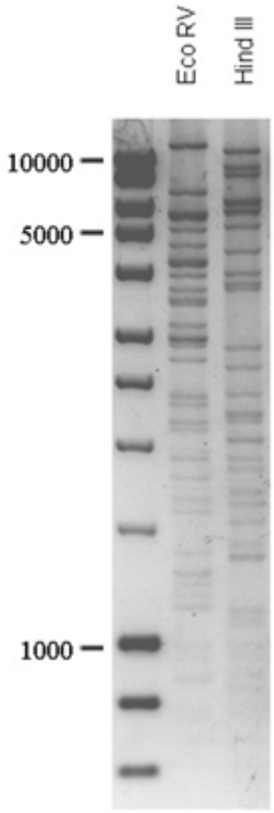


Fig. 2. Restriction digest profiles of phages ϕL47 using the endonuclease restriction enzymes EcoRV and HindIII. Digested DNA was run a 0.7% (wt./ vol.) agarose at 50 V overnight at 4 ^{o}C.

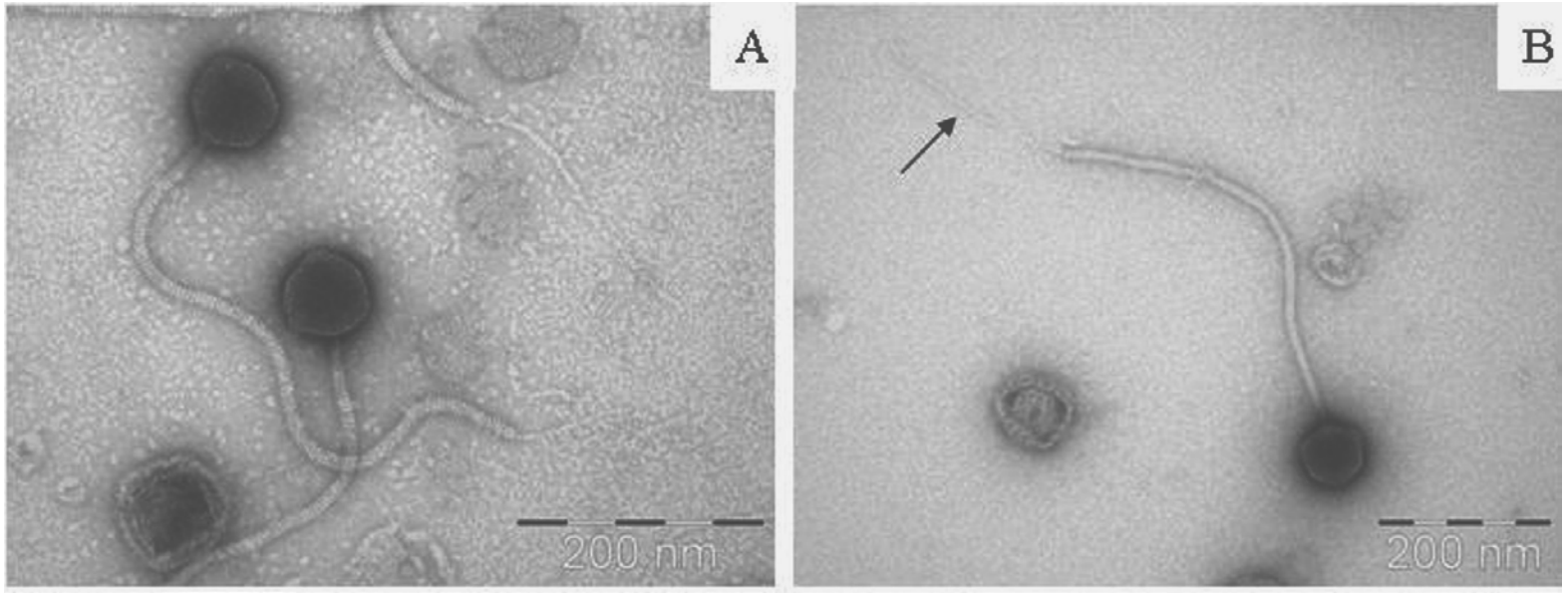
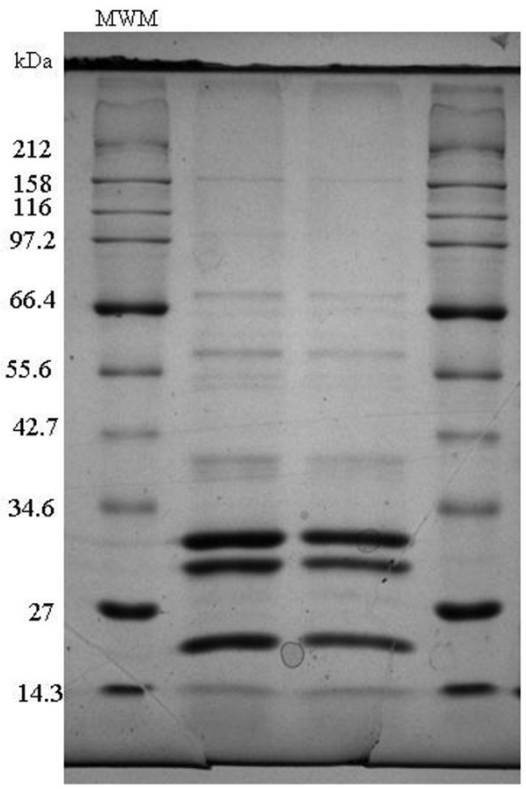


Fig. 3. (A) Scanning electron micrograph image of ϕL47 negatively stained using 2% uranyl acetate. (B) Single view of ϕL47. Arrow shows the elongated tail fiber.



Putative function	Predicted molecular weight from geome analysis (kDa)	ORF
Tape measure protein	357.8	166
Tail protein	156.2	164
–		
Tail protein	65.9	157
Structural protein	58.1	182
Structural protein	46.9	181
Host specificity	41.8	160
Structural protein	38.3	179
Base plate	31.4	159
Tail component	28.6	165
Structural	28.5	175
Structural	25.1	172
Structural protein	22.9	174
Structural protein	19.3	180
Structural protein	18.7	168
Structural protein	17.5	158
Structural protein	16.9	177

Fig. 4. Analysis of the structural proteins of ϕL47 obtained from ammonium acetate precipitations. A broad range ladder (New England Biolabs) was used as a relative molecular weight marker (MWM). Molecular masses of structural proteins was predicted from genome sequence data using the Compute pI/ Mw tool (Gasteiger *et al*., 2005).

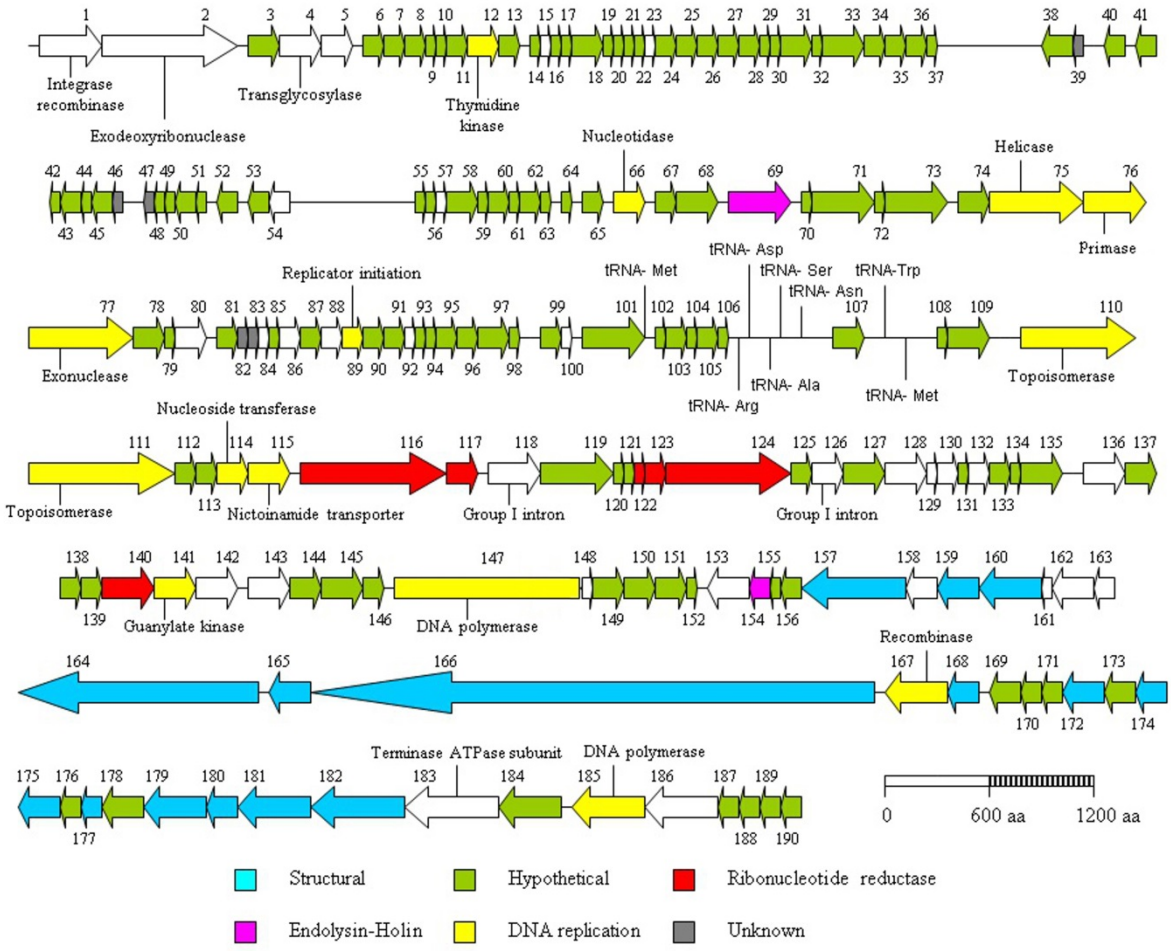


Fig. 5. Genome organisation of the lactococcal phage ϕL47. Each arrow represents an opening reading frame (ORF), the orientation of which shows the direction of transcription. The predicted function of each ORF was determined by bioinformatic analyses (Carver *et al.*, 2005).

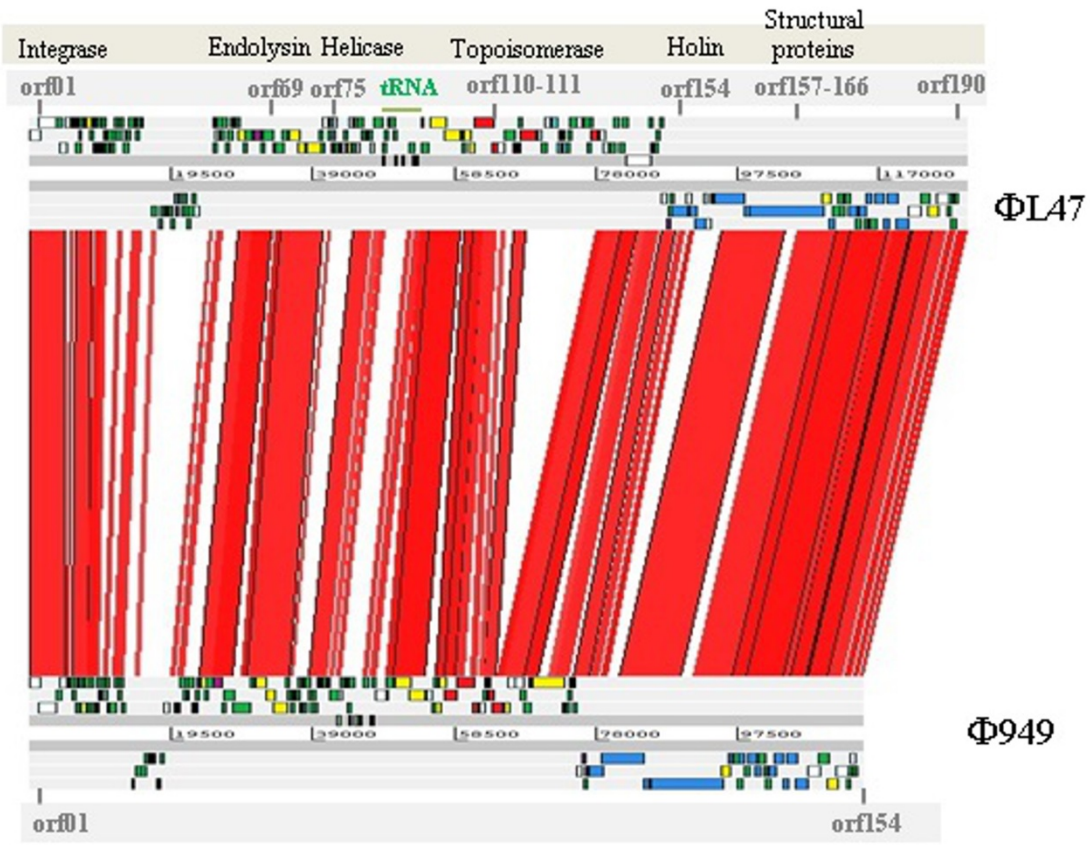


Fig. 6. BLASTn-based alignment (nucleotide identity >90% shown) of the genome sequences of *L. lactis* phage ϕL47 and ϕ949 as displayed by Artemis comparison tool (ACT). Red lines between genomes indicate orthologues in the same orientation. Coding sequences are colour coded as described in Fig. 4; structural (blue), hypothetical (green), DNA replication (yellow), ribonucleotide reductase (red), endolysin-holin (mauve), unknown (grey).

Chapter 5

Location, location, location:

examination of niche specialisation of *Lactococcus lactis* through comparative genome analysis

5.1. Abstract

Lactococcus lactis is a member of the lactic acid bacteria (LAB) and is used extensively in the production of fermented dairy foods such as cheese and sour cream. Although it is associated with dairy fermentations, it is hypothesised that lactococcal strains in use today are descended from plants. To identify and characterise specific genes and gene sets which may be important in niche adaptation, comparative genome analysis was performed with *L. lactis* isolates from corn (DPC6853), from the bovine rumen (DPC6856) and from grass (DPC6860), and compared to other lactococcal isolates from dairy and non-dairy environments. Non-conserved genes or gene clusters were identified through a combination of functional genome distribution (FGD) analysis (Altermann, 2012) and manual curation with Artemis (Rutherford *et al.,* 2000). Systems which are typically found in dairy organisms, such as genes encoding lactose utilisation, were identified in all non-dairy strains. Genes encoding defences to bacteriophages were identified in non-dairy strains which may contribute to resistance or insensitivity to dairy lactococcal phages as observed in Chapter 4. Various phenotypic and genotypic differences were found between DPC6860 and DPC6853 such as cobalt sensitivity and capacity to ferment xylose, suggesting that DPC6856 is not transient but a common rumen inhabitant. DPC6856 appears to possess specific genes/ gene clusters which may contribute to its survival in the rumen environment. Overall DPC6860 appears to be more metabolically robust. This study highlights the contribution of niche adaptation to intra-species diversity and capacity of *L. lactis* to evolve to its environment.

5.2. Introduction

Lactococcus lactis is a member of the lactic acid bacteria (LAB) and a key component of starter systems used in the production of fermented dairy products such as cheese and sour cream. It is widely held that dairy strains, commonly used in modern cheese production, originated from plant material (Kelly *et al.*, 2010). Genome sequence analysis of *L. lactis* has revealed that adaptation to the dairy environment is promoted by the loss of dispensable genes involved in amino acid biosynthesis and plant sugar utilisation, and the acquisition of genes enabling lactose fermentation and casein utilisation (Bolotin *et al.*, 2001; Wegmann *et al.*, 2007; Kelly *et al.*, 2010). In contrast to their dairy counterparts, strains from non-dairy environments exhibit diverse metabolic capabilities and lifestyle characteristics required for niche specialisation. The non-dairy strain *L. lactis* A12 is capable of fermenting raffinose and arabinose, which reflects the adaptation of this strain to the sourdough environment where it rapidly uses sugars which are more abundant due to their slow utilisation by yeasts and lactobacilli (Passerini *et al.*, 2013). Another non-dairy strain *L. lactis* KF147, isolated from mung bean sprouts, possesses numerous gene clusters which may enable it to survive in its environment, such as EPS production and utilisation of plant carbohydrates (Siezen *et al.*, 2008). Furthermore, comparative analysis of LAB from different environments has also identified gene clusters which could be assigned to a specific niche (O' Sullivan *et al.*, 2009).

Owing to the limited number of established dairy strains used in production today (Marshall, 1991), there is an increasing demand for novel strains with concerted efforts to mine the microbiota of natural environments, such as raw milk and plants, for strains of technological interest (Klijn *et al.*, 1995; Kelly *et al.*, 1998; Nomura *et al.*, 2006). In the United States, the documented safe use of lactic acid bacteria (LAB)

in food has led to many species being designated as generally regarded as safe (GRAS) including *L. lactis*. Although species from diverse environments may possess GRAS status, they may nevertheless encode specific genes, which although may be considered niche adaptation factors (Hill, 2012), could also contribute to these organisms being unsuitable for food applications. Therefore strains isolated from diverse environments with potential for food applications should be carefully assessed prior to their use in food production.

In this study, we report on the whole genome sequencing of three *L. lactis* isolates from different non-dairy environments, i.e. grass (DPC6860), bovine rumen (DPC6856) and corn (DPC6853). Through comparative genome analysis, the sequences were examined for non-conserved ORFs which may confer a competitive advantage to the strains in their respective environments. In particular, traits which may be of technological interest in dairy fermentations and those which may play an important role in niche adaptation were assessed. Specific traits present in DPC6856 were also examined which may be important in the adaptation of this strain to the bovine rumen.

5.3. Materials and methods

5.3.1. Bacterial strains and growth conditions

Lactococcus lactis strains used in this study were stored at -80 ^{o}C in M17 (Oxoid, Hampshire, England) containing 50% (vol./ vol.) glycerol supplemented with the appropriate sugar: 0.5% (wt./ vol.) glucose (GM17) (Sigma Aldrich, Dublin, Ireland) for MG1363 and IL1403 and 0.5% (wt./ vol.) lactose (LM17) (VWR, Leuven, Belgium) for the remaining strains. Prior to use, strains were cultured on GM17 or LM17 agar plates as required and incubated aerobically at 30 ^{o}C for 48 hrs. Individual colonies were suspended in GM17 or LM17 and cultured overnight at 30 ^{o}C under aerobic conditions for analysis.

5.3.2. Plasmid profile analysis

Plasmid DNA was isolated according to the method described by O'Sullivan and Klaenhammer (1993) from 2 ml overnight cultures. Plasmid DNA was run on a 0.7% (wt./ vol.) agarose gel for 18 hrs at 4 ^{o}C containing 200 ng/ ml of ethidium bromide (Sigma Aldrich, Dublin, Ireland). DNA was visualised using a transilluminator and an image captured using a Kodak photo imager.

5.3.3. Draft genome sequencing and annotation

Based on MLSA performed in chapter 2, three strains were selected for genome sequencing: *L. lactis* ssp. *cremoris* DPC6856 (bovine rumen) and DPC6860 (grass), and *L. lactis* ssp. lactis DPC6853 (corn). Whole genome sequencing was performed using the Roche 454 sequencing platform as described in Chapter 2. Contigs were assembled using the Seqman NGen program from the DNAstar software suite (Madison, Wisconsin, USA). Contigs less than 1,000 nucleotides were removed from

analysis. Draft assemblies were annotated using an updated version of the GAMOLA software package (Altermann and Klaenhammer, 2003). Briefly, individual contigs were concatenated with a non-bleeding spacer sequence and open reading frames (ORFs) were predicted using Prodigal (Hyatt *et al.*, 2010) and RAST (Aziz *et al.*, 2008). Each ORF was then subjected to BlastP (standalone BLAST (Altschul *et al.*, 1990), non-redundant NCBI database), COG (updated 2008 COG database release), Pfam (HMMER3 (Eddy, 2009), database release 27), and TIGRfam (database release 14.0) analyses to predict biological function. Results were then combined into an auto-annotation in Genbank format. Contig mapping was performed using Mauve against the reference genomes *L. lactis* IL1403 and *L. lactis* MG1363 (Darling *et al.*, 2010). Contigs which did not map with the reference genomes Il1403 and MG1363 were manually curated using the Artemis program (Rutherford *et al.*, 2000) following ORF calling.

5.3.4. Mining of non-conserved ORFs

Nineteen publicly available complete and draft *L. lactis* genomes were downloaded in Genbank or FASTA format from the NCBI genome database. Draft genomes were automatically concatenated with a non-bleeding spacer sequence and the respective gene models predicted and saved in Genbank format using GAMOLA2 as described above. All 19 *Lactococcus* genomes were then subjected to FGD analysis (Altermann, 2012) using compACTor v0.12. The 19 ORFeomes were compared to each other in a round-robin analysis using individual BLASTp analyses. Genome clusters were selected for further analysis using the FGDfinder v0.013 tool. Briefly, two clusters of selected genomes were compared to each other by identifying cluster conserved ORFs at e-value thresholds of 1e-40 and 1e-20, respectively. Those

conserved ORFs were then compared to the ORFeomes of the respective subject genome cluster and ORFs without homologs above e-value levels of 1e-10 were considered cluster specific. A maximum of one and two mismatches per comparison between clusters were allowed for. The resulting ORF-sets were then analysed within their respective genomic context for biological function. Two separate comparisons were performed using FGD, with DPC strains as a single cluster compared to dairy strains (UC509.9, HP^T, LD61, A76, SK11, CNCM I-1631, TIFN7, TIFN2, TIFN5, TIFN3, TIFN1, TIFN4) and both dairy and non-dairy strains (KF147, IO-1, CV56, KW2, A12, NCDO2118, KLDS 4.035 UC509.9, HPT, LD61, A76, SK11, CNCM I-1631, TIFN7, TIFN2, TIFN5, TIFN3, TIFN1, TIFN4).

5.3.5. Minimal media to assess carbohydrate fermentation

Modified M17 medium was made by the omission of meat extract and any additional sugar source. Controlled defined M17 (CDM17) was subsequently used as a medium to examine the growth of dairy and non-dairy *Lactococcus* strains in the presence of different carbohydrate substrates. CDM17 contained the following as g/ L: tryptone (Sigma Aldrich) 5.0, bacteriological peptone (Oxoid) 5.0, yeast extract (Sigma Aldrich) 2.5, ascorbic acid (Sigma Aldrich) 0.5, magnesium sulphate (Sigma Aldrich) 2.5, disodium-β-glycerophosphate (Sigma Aldrich) 19. The pH of the media was adjusted to between 6.3 and 6.5 and sterilised by autoclaving at 121 ^oC for 15 mins. Growth of *Lactococcus* strains in CDM17 without the addition of a sugar did not exceed an OD_{600nm} of 0.25 after 48 hrs.

5.3.6. Carbohydrate fermentation

Fermentation of carbohydrates was assayed by optical density (OD) measurements using a Synergy 2 plate reader (BioTek Instruments Inc., Vermont,

USA). Carbohydrate solutions were prepared by the addition of a carbohydrate (1% wt./ vol.) to the CDM17 medium and filter-sterilised using a 0.45 µm filter (Sarstedt, Wexford, Ireland). 200 µl of supplemented CDM17 was dispensed into a 96 well micro-titre plate and inoculated with 1% (vol./ vol.) of overnight bacterial culture grown in LM17 at 30 ºC. The inoculated samples were grown at 30 ºC and OD readings were taken at 48 hrs. Each carbohydrate was examined in triplicate for duplicate experiments and the average of both experiments used for analysis. The scoring of sugar fermentation was as follows: >0.8 OD_{600nm} ++++; 0.5-0.8 OD_{600nm} +++; 0.2-0.5 OD_{600nm} ++; 0.1-0.2 OD_{600nm} +; <0.1 OD_{600nm} −.

5.3.7. Growth inhibition

Insensitivity to Cadmium (Cd^{2+}) and Cobalt (Co^{2+}) was determined by measuring OD using a 96 well micro-titre plate (Sarstedt, Wexford, Ireland. Strains were grown in 200 µl volumes containing increasing concentrations of $CdCl_2$ and $CoCl_2$ from 0 to 0.6 mM and 0 to 6 mM respectively, at 30 ºC. Following inoculation at 1% with cultures grown in the absence of heavy metals, growth was monitored over 24 hrs with absorbance measured using a Synergy 2 plate reader (BioTek Instruments Inc., Vermont, USA). Growth at 6 ºC was monitored over 20 days, measuring OD using a 96 well micro-titre plate as above.

5.3.8. Lysine decarboxylase activity

Lysine decarboxylase broth was used to determine whether non-dairy strains could produce cadaverine from lysine, which was prepared as described previously (Falkow, 1958). Briefly, the solution was composed of gelatin peptone (5 g/ L), L-lysine (5 g/ L), yeast extract (3 g/ L), glucose (1 g/ L) and bromocresol purple (0.02

g/ L) which was autoclaved at 121 ^{o}C for 15 mins. Prior to use, the solution was stored at 2-8 ^{o}C. Bacteria were added to 10 ml of broth to create a 1% inoculum and incubated at 30 ^{o}C for 48 hrs. A yellow colour was indicative of a negative reaction and a purple colour for a positive reaction.

5.4. Results and discussion

The selective pressure imparted by a specific environment is a key driver in the genomic diversity observed between strains of the same species originating from different habitats. Previously, the tandem use of phenotypic and genotypic data has successfully identified novel genes and gene clusters in plant isolates absent in dairy *L. lactis* strains (Siezen *et al.*, 2008). Due to advances in high throughput sequencing and reductions in sequencing costs, it has now become feasible to sequence whole genomes of different strains of the same species with a view to evaluating the diversity between strains. In all, the NCBI database lists 23 sequenced *L. lactis* genomes [accessed March 2014] with an increasing number of genomes of strains isolated from outside the dairy environment, including strain A12 from sourdough bread (Passerini *et al.*, 2013), IO-1 from drain water (Kato *et al.*, 2012), CV56 from human vaginal samples (Gao *et al.*, 2011) and KW2 from fermented corn (Kelly *et al.*, 2013). Bioinformatic tools such as functional genome distribution (FGD) have been developed (Altermann, 2012) enabling the identification of non-conserved ORFs between groups of strains based on amino acid similarity, while also taking into account genetic alterations associated with niche adaptation. More simply, the analysis of non-alignment contigs with reference genomes can be useful in the identification of genetic islands which may be important in adaptation to a specific environment (Siezen *et al.*, 2008).

5.4.1. Genome comparison of DPC strains from grass, corn and the bovine rumen

Draft genome sequences were generated for three non-dairy *L. lactis* strains isolated from grass (DPC6860), corn (DPC6853) and the bovine rumen (DPC6856). Following quality control and assembly, 116 contigs were generated for DPC6853,

163 contigs for DPC6856 and 104 contigs for DPC6860 (Table 1). The estimated G+C content of strains DPC6853, DPC6856 and DPC6860 were approximately 35%, 35.3% and 35.6% respectively, comparable to those of sequenced *L. lactis* strains (34.8%-36.7%). FGD was performed on DPC strains in comparison to: (i) dairy strains (UC509.9, HP^T, LD61, A76, SK11, CNCM_I-1631, TIFN7, TIFN2, TIFN5, TIFN3, TIFN1, TIFN4) and (ii) dairy and non-dairy strains (KF147, IO-1, CV56, KW2, A12, NCDO2118, KLDS 4.035) to identify ORFs specific to DPC strains. For the purposes of our analysis, non-conserved ORFs refers to ORFs that were identified in the group comprising the non-dairy DPC strains, but were absent from the groups containing dairy lactococcal strains or all lactococcal strains (dairy and non-dairy). Mining of non-conserved genes within the genomes of the DPC strains using FGD with two mismatches, identified 310 ORFs which were absent from the dairy lactococcal group, the majority of which were identified as hypothetical proteins or phage-associated proteins. Similarly, comparison between the three DPC strains and all sequenced lactococcal strains (both dairy and non-dairy), using two mismatches, identified 237 non-conserved ORFs. Strains IL1403 and MG1363 were omitted from analysis due to the plasmid cured nature of these strains and the loss of associated phenotypes, which may be important in niche adaptation. *L. lactis* Dephy1 was excluded from analysis as the origin of the strain is unknown. In tandem with this analysis, the use of Artemis aided in the identification of novel gene clusters. Many of the non-conserved ORFs which were identified in DPC strains were located on non-alignment contigs which suggests these genes are plasmid encoded or may correspond to genetic islands. For detailed annotation of contigs which did not align with reference genomes MG1363 and IL1403 see section 5.8, appendices 5-1, 5-2 and 5-3.

Many dairy lactococcal strains harbour plasmids which can encode a variety of important technological traits such as lactose and casein metabolism, bacteriophage resistance, bacteriocin production and citrate utilisation (Mills *et al.*, 2006). In a recent analysis of lactococcal isolates from grass, two of 12 strains harboured a single plasmid, both of which appeared to be ~85 kb (Alemayehu *et al.*, 2014). Analysis of the plasmid complement of the DPC non-dairy isolates demonstrated the presence of varying numbers of plasmids in each strain, with DPC6853 (corn) possessing two large plasmids of ~42kb and ~75 kb, and DPC6856 and DPC6860 containing approximately seven and four plasmids, respectively. Alignment of sequenced contigs with strains IL1403 and MG1363 using Mauve identified a higher number of contigs which did not align with these reference strains in DPC6856 and DPC6860 compared to DPC6853. This corresponded with a larger plasmid complement and a larger genome size of approximately 2.9 mb for DPC6856 (Table 1).

5.4.2. Technologically important traits

The presence of genes involved in casein and lactose utilisation have been reported previously in non-dairy lactococcal strains (Liu *et al.*, 2010; Passerini *et al.*, 2010). Similarily, previous work performed for this thesis has demonstrated a number of important technological characteristics of the non-dairy DPC strains, including the capacity to grow in milk (Chapter 2), the capacity to diversify cheese flavour (Chapter 3) and the observation of enhanced insensitivity to dairy lactococcal phages (Chapter 4). Access to the genome sequences of selected non-dairy DPC strains allows for the examination of the genetic determinants responsible for these important industrial traits, which may dictate their future use in dairy applications.

5.4.2.1. Milk-specific gene sets

Analysis of non-alignment contigs of the non-dairy DPC strains with reference *L. lactis* strains MG1363 and IL1403 revealed the presence of gene clusters associated with protein and lactose utilisation. The grass isolate DPC6860 was found to harbour the proteinase *prtP* and its maturation protein *prtM* with neither of these genes identified in draft sequences of DPC6853 or DPC6856. Analysis of the proteolytic system of 39 *L. lactis* strains by comparative genome hybridisation showed that six strains, from a total of 20 strains from plant environments, possessed *prtP* and *prtM* (Liu *et al.*, 2010). This was attributed mainly to the presence or absence of plasmids. It is most likely that *prtP* and *prtM* are also plasmid-encoded in DPC6860, and that casein utilization is linked to traits that may be of benefit to the cell in the grass environment via this plasmid. Previously, analysis of a bank of dairy *L. lactis* strains revealed *prtP* to be present in only 60.87% of the analysed strains (Passerini *et al.*, 2010). The authors of the study therefore concluded that the possession of *prtP* was not solely responsible for the emergence of industrial dairy strains but that the presence of this gene in a bacterial subpopulation may be essential (Passerini *et al.*, 2010). In *L. lactis*, one transport system has been identified for the intracellular uptake of di-/ tripeptides with two systems for oligopeptide transport (Lamarque *et al.*, 2011). Genes corresponding to the oligopeptide transporters (Opp, Opt and a smaller peptide transporter, DtpT) involved in peptide transport were identified in all of the non-dairy *L. lactis* strains sequenced in this study. The possession of these genes may enable non-dairy strains to grow in milk by scavenging for peptides released by prt^+ organisms and may play an important role in the initial adaptation of non-dairy strains to milk.

All three DPC strains possessed the *lacABCDFEGX* gene cluster necessary for lactose metabolism, as found in dairy-associated lactococci (Ainsworth *et al*., 2014). The *lacABCDFEGX* operon encodes a phosphoenolpyruvate-phosphotransferase system (PEP-PTS) (*lacEF*), tagatose-6-phosphate enzymes (*lacABCD*) and a phospho-β-galactosidase (*lacG*). The identification of this gene cluster corresponds with the observations made in Chapter 2 that each of the isolates was capable of growth in milk, albeit at varying efficiencies. In the case of the grass isolate DPC6860, the lactose operon was located upstream from the genes encoding *prtP* and *prtM*. Previously, two *L. lactis* strains from non-dairy environments have been reported to possess *lacE*, forming part of the lactose PEP-PTS, which was absent in 34.78% of the dairy associated strains analysed (Passerini *et al.*, 2010). In the non-dairy DPC strains examined in this study, *lacE* was present in all genomes and an operon encoding all other components of the *lacABCDFEGX* operon was identified. This implies that these non-dairy strains may be more readily equipped to grow in milk than strains which lack this complete gene cluster, and suggests that upon chance introduction to milk, they may dominate the milk microbiota.

5.4.2.2. Bacteriophage resistance

Previously dairy lactococcal phages were found, on the whole, to be unable to lyse non-dairy lactococcal strains isolated in this study; which in turn led to the isolation of ϕL47 using *L. lactis* DPC6860 as a host (see Chapter 4). Although ϕL47 was capable of lysing three other grass *L. lactis* strains, it was unable to lyse either the corn isolate DPC6853 or the rumen isolate DPC6856. The subsequent genome analysis of non-dairy strains identified putative phage resistance genes some of which were located on contigs which did not correspond to the reference *L. lactis* genomes,

and may correspond to plasmid-encoded resistance mechanisms. Using the CRISPR recognition tool (Bland *et al.*, 2007), no CRISPR/ Cas system was identified in either of the three strains analysed in this study. CRISPR/ Cas systems are rare in *L. lactis* and to date only a single plasmid-encoded system has been identified (Millen *et al.*, 2012).

5.4.2.2.1. Diverse anti-phage systems of non-dairy isolates

Bacteria utilise many different mechanisms to overcome phage attack, amongst which are restriction-modification (R-M) and abortive infection (Abi) systems. Type I R-M systems are composed of three subunits: HsdR (restriction), HsdM (methylation) and HsdS (specificity), which act to cleave invading foreign genetic material, while protecting host DNA (Stern and Sorek, 2011). Three type I R-M systems were identified in *L. lactis* DPC6853, of which the HsdS subunits showed no significant matches to other HsdS subunits in *L. lactis*, but showed 42-64% amino acid identity to specificity proteins of other species including *Solobacterium*, *Gardnerella* and *Candidatus* species (Fig. 1). In addition, one of these systems appears to form a chimeric type I R-M system containing three HsdS subunits (Fig. 1). The stacking of multiple *hsdS* genes may be beneficial to a cell by creating an increased level of recognition by the R-M system, leading to an overall higher level of phage resistance (Seegers *et al.*, 2000). Common S-subunits contain two variable and three conserved regions, and typically consist of more than 400 amino acids (Adamczyk-Popławska *et al.*, 2003). In the chimeric type I system of DPC6853, the HsdS protein immediately upstream of the HsdM protein is composed of 468 amino acids and contains two target recognition domains. Contrastingly, the other two specificity subunits of DPC6853 consist of 288 and 200 amino acids respectively and

consist of only one target recognition domain. It has been demonstrated previously that HsdS genes can recombine through the core domain, which in turn can create HsdS subunits with different specificities (Gubler *et al.*, 1992). Although two subunits identified in this system consisted of a single conserved domain, they may contribute to alternative specificity through homologous recombination with whole specificity proteins, as suggested in *L. lactis* IL594 (Górecki *et al.*, 2011)

Type II R-M systems possess both restriction endonuclease and methyltransferase subunits but differ from type I systems by the absence of a specificity protein (or HsdS subunit). *L. lactis* DPC6860 encodes a type II R-M system with high % amino acid similarity to systems found in *Enterococcus faecalis*, *Streptococcus equi* and *Lactobacillus gasseri*. The downstream gene encodes a protein of 676 amino acids which has a 60% amino acid identity to a type II restriction endonuclease of *Streptococcus equi* ssp. *ruminatorum* CECT 5772. The upstream gene shares 64% amino acid identity to an adenine-specific methyltransferase gene of *Enterococcus faecalis* MTUP9, and possesses a D12 class methyltransferase (pfam02086) and an adenine-specific DNA methylase (COG3392) domain. Of note, mining of non-conserved ORFs compared to lactococcal groups (dairy; dairy and non-dairy) identified numerous prophage-related proteins with significant similarity to phage-related proteins of *Lactococcus garvieae*. Future analysis of *L. garvieae* phages may shed further light on the ability of phages to overcome host defences and aid in the development of cultures with enhanced phage insensitivity.

5.4.2.3. Flavour formation

The release of amino acids from large milk proteins and their subsequent conversion to volatile aroma compounds is crucial for the development of flavour in

semi-hard cheese varieties such as Gouda (Smit *et al.*, 2005). Following the internalisation of peptides released by the extra-cellular proteinase, peptidases function in the release of amino acids. Comparative genome analysis revealed that the non-dairy DPC strains possess a peptidase complement similar to that of dairy lactococci (data not shown). Following the liberation of aromatic and branched chain amino acids by peptidases, transamination is the first step in the conversion to volatile compounds, which leads to the formation of an α-keto acid (Smit *et al.*, 2005). As described in Chapter 2, non-dairy DPC isolates possessed increased amino acid transferase activity for phenylalanine. In addition to a putative aromatic aminotransferase, *araT*, five putative aspartate aminotransferases were identified in DPC6856 and DPC6853, and four in DPC6860. In strains MG1363 and IL1403 three and two aspartate aminotransferases have been identified respectively. Previously, the aspartate aminotransferase of *Brevibacterium linens* was shown to function in the transamination of aspartate, but was also active on aromatic amino acids (Yvon and Rijnen, 2001). Further analysis revealed that in strains DPC6853 and DPC6856, two of the putative aspartate aminotransferases possess PRK07309 domains found in aromatic aminotransferases, while only one ORF, in strain DPC6860, harboured a similar domain. Therefore, the enhanced transferase activity of non-dairy isolates could be in part due to an increased number of aminotransferases which are active on phenylalanine. In a cheese matrix, the conversion products of phenylalanine form rosy aroma notes which may be considered off-flavours (van Kranenburg *et al.*, 2002; Marilley and Casey, 2004). This may account for the increased association of non-dairy strains with off-flavours as determined by sensory analysis in cheese models (Chapter 3); however, this is speculation without volatile analysis of cheeses.

Transamination of the branched chain amino acid (BCAA) leucine forms α-isocaproic acid which can be decarboxylated to the potent flavour compound 3-methylbutanal. A putative α-keto acid decarboxylase, which functions in the formation of this compound, was identified in the corn isolate DPC6853 and the rumen isolate DPC6856, with 99% amino acid identity to that of *L. lactis* B1157. In cheddar cheese, 3-methylbutanal forms a dark chocolate, malt aroma and is a key flavour component of Proosdij cheese (Singh *et al.*, 2003; Smit *et al.*, 2005). Interestingly, in milk fermentations (Chapter 3), neither DPC6853 nor DPC6856 were strongly associated with 3-methylbutanal. However, this analysis was performed in a milk model which, although it offers insight into the capacity of a strain to produce volatile compounds, differs considerably from a cheese environment. In a cheese matrix, if functional, this enzyme could increase the production of 3-methylbutanal and contribute to flavour formation.

5.4.3. Niche specific traits

In comparison to milk, the composition of non-dairy environments differs greatly, particularly in terms of carbohydrate composition and protein concentration. Although mainly associated with milk, *L. lactis* has been isolated from a variety of different niches, from drain water to sourdough bread to mung bean sprouts (Siezen *et al.*, 2008; Kato *et al.*, 2012; Passerini *et al.*, 2013). It is postulated that *L. lactis* came to inhabit milk from grass via cattle. A number of studies have examined the genomes of non-dairy isolates (Siezen *et al.*, 2008; Passerini *et al.*, 2013; Kato *et al.*, 2012) however to date, none have examined *L. lactis* isolates from the bovine rumen or grass. Indeed, comparison of DPC6856 (rumen) and DPC6860 (grass) may provide

information on whether *L. lactis* is a normal rumen inhabitant or simply a transient in this environment.

5.4.3.1. Diverse carbohydrate fermentation capacity of non-dairy isolates

Plant material presents a variety of different carbohydrates that are absent from the milk environment. Consequently, non-dairy lactococci have been found to ferment different plant-associated carbohydrates, with industrial dairy strains showing significantly reduced capacity in this regard (Siezen *et al.*, 2008; Alemayehu *et al.*, 2014). Mining of non-conserved ORFs within DPC lactococcal strains compared to dairy strains resulted in the identification of different genes associated with carbohydrate/ sugar utilisation which were further analysed using Artemis (Rutherford *et al.*, 2000). This, in turn, allowed for the elucidation of specific gene clusters in each strain.

5.4.3.1.1. Specific carbohydrate utilisation genes in grass and corn isolates

Xylan is predominantly comprised of 1,4-β-D-xylose monomers which can be substituted with arabinosyl or acetyl groups amongst others . Acetyl xylan esterases function in hydrolysing the ester bonds of acetyl groups in natural xylan which can be found in both cereals and annual plants (Whistler and Richards, 1970). For efficient action of endoxylanases on xylan to take place, the presence of an acetyl xylan esterase is crucial (van den Brink and de Vries, 2011). A putative acetyl xylan esterase was identified in strains DPC6853 (corn) and DPC6860 (grass) (Fig. 2A). Upstream from this ORF, a putative glycoside hydrolase family 113 protein was identified, from a family of hydrolases that exhibit β-mannase activity. These hydrolases cleave β-1,4,mannosidic linkages in glucomannan and galactomannan (Lombard *et al.*, 2014),

found in some plant cell walls. Other ORFS within this gene cluster encode for sugar uptake and β-glucosidase activity (Fig. 2A).

A putative arabinogalactan endo-1,4-β-galactosidase was identified in strain DPC6860 which exhibited 57% amino acid identity to that of *Bacillus coagulans*, with no homologues in *L. lactis*. This ORF is flanked by two transposases which may contribute to the mobilisation of this gene. Sequence analysis revealed two glycosyl hydrolase family 53 protein domains (pfam07745) and two COG3867 domains found in arabinogalactan endo-1,4-beta-galactosidases. Similar to plants, grasses possess highly branched arabinogalactans involved in primary cell wall formation (Carpita, 1996) and it is presumed that this protein may play a role in the utilisation of plant carbohydrates by DPC6860 in the grass environment.

Only strain DPC6853 could grow in the presence of sucrose (Table 2) and a gene cluster encoding sucrose metabolism was identified, which included a putative fructokinase, PTS protein and hydrolase (Fig. 2D). Both DPC6860 and DPC6853 could grow using D-ribose which corresponded to the identification of a putative ribokinase, ribose associated transporter proteins and transcriptional regulator on the genome.

5.4.3.1.2. Insights into carbohydrate utilisation by rumen-associated *L. lactis*

Little is known about *L. lactis* and its inhabitation of the bovine rumen. Due to the large consumption of grass by cattle, these organisms may be transient and simply pass through the bovine digestive system. On the other hand, *L. lactis* may colonise the rumen whereby it survives by fermenting plant material ingested by the animal. Presumably *L. lactis* came to inhabit the rumen through bovine consumption of grass

containing these organisms, which are equipped to utilise grass-derived carbohydrates.

5.4.3.1.2.1. Common carbohydrate-degrading enzymes associated with rumen and grass niches

Comparative analysis of both the grass isolate DPC6860 and the rumen isolate DPC6856 identified gene clusters involved in carbohydrate utilisation, which are common to both strains, supporting the hypothesis that DPC6856 originated from a grass environment. Xyloglucans are amongst the most common hemicelluloses found in cell walls of most land plants and are comprised of a D-glucose backbone substituted with D-xylose. In both DPC6860 and DPC6856, a four gene cluster, encoding a β-xylosidase, endoglucanase, permease and transcriptional regulator, was identified (Fig. 2B), absent in assembled contigs of the corn isolate DPC6853. The putative β-xylosidase protein possesses a glycosyl hydrolase family 43 protein domain, found in numerous enzymes which exhibit both α-L-arabinofuranosidase and β-D-xylosidase activity (Lombard, Ramulu *et al.*, 2014). The action of β-D-xylosidases releases D-xylose monomers from xylan backbones. The upstream ORF encodes a protein with 99% amino acid identity to an endoglucanase of *L. lactis* GE124. A domain search of this ORF revealed the presence of a cellulase (glycosyl hydrolase family 5; pfam00150) and endoglucanase domain (COG2730) which may contribute to the breakdown of plant cell wall components such as cellulose. For the liberation of glucose from the xyloglucan backbone to take place, α-1,6-linked D-xylose units must be removed by an α-xylosidase. Strains DPC6860 and DPC6856 both possess a gene cluster encoding a transcriptional regulator, α-xylosidase and transport associated proteins (Fig. 2B).

In addition to the xylosidase-associated genes of both strains, a three gene cluster was identified which encodes a β-glucosidase, a 6-phospho-β-glucoisodase and a sugar kinase (Fig. 2B). These ORFs possess only two significant matches in *L. lactis*, the dairy organisms MG1363 and NZ9000, and the putrescine producing strain GE124. The putative β-glucosidase possesses a glycosyl hydrolase family 3 domain, which groups together β-D-xylopyranosidases, β-D-glucosidases and α-L-arabinofuranosidases, and is involved in cellulose degradation (van den Brink and de Vries, 2011; Lombard *et al.*, 2014). Within the xylose utilisation gene cluster, an additional ORF was identified between the xylose isomerase and the xylulose kinase in strains DPC6860 and DPC6856 (Fig. 2C). Sequence analysis of this ORF revealed the presence of a glycosyl hydrolase family 11 domain that exhibited 100% amino acid identity to a endo-1,4-β-D-xylanase found in *L. lactis* A76. During xylan breakdown, an endo-1,4-β-D-xylanase functions in splitting xylan into xylose polymers of different sizes which are subsequently taken into the cell .

5.4.3.1.2.2. Specialisation to the rumen environment

Despite the similarities that were observed between DPC6860 and DPC6856, notable differences were also evident, suggesting that DPC6856 may not be transient, but a common rumen inhabitant. A gene cluster involved in carbohydrate metabolism was identified in DPC6856 which was not identified in the grass strain, DPC6860 (Fig. 2E). Within this gene cluster, a putative esterase/ lipase was identified, which possessed an Aes domain (COG0657) found in esterases/ lipases. Adjacent to this ORF, a putative lyxose ketol isomerase was identified which has only two significant matches in *L. lactis*, strains IO-1 and KW2, isolated from drain water and fermented corn respectively. This enzyme catalyses the reversible conversion of D-lyxose to D-

xylulose which can be further converted to D-xylulose-5-phosphate, and subsequently enter the pentose phosphate cycle. D-lyxose isomerase from *Cohnella laevoribosii* RI-39 sp. nov, also participates in the reversible conversion of L-ribose to L-ribulose and D-mannose to D-fructose (Cho *et al.*, 2007; Kwon *et al.*, 2010). Three genes within this cluster were identified as PTS components with fructose specificity, while a putative fructose bisphosphate aldolase was also identified. Mannans such as galactomannans and glucomannans are composed of D-mannose and can be found in plant cell walls. Following the release of D-mannose monomers, these sugars can be converted to D-fructose by the lyxose ketol isomerase and taken into the cell via a PTS. Together this gene-set may function in the utilisation of D-mannose or other monosaccharides released into the rumen by the action of other, surrounding organisms.

DPC6856 exhibited weak growth in the presence of both L-arabinose and melibiose, in which DPC6860 and DCP6853 showed no growth. After D-xylose, L-arabinose is the most abundant monosaccharide in plants and is involved in pectin and hemicellulose formation (Seiboth and Metz, 2011). Similar to that identified in strain KF147 (mung bean sprouts), a gene cluster encoding for arabinose fermentation was identified in DPC6856, comprising seven ORFs (Siezen *et al.*, 2008). Melibiose is an α-galactoside which are common plant oligosaccharides. An α-galactosidase is required for fermentation of melibiose by cleaving an α-1-6 glycosidic bond (Boucher *et al.*, 2002); however, no putative α-galactosidase was found in the DPC6856 genome. Due to the draft status of the genome, the observation that DPC6856 was able to grow weakly in the presence of melibiose, prompts us to hypothesise that this strain may encode an α-galactosidase which may have been present in a sequence gap. In terms of sugar fermentation, the most striking difference between DPC6860 and

DPC6856 is the inability of the rumen isolate to grow in the presence of xylose as the sole sugar source. Disruption of *xylT* in *Clostridium acetobutylicum* significantly reduced the cells ability to grow in the presence of xylose as a sole sugar source (Gu *et al.*, 2010). Both D-xylose proton symporters appear to be pseudogenes in strains DPC6853 and DPC6856 which may account for their inability to utilise D-xylose. Similarily, ORFs encoding a putative arabinogalactan endo-1,4-beta-galactosidase and acetyl xylan esterase were absent in DPC6856.

Overall, DPC6860 (grass) and DPC6856 (rumen) share common genes involved in carbohydrate utilisation. However, the absence of certain genes involved in complex carbohydrate degradation (i.e. arabinogalactan endo-1,4-beta-galactosidase) and simple sugar utilisation (i.e. xylose utilisation) hints that these may not be required in the rumen owing to the contribution of the rumen microflora to breakdown of grass-associated sugars. Similarily, DPC6856 possessed additional gene sets encoding for arabinose utilisation and simple sugar utilisation absent in DPC6860. Other genetic features were also identified which support this hypothesis; these are discussed below.

5.4.3.2. Surviving the cold: genetic determinants associated with growth at low temperatures

At low temperatures, bacteria need to overcome numerous hurdles including increased instability of RNA and DNA structures (Cartier *et al.*, 2010). During a temperature change from hot to cold, cold shock proteins (CSPs) are important in reducing the negative impact that this environmental alteration can have on the cell (van de Guchte *et al.*, 2002). Analysis of non-dairy strains identified four, six and nine CSPs in strains DPC6853, DPC6860 and DPC6856 respectively (Table 3). In the dairy *L. lactis* strains IL1403 and MG1363, two and seven CSPs have been

identified respectively (Bolotin *et al.*, 2001; Wouters *et al.*, 2001). To examine whether these genes contributed to the survival of the non-dairy DPC strains at low temperatures, growth was monitored over 20 days at 6 ^{o}C with measurements taken every 48 hrs (Fig. 3). The grass isolate DPC6860 grew relatively well at 6 ^{o}C with strains DPC6853 and DPC6856 reaching a higher OD than the dairy strains, MG1363 and IL1403. The increased number of CSPs in DPC6856 did not appear to enhance its ability to grow at 6 ^{o}C and DPC6860 demonstrated the highest capacity to grow at this temperature, possessing six CSPs. The higher number of CSPs in MG1363 may account for the differential growth rate at 6 ^{o}C between the two dairy strains (IL1403 and MG1363) observed in this study. In *Bacillus subtilis,* all three CSPs are induced under cold conditions in contrast to *E. coli* where only four of the nine CSPs are induced (Sanders *et al.*, 1999). Therefore we may only speculate as to how many of these CSPs identified in DPC6856 are induced during cold conditions; however, transcriptome analysis of non-dairy lactococci grown under cold conditions could shed further light on if and when these genes are expressed.

Mining of non-conserved ORFs in the non-dairy DPC strains compared to *L. lactis* from dairy and non-dairy origins identified a DEAD/ DEAH-box helicase in strains DPC6853 and DPC6856 with a highest significant match to *Listeria monocytogenes*. In *L. monocytogenes* EGD-e, the growth of DEAD-box negative mutants was impaired at 3 ^{o}C suggesting these genes play a role in survival at lower temperatures (Markkula *et al.*, 2012). Interestingly, DPC6860 lacked the DEAD-box helicase identified in other DPC strains but showed the fastest growth rate at 6 ^{o}C. This suggests that this gene does not contribute to growth at 6 ^{o}C but may play a role in survival at lower temperatures than 6 ^{o}C or when exposed to other environmental stresses.

5.4.3.3. Surviving low pH: Biogenic amine production

5.4.3.3.1. Cadaverine production

Biogenic amines are nitrogen-containing organic bases of low molecular weight, which are formed by the decarboxylation of amino acids (Alvarez and Moreno-Arribas, 2014). The consumption of foods containing high levels of biogenic amines may lead to a number of health issues such as respiratory disorders and tachycardia (for review see Ladero *et al*., 2010). Strain DPC6853 was found to possess a putative lysine decarboxylase gene which exhibits a 66% amino acid identity with the lysine decarboxylase gene of *Enterococcus faecalis* and *Streptococcus agalactiae*. This ORF contained a PGGXGTXXE motif (Fig. 4), which is highly conserved amongst a number of bacterial proteins, some of which are annotated as lysine decarboxylases, and possessed a Rossmann fold nucleotide binding domain (COG1611), involved in cofactor binding. The production of cadaverine by plant-associated organisms has been correlated with enhanced root development in rice and soybean seedlings (Gamarnik and Frydman, 1991; Cassan *et al.*, 2009). The isolation of DPC6853 from corn hints that the putative production of cadaverine by this organism may contribute to the development of the corn plant. Growth of DPC6853 in lysine decarboxylase broth was assessed after 48 hrs and indicated a negative result for lysine decarboxylation. However, the presence of the gene indicates the potential for production of cadaverine, and in food products, biogenic amine production is affected by numerous environmental conditions including temperature, pH, salt concentration and oxygen availability (EFSA, 2011). Therefore, under certain conditions, DPC6853 may still produce cadaverine and this may negatively impact the suitability of this strain for use in food production.

5.4.3.3.2. Putrescine production

Putrescine is a biogenic amine that is formed via two separate pathways. The first involves the conversion of arginine to ornithine which is subsequently decarboxylated to putrescine. The second involves the formation of agamatine via decarboxylation, and the formation of putrescine by deimination. The second pathway has been identified in *L. lactis* from dairy sources and also in *L. lactis* KF147 (Ladero *et al.*, 2011). All three strains examined in this study possessed the same pathway for putrescine production which, in turn, may have implications for the use of these strains in food production. The consumption of putrescine is not held to have any direct toxic effects; however, it has been implicated in contributing to malignancy, both directly and indirectly (Ladero *et al.*, 2010), and is capable of augmenting the harmfulness of other biogenic amines. Analysis of 'wild-type' *L. lactis* strains IO-1 (drain water), KLDS 4.032 (fermented horse milk) and KW2 (fermented corn), also showed the presence of complete pathways for putrescine production by deimination of agamatine, and suggests that this pathway may play a role in survival at low pH in diverse environments. Using the RAST SEED viewer (Altschul *et al.*, 1997), no histidine or tyrosine decarboxylase genes were identified in either of the three lactococcal isolates. The absence of these genes supports the use of primers in Coton and Coton (2005) and Coton *et al.* (2004) for the identification of these genes in non-dairy lactococci.

5.4.3.4. Resistance and immunity to antimicrobial compounds

The spread of resistance amongst microbial populations is a serious concern in the fight against antibiotic resistant bacteria. If antibiotic resistance genes are linked

to mobile genetic elements, the strain may be deemed unsuitable for use in food production. Phenotypic analysis performed in Chapter 2 showed that *L. lactis* ssp. *lactis* DPC6853 possessed an increased resistance to tetracycline in comparison to other non-dairy isolates. This resistance was above the cut-off values for antimicrobials set out by EFSA (2012). A tetracycline resistance protein identical to that of *L. monocytogenes* LM78, *Enterococcus faecalis* EnGen0311 and *Lactococcus garvieae* BCC43578 was found on the genome of DPC6853. This protein is commonly associated with plasmids or transposable elements, and serves in ribosomal protection. The contig containing this gene did not align with the reference genome of IL1403 and may form part of one of the DPC6853 plasmids. Sequence analysis of this contig also identified a hypothetical protein with 100% amino acid identity to *L. garvieae* and a mobile element protein identical to *Enterococcus faecalis*. The association of this protein with a transposable element suggest that this gene may be transferred to other bacteria if used in food production and contribute to the horizontal spread of antibiotic resistance.

Mining of non-conserved ORFs using FGD also identified a CAAX amino terminal protease family protein in all of the tested DPC strains absent in dairy lactococci. A database search of this protease using BLASTp showed that these proteins were similar to proteins identified in two non-dairy associated *L. lactis* strains (Dephy1 and AI06) with 100% query coverage. In contrast, these proteins were smaller by 16 residues in the industrial dairy *L. lactis* strains NZ9000 and MG1363. A domain search of this protein identified an Abi domain (pfam02517) found in proteases associated with genes involved in bacteriocin production, which are implicated to play a role in the immunity of the bacteriocin-producing strain to its own bacteriocin (Kjos *et al.*, 2010). In all DPC strains, these proteases did not appear to

be associated with bacteriocin production genes. Therefore, in non-dairy isolates, these genes may confer these organisms with the ability to resist attack by antimicrobial compounds produced by other organisms.

5.4.4. Heavy metal ion transport

5.4.4.1. Increased insensitivity to cobalt

Cobalt is an essential component of many enzymes, but in the plant environment soluble Co^{2+} is normally present at minute concentrations (Rodionov *et al.*, 2006). Therefore, bacteria must utilise a high-affinity uptake system to scavenge for this cation to meet their requirements. Two different Co^{2+} transport systems were identified in strains DPC6860 and DPC6853 (Fig. 5). For the microbial transport of Co^{2+}, CbiMNQO has been identified as the most prevalent transport system in LAB (Rodionov *et al.*, 2006). This system is composed of an ATPase (CbiO), two trans-membrane elements (CbiM/ Q) and a small membrane bound element (CbiN). In strain DPC6860, three ORFs contained features that are consistent with the different elements which comprise a Co^{2+} transporter (Fig. 5). ORF 6 possessed a CbiQ domain (COG0619) while both ORFs 7 and 8 possessed CbiO domains (COG1122) suggesting they function in ATP binding. ORF 4 shared a 96% amino acid identity to a hypothetical protein of *Streptococcus macedonicus* while ORF 5 encodes a putative membrane protein with a 99% amino acid similarity to that of *Lactobacillus buchneri* CD034. Topological analysis of ORF 5 using TOPCONS, showed the presence of six transmembrane motifs which suggests that gene product may form a transmembrane component of this system similar to CbiM. TOPCONS search of ORF 4 revealed a single transmembrane motif suggesting that although it is associated with the cell membrane, it has no features in common with CbiN; however, comparative alignment

of this contig showed that ORF 4 had no similarity to any protein identified in the corresponding CbiMNQO system of other bacteria (Fig. 5).

A similar gene cluster was also identified in DPC6853 located on contig C24 which contains three genes flanked by transposases. In comparison to DPC6860, this system appeared to be more streamlined and is composed of three ORFs (Fig. 5). BlastP analysis showed these ORFs possessed only two significant matches in *L. lactis* strains A12 and UC509.9. ORF 4 encodes a small protein consisting of 114 residues, which shares a 97% amino acid identity to an integral membrane protein from *L. lactis* A12. ORF 5 exhibits a 99% identity to a transport protein from *L. lactis* A12 and contains a COG0390 domain found in the permease component of ABC-type transporters. Comparable with CbiM, this protein also contains seven transmembrane helices. ORF 6 exhibits a 99% identity to an ATPase like protein from *L. lactis* A12 and contains numerous domains associated with the ATP-binding components of different ABC transporters, including CbiO. To examine the functionality of these systems, non-dairy DPC strains were examined for their ability to grow in LM17 media containing up to 6 mM $CoCl_2$ for 24 hrs (Fig. 6). *L. lactis* DPC6853 was capable of growth in media containing 6 mM Co^{2+}, reaching an OD_{600nm} of 0.67 in comparison to 0.37 for DPC6860. The presence of a Co^{2+} transport system previously identified in the plant-associated strain *L. lactis* DPC3901, was hypothesised to allow the cell to maintain homeostasis at higher concentrations of this heavy metal (Fallico *et al.*, 2011).

5.4.4.2. Cobalt sensitivity of DPC6856

In contrast to the relatively high level of resistance of strains DPC6853 and DPC6860, strain DPC6856 was unable to grow at Co^{2+} concentrations above 0.2 mM.

Examination of the sensitivity of rumen-associated bacteria to heavy metals showed that after mercury and cadmium, cobalt was the most toxic to these organisms (Salem *et al.*, 2011). The two transport systems identified in DPC6853 and DPC6860 may equip these organisms to scavenge for Co^{2+} in plant environments; however, no such genetic machinery may be needed in DPC6856 due to the environment that it inhabits. Ruminant diets require a sufficient supply of Co^{2+} which is essential for the biosynthesis of vitamin B_{12} (Tiffany *et al.*, 2003). Cattle acquire the majority of their minerals through forage consumption but also ingest small amounts through water and soil (McDowell, 1996). Four putative *corA* transporters were annotated in strain DPC6856, which has not previously been reported in *L. lactis*. Therefore the constant influx of plant matter and water may provide a higher concentration of Co^{2+} than in the plant environment and negate the requirement of a high affinity transport system.

5.4.4.3. Cadmium

To examine whether non-dairy isolates could also grow in the presence of cadmium (Cd^{2+}), strains were grown for 24 hrs in media containing 0.1-0.6 mM $CdCl_2$. All strains were sensitive to Cd^{2+} at concentrations of >0.4 mM with only strain DPC6853 capable of growth at 0.3 mM. A cadmium efflux ATPase (100% identity to *L. lactis* ssp. *lactis*) and a cadmium efflux system accessory protein (67% identity to *Enterococcus faecium*) were identified in strain DPC6853, in addition to a heavy metal transporting ATPase, which may contribute to this organism surviving relatively low concentrations of cadmium. Although a putative *cadA* and *cadC* were identified in DPC6860, these genes did not appear to affect growth in the presence of cadmium. Whereas Co^{2+} is involved in certain cellular processes, Cd^{2+} is not identified as possessing any metabolic role (Rossbach *et al.*, 2000), and is toxic to

cells at low concentrations. Cd^{2+} is used in various pesticides and fertilisers (Bruins *et al.*, 2000) and the use of this cation in corn production may account for the resistance of DPC6853 to this heavy metal.

5.4.5. Extra-cellular secretions

5.4.5.1. WXG100 secretion system (Wss)

Protein secretion is considered essential for pathogenic organisms to (i) take full advantage of nutrient resources and (ii) avoid host defence systems (Houben *et al.*, 2012). The presence of a WXG100 family protein coupled with a protein possessing a FtsK-SPOIIIE domain are the only conserved parts of a type VII secretion system amongst a wide range of *Firmicutes* and *Actinobacteria* (Abdallah *et al.*, 2007; Bitter *et al.*, 2009). Strains DPC6856 and DPC6860 encode a putative WXG100 protein of 96 residues, which possessed both WXG100_ESAT6 (TIGR03930) and WXG100 (pfam06013) domains (Fig. 7). Of note, this protein was not identified in any subspecies *lactis* strain, and could prove a potential tool for rapid identification of *cremoris* subspecies genotypes in the future. Downstream from each of these ORFs, a putative ATPase was identified which is required for secretion of the WXG100 family protein, that were annotated as EssC proteins. The observation that this secretion system is only found in four *L. lactis* strains, highlights the potential role of this system in survival outside of the milk environment. The identification of WXG100 proteins in industrial dairy strains, suggests that this system is dispensable in milk, and a complete system may have been present in these strains prior to adaptation to this environment.

5.4.5.2. Exopolysaccharide (EPS) production

For colonisation of a plant environment, EPS biosynthesis is considered crucial in organisms such as *Gluconacetobacter diazotrophicus* PAL5 (Meneses *et al.*, 2011). The EPS phenotype enables bacteria to irreversibly bind to root surfaces and contribute to biofilm formation, while also contributing to protection from plant defences (Rodríguez-Navarro *et al.*, 2007; Meneses *et al.*, 2011). A gene cluster encoding EPS biosynthesis was identified in strain DPC6860 (grass) (Fig. 8) spanning over 15 kb, with an identical gene order to that identified in *L. lactis* SMQ-461 (Dabour and LaPointe, 2005). Genes involved in EPS production were identified on separate contigs in strain DPC6853 (corn) (contigs C107, C21, C1). The gene cluster identified in strain DPC6860 showed highest amino acid similarity to *Lactococcus* species *lactis*, *raffinolactis* and *garvieae* and to *Streptococcus thermophilus*. A putative *epsG* possessed no significant matches in *L. lactis* with the closest match to a hypothetical protein of *Butyrivibrio* sp. LC3010 (64% aa identity; e-value 6e-103). This protein possessed a Succinoglycan-BP-ExoA domain which is involved in the biosynthesis of succinoglycan; an acidic EPS which is important in the penetration of root nodules. In the cheese industry, certain strains are capable of producing an EPS which may contribute to the texture of the product, particularly in low fat systems (Costa *et al.*, 2010). Similar to strain DPC6860, *L. lactis* KF147 (isolated from Mung bean sprouts), possessed a gene cluster encoding EPS biosynthesis and this phenotype may be a remnant of when *L. lactis* occupied a plant environment (Siezen *et al.*, 2008).

5.4.5.3. *yydFGHIJ*-like gene cluster of DPC6856

Inhibitory compounds produced by a microbe are implicated in having a significant impact on the survival of that strain in the rumen (Kalmokoff *et al.*, 1996). Therefore the possession of the genetic equipment necessary for antimicrobial

production would be of great benefit for a cell to thrive in this environment. In *L. lactis* DPC6856, contig R163 carries a *yydFGHIJ*-like gene cluster (Fig. 9) similar to that found in *Bacillus subtilis* (Butcher *et al.*, 2007). This operon appears to possess all of the necessary genes for the production and secretion of a small antimicrobial like peptide. Based on analyses performed using the Antimicrobial Peptide Database (http://aps.unmc.edu/AP/main.html) (Wang, *et al.*, 2009, Wang and Wang, 2004), ORF 6 encodes a small protein (68 amino acids) which possesses antimicrobial associated properties; namely the capacity to form an α-helix structure with 12 residues on the same surface, a +2.4 charge at a neutral pH and is 36% hydrophobic. Alignment of this protein using the Antimicrobial peptide database (Wang *et al*., 2009, Wang and Wang 2004) showed that this peptide exhibited the greatest similarity to a cyto-insectotoxin of *Lachesana tarabaevi* (Fig. 11) which possesses activity against *Arthrobacter globiformis*, *Bacillus subtilis* and *Pseudomonas aeruginosa* that can be commonly found in soil.

BLASTp analysis of ORF 8 revealed the presence of an ATP-binding cassette (cd03228) domain, conserved in multidrug resistance protein-like transporters. ORF 9 encodes a putative papain cysteine protease family protein with 27% amino acid identity to *Carnobacterium maltaromaticum*. This protein contained a Pept-C1 domain (smart00645) indicative of a papain family cysteine protease which may function in peptide cleavage. ORF 10 encodes a protein (344 amino acids) exhibiting a 62% amino acid identity to a hypothetical protein of *Streptococcus* sp. HSISS3. Consistent with the YydG protein of *Bacillus subtilis* (Butcher *et al.*, 2007) this protein possesses a rSAM-yydG (TIGR04078) and a Fe-S oxidoreductase (COG0535) domain, and may function in peptide modification. TOPCONS analysis of ORFs 11 and 12 identified one and five trans-membrane motifs respectively and suggests that

these proteins may function as permease subunits of the system. In *Bacillus subtilis*, the modified peptide produced by the yyd*FGHIJ* operon evokes a response by the LiaRS two component system which also responds to disruption of the cell envelope caused by antibiotics such as vancomycin and nisin (Cao *et al.*, 2002; Mascher *et al.*, 2003; Mascher *et al.*, 2004). Overall, this gene cluster identified in DPC6856 may play a role in inhibiting the growth of other bacteria and may be of importance to this strain in adapting to the rumen environment.

5.4.6. Unique presence of *cobQ* in *L. lactis*

Vitamin B_{12} or cobalamin is of crucial importance in cellular growth, acting as a coenzyme in numerous metabolic reactions. For the formation of this molecule, approximately 30 enzymatic steps are required which can proceed under aerobic or anaerobic conditions. In ruminants, Vitamin B_{12} is synthesised by the microbes that inhabit the rumen (Wolin *et al.*, 1997). A putative cobyric acid synthase was identified in strain DPC6856, which is absent in all other *L. lactis* genomes. This ORF exhibited a 67% amino acid identity to *cobQ* of *Bacillus cereus* which is involved in the conversion of cobyrinic acid to vitamin B_{12} and comprises part of the porphyrin metabolic pathway. This gene possesses a G-C content of 25.95% much lower than the 35-36% G-C content of *Lactococcus lactis*, and is a clear contender for horizontal gene transfer. Previously, analysis of the ORFeomes of *Clostridium* and *Ruminococcus* species identified *cobQ* in *Clostridium* species only which the author stated, highlights the ongoing adaptation of *Ruminococcus* to the rumen environment (Altermann, 2012).

5.5. Conclusions

In conclusion, draft genome sequencing and FGD offered a rapid means of identifying genes/ gene clusters specific to non-dairy isolates. Similarily, analysis of non-alignment contigs identified genetic regions which may correspond to plasmid DNA or genomic islands. Non-dairy isolates possessed traits associated with dairy *L. lactis* strains that would be beneficial in the milk environment (lactose and casein utilisation). These strains appeared to possess diverse mechanisms for overcoming phage attack, and may be beneficial in dairy fermentations. In addition, these strains also possessed niche-specific genes which may play an important role in adaptation to their respective environment, most notably genes involved in carbohydrate utilisation and heavy metal resistance.

Based on genome analyses, *L. lactis* DPC6856 has acquired traits which may enable it to thrive in the rumen while concurrently forfeiting others such as xylose fermentation. In all, the acquisition and loss of specific traits is reminiscent of the domestication of *L. lactis* to milk, and suggests that this organism is a typical inhabitant of the rumen environment. Overall, non-dairy strains exhibited enhanced genetic diversity and demonstrated the capacity to thrive under harsher conditions than those presented in an industrial dairy setting. Further analysis of specific gene sets identified in this study will shed further light on the role that they may play in niche specialisation and may be beneficial in understanding the capacity of *L. lactis* to respond to its environment.

5.6. Acknowledgements

Thanks to A. Casey for assistance in genome assembly, annotation and analysis.

Thanks to E. Altermann for bioinformatic analyses and identification of non-conserved ORFs.

Thanks to P. Cotter for genome sequencing.

5.7. References

Abdallah, A. M., Van Pittius, N. C. G., Champion, P. a. D., Cox, J., Luirink, J., Vandenbroucke-Grauls, C. M., Appelmelk, B. J., and Bitter, W. (2007). Type VII secretion—mycobacteria show the way. *Nat. Rev. Microbiol.* 5**,** 883-891.

Adamczyk-Popławska, M., Kondrzycka, A., Urbanek, K., and Piekarowicz, A. (2003). Tetra-amino-acid tandem repeats are involved in HsdS complementation in type IC restriction–modification systems. *Microbiol.* 149**,** 3311-3319.

Ainsworth, S., Stockdale, S., Bottacini, F., Mahony, J., and Van Sinderen, D. (2014). The *Lactococcus lactis* plasmidome: much learnt, yet still lots to discover. *FEMS Microbiol. Rev.* 38, 1066-1088. DOI: 10.1111/1574-6976.12074.

Alemayehu, D., Hannon, J. A., McAuliffe, O., and Ross, R. P. (2014). Characterization of plant-derived lactococci on the basis of their volatile compounds profile when grown in milk. *Int. J. Food Microbiol.* 172**,** 57-61. DOI: 10.1016/j.ijfoodmicro.2013.11.024.

Altermann, E. (2012). Tracing lifestyle adaptation in prokaryotic genomes. *Front. Microbiol.* 3 (48), 1-17.

Altermann, E., and Klaenhammer, T. R. (2003). GAMOLA: a new local solution for sequence annotation and analyzing draft and finished prokaryotic genomes. *Omics A J. Integrative Biol.* 7**,** 161-169.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215**,** 403-410.

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic acids Res.* 25**,** 3389-3402.

Alvarez, M. A., and Moreno-Arribas, M. (2014). The problem of biogenic amines in fermented foods and the use of potential biogenic amine-degrading microorganisms as a solution. *Trends Food Sci. Tech.* 39, 146-155

Aziz, R. K., Bartels, D., Best, A. A., Dejongh, M., Disz, T., Edwards, R. A., Formsma, K., Gerdes, S., Glass, E. M., and Kubal, M. (2008). The RAST Server: rapid annotations using subsystems technology. *BMC genomics.* 9**,** 75.

Bitter, W., Houben, E. N., Bottai, D., Brodin, P., Brown, E. J., Cox, J. S., Derbyshire, K., Fortune, S. M., Gao, L.-Y., and Liu, J. (2009). Systematic genetic nomenclature for type VII secretion systems. *PLoS Pathog.* 5**,** e1000507.

Bland, C., Ramsey, T. L., Sabree, F., Lowe, M., Brown, K., Kyrpides, N. C., and Hugenholtz, P. (2007). CRISPR recognition tool (CRT): a tool for automatic detection of clustered regularly interspaced palindromic repeats. *BMC bioinformatics.* 8**,** 209.

Bolotin, A., Wincker, P., Mauger, S., Jaillon, O., Malarme, K., Weissenbach, J., Ehrlich, S. D., and Sorokin, A. (2001). The complete genome sequence of the lactic acid bacterium *Lactococcus lactis* ssp. *lactis* IL1403. *Genome Res.* 11**,** 731-753.

Boucher, I., Parrot, M., Gaudreau, H., Champagne, C. P., Vadeboncoeur, C., and Moineau, S. (2002). Novel food-grade plasmid vector based on melibiose fermentation for the genetic engineering of *Lactococcus lactis*. *Appl. Environ. Microbiol.* 68**,** 6152-6161.

Bruins, M. R., Kapil, S., and Oehme, F. W. (2000). Microbial resistance to metals in the environment. *Ecotox. Environ. Safe.* 45**,** 198-207.

Butcher, B. G., Lin, Y.-P., and Helmann, J. D. (2007). The *yydFGHIJ* operon of *Bacillus subtilis* encodes a peptide that induces the LiaRS two-component system. *J. Bacteriol.* 189**,** 8616-8625.

Cao, M., Wang, T., Ye, R., and Helmann, J. D. (2002). Antibiotics that inhibit cell wall biosynthesis induce expression of the *Bacillus subtilis* σW and σM regulons. *Mol. Microbiol.* 45**,** 1267-1276.

Carpita, N. C. (1996). Structure and biogenesis of the cell walls of grasses. *Annu. Rev. Plant Biol.* 47**,** 445-476.

Cartier, G., Lorieux, F., Allemand, F., Dreyfus, M., and Bizebard, T. (2010). Cold adaptation in DEAD-box proteins. *Biochemistry.* 49**,** 2636-2646.

Cassan, F., Maiale, S., Masciarelli, O., Vidal, A., Luna, V., and Ruiz, O. (2009). Cadaverine production by *Azospirillum brasilense* and its possible role in plant growth promotion and osmotic stress mitigation. *Eur. J. Soil Biol.* 45**,** 12-19.

Cho, E.-A., Lee, D.-W., Cha, Y.-H., Lee, S.-J., Jung, H.-C., Pan, J.-G., and Pyun, Y.-R. (2007). Characterization of a Novel D-Lyxose Isomerase from *Cohnella laevoribosii* RI-39 sp. nov. *J. Bacteriol.* 189**,** 1655-1663.

Costa, N., Hannon, J., Guinee, T., Auty, M., McSweeney, P., and Beresford, T. (2010). Effect of exopolysaccharide produced by isogenic strains of *Lactococcus lactis* on half-fat Cheddar cheese. *J. Dairy Sci.* 93**,** 3469-3486.

Coton, E., and Coton, M. (2005). Multiplex PCR for colony direct detection of Gram-positive histamine-and tyramine-producing bacteria. *J. Microbiol. Meth.* 63**,** 296-304.

Coton, M., Coton, E., Lucas, P., and Lonvaud, A. (2004). Identification of the gene encoding a putative tyrosine decarboxylase of *Carnobacterium divergens* 508. Development of molecular tools for the detection of tyramine-producing bacteria. *Food Microbiol.* 21**,** 125-130.

Dabour, N., and 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.

Darling, A. E., Mau, B., and Perna, N. T. (2010). progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PloS one.* 5**,** e11147.

Eddy, S. R. (2009). A new generation of homology search tools based on probabilistic inference. *Genome Inform*. 23, 205-211.

EFSA Panel on Additives and Products or Substances used in Animal Feed (FEEDAP). (2012). Guidance on the assessment of bacterial susceptibility to antimicrobials of human and veterinary importance. *EFSA Journal*. 10 (6), 2740. [10 pp.] DOI: 10.2903/j.efsa.2012.2740.

EFSA Panel on Biological Hazards (BIOHAZ). (2011). Scientific Opinion on risk based control of biogenic amine formation in fermented foods. *EFSA Journal* 9(10), 2393. [93 pp.]. DOI: 10.2903/j.efsa.2011.2393.

Fallico, V., McAuliffe, O., Fitzgerald, G. F., and Ross, R. P. (2011). Plasmids of raw milk cheese isolate *L. lactis* subsp. *lactis* biovar. *diacetylactis* DPC3901 suggest a plant-based origin for the strain. *Appl. Environ. Microbiol.* 77, 6451-6462.

Gamarnik, A., and Frydman, R. B. (1991). Cadaverine, an essential diamine for the normal root development of germinating soybean (Glycine max) seeds. *Plant Physiol.* 97**,** 778-785.

Gao, Y., Lu, Y., Teng, K.-L., Chen, M.-L., Zheng, H.-J., Zhu, Y.-Q., and Zhong, J. (2011). Complete genome sequence of *Lactococcus lactis* ssp. *lactis* CV56, a probiotic strain isolated from the vaginas of healthy women. *J. Bacteriol.* 193**,** 2886-2887. DOI: 10.1128/JB.00358-11.

Górecki, R. K., Koryszewska-Bagińska, A., Gołębiewski, M., Żylińska, J., Grynberg, M., and Bardowski, J. K. (2011). Adaptative potential of the *Lactococcus lactis* IL594 strain encoded in its 7 plasmids. *PloS one.* 6**,** e22238. DOI: 10.1371/journal.pone.0022238.

Gu, Y., Ding, Y., Ren, C., Sun, Z., Rodionov, D. A., Zhang, W., Yang, S., Yang, C., and Jiang, W. (2010). Reconstruction of xylose utilization pathway and regulons in Firmicutes. *BMC genomics.* 11**,** 1-14.

Gubler, M., Braguglia, D., Meyer, J., Piekarowicz, A., and Bickle, T. A. (1992). Recombination of constant and variable modules alters DNA sequence recognition by type IC restriction-modification enzymes. *EMBO J.* 11**,** 233-240.

Hill, C. (2012). Virulence or niche factors: what's in a name? *J. Bacteriol.* 194**,** 5725-5727.

Houben, E. N., Bestebroer, J., Ummels, R., Wilson, L., Piersma, S. R., Jiménez, C. R., Ottenhoff, T. H., Luirink, J., and Bitter, W. (2012). Composition of the type VII secretion system membrane complex. *Molecular Microbiol.* 86**,** 472-484.

Hyatt, D., Chen, G.-L., Locascio, P. F., Land, M. L., Larimer, F. W., and Hauser, L. J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC bioinformatics.* 11**,** 119.

Kalmokoff, M., Bartlett, F., and Teather, R. (1996). Are ruminal bacteria armed with bacteriocins? *J. Dairy Sci.* 79**,** 2297-2306.

Kato, H., Shiwa, Y., Oshima, K., Machii, M., Araya-Kojima, T., Zendo, T., Shimizu-Kadota, M., Hattori, M., Sonomoto, K., and 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. DOI: 10.1128/JB.00074-12.

Kelly, W. J., Altermann, E., Lambie, S. C., and Leahy, S. C. (2013). Interaction between the genomes of *Lactococcus lactis* and phages of the P335 species. *Front. Microbiol.* 4, 1-9.

Kelly, W. J., Davey, G. P., and Ward, L. J. H. (1998). Characterization of lactococci isolated from minimally processed fresh fruit and vegetables. *Int. J. Food Microbiol.* 45**,** 85-92.

Kelly, W. J., Ward, L. J. H., and Leahy, S. C. (2010). Chromosomal diversity in *Lactococcus lactis* and the origin of dairy starter cultures. *Genome Biol. Evol.* 2**,** 729-744.

Kjos, M., Snipen, L., Salehian, Z., Nes, I. F., and Diep, D. B. (2010). The Abi proteins and their involvement in bacteriocin self-immunity. *J. Bacteriol.* 192**,** 2068-2076.

Klijn, N., Weerkamp, A. H., and De Vos, W. M. (1995). Detection and characterization of lactose-utilizing *Lactococcus* spp. in natural ecosystems. *Appl. Environ. Microbiol.* 61**,** 788-792.

Kwon, H.-J., Yeom, S.-J., Park, C.-S., and Oh, D.-K. (2010). Substrate specificity of a recombinant D-lyxose isomerase from *Providencia stuartii* for monosaccharides. *J. Biosci. Bioeng.* 110**,** 26-31.

Ladero, V., Calles-Enriquez, M., Fernández, M., and A Alvarez, M. (2010). Toxicological effects of dietary biogenic amines. *Curr. Nut. Food Sci.* 6**,** 145-156.

Ladero, V., Rattray, F. P., Mayo, B., Martín, M. C., Fernández, M., and Alvarez, M. A. (2011). Sequencing and transcriptional analysis of the biosynthesis gene cluster of putrescine-producing *Lactococcus lactis*. *Appl. Environ. Microbiol.* 77**,** 6409-6418. DOI: 10.1128/AEM.05507-11.

Lamarque, M., Aubel, D., Piard, J.-C., Gilbert, C., Juillard, V., and Atlan, D. (2011). The peptide transport system Opt is involved in both nutrition and environmental sensing during growth of *Lactococcus lactis* in milk. *Microbiol.* 157**,** 1612-1619.

Liu, M., Bayjanov, J. R., Renckens, B., Nauta, A., and Siezen, R. J. (2010). The proteolytic system of lactic acid bacteria revisited: a genomic comparison. *BMC Genomics.* 11**,** 36. DOI: 10.1186/1471-2164-11-36.

Lombard, V., Ramulu, H. G., Drula, E., Coutinho, P. M., and Henrissat, B. (2014). The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic acids Res.* 42**,** D490-D495.

Marilley, L., and Casey, M. (2004). Flavours of cheese products: metabolic pathways, analytical tools and identification of producing strains. *Int. J. Food Microbiol.* 90**,** 139-159.

Markkula, A., Mattila, M., Lindström, M., and Korkeala, H. (2012). Genes encoding putative DEAD-box RNA helicases in *Listeria monocytogenes* EGD-e are needed for growth and motility at 3° C. *Environ. Microbiol.* 14**,** 2223-2232.

Marshall, V. M. (1991). Inoculated ecosystem in a milk environment. *J. Appl. Bacteriol.* 73**,** 127S-135S.

Mascher, T., Margulis, N. G., Wang, T., Ye, R. W., and Helmann, J. D. (2003). Cell wall stress responses in *Bacillus subtilis*: the regulatory network of the bacitracin stimulon. *Mol. Microbiol.*50**,** 1591-1604.

Mascher, T., Zimmer, S. L., Smith, T.-A., and Helmann, J. D. (2004). Antibiotic-inducible promoter regulated by the cell envelope stress-sensing two-component system LiaRS of Bacillus subtilis. *Antimicrob. Agents Ch.* 48**,** 2888-2896.

McDowell, L. R. (1996). Feeding minerals to cattle on pasture. *Anim. Feed Sci. Tech.* 60**,** 247-271.

Meneses, C. H., Rouws, L. F., Simões-Araújo, J. L., Vidal, M. S., and Baldani, J. I. (2011). Exopolysaccharide production is required for biofilm formation and plant colonization by the nitrogen-fixing endophyte *Gluconacetobacter diazotrophicus*. *Mol. Plant Microbe In.* 24**,** 1448-1458.

Millen, A. M., Horvath, P., Boyaval, P., and Romero, D. A. (2012). Mobile CRISPR/Cas-mediated bacteriophage resistance in *Lactococcus lactis*. *PloS One.* 7**,** e51663.

Mills, S., McAuliffe, O. E., Coffey, A., Fitzgerald, G. F., and Ross, R. P. (2006). Plasmids of lactococci–genetic accessories or genetic necessities? *FEMS Microbiol. Rev.* 30**,** 243-273. DOI: 10.1111/j.1574-6976.2005.00011.x.

Nomura, M., Kobayashi, M., Narita, T., Kimoto-Nira, H., and Okamoto, T. (2006). Phenotypic and molecular characterization of *Lactococcus lactis* from milk and plants. *J. Appl. Microbiol.* 101**,** 396-405. DOI: 10.1111/j.1365-2672.2006.02949.x.

O'Sullivan, O., O'Callaghan, J., Sangrador-Vegas, A., Mc Auliffe, O., Slattery, L., Kaleta, P., Callanan, M., Fitzgerald, G. F., Ross, R. P., and Beresford, T. (2009). Comparative genomics of lactic acid bacteria reveals a niche-specific gene set. *BMC Microbiol.* 9 (50), 1-9.

O'Sullivan, D. J., and Klaenhammer, T. R. (1993). Rapid mini-prep isolation of high-quality plasmid DNA from *Lactococcus* and *Lactobacillus* spp. *Appl. Environ. Microbiol.* 59**,** 2730-2733.

Passerini, D., Beltramo, C., Coddeville, M., Quentin, Y., Ritzenthaler, P., Daveran-Mingot, M. L., and Le Bourgeois, P. (2010). Genes but not genomes reveal bacterial domestication of *Lactococcus lactis*. *PloS One.* 5**,** e15306.

Passerini, D., Coddeville, M., Le Bourgeois, P., Loubière, P., Ritzenthaler, P., Fontagné-Faucher, C., Daveran-Mingot, M.-L., and Cocaign-Bousquet, M. (2013). The carbohydrate metabolism signature of *Lactococcus lactis* strain A12 reveals its sourdough ecosystem origin. *Appl. Environ. Microbiol.* 79**,** 5844-5852. DOI: 10.1128/AEM.01560-13.

Rodionov, D. A., Hebbeln, P., Gelfand, M. S., and Eitinger, T. (2006). Comparative and functional genomic analysis of prokaryotic nickel and cobalt uptake transporters: evidence for a novel group of ATP-binding cassette transporters. *J. Bacteriol.* 188**,** 317-327.

Rodríguez-Navarro, D. N., Dardanelli, M. S., and Ruíz-Saínz, J. E. (2007). Attachment of bacteria to the roots of higher plants. *FEMS Microbiol. Lett.* 272**,** 127-136.

Rossbach, S., Kukuk, M. L., Wilson, T. L., Feng, S. F., Pearson, M. M., and Fisher, M. A. (2000). Cadmium-regulated gene fusions in *Pseudomonas fluorescens*. *Environ. Microbiol.* 2**,** 373-382.

Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M.-A., and Barrell, B. (2000). Artemis: sequence visualization and annotation. *Bioinformatics.* 16**,** 944-945.

Salem, A., Ammar, H., Lopez, S., Gohar, Y., and González, J. S. (2011). Sensitivity of ruminal bacteria isolates of sheep, cattle and buffalo to some heavy metals. *Anim. Feed Sci. Tech.* 163**,** 143-149.

Sanders, J. W., Venema, G., and Kok, J. (1999). Environmental stress responses in *Lactococcus lactis*. *FEMS Microbiol. Rev.* 23**,** 483-501.

Seegers, J. F., Van Sinderen, D., and Fitzgerald, G. F. (2000). Molecular characterization of the lactococcal plasmid pCIS3: natural stacking of specificity subunits of a type I restriction/modification system in a single lactococcal strain. *Microbiol.* 146**,** 435-443.

Seiboth, B., and Metz, B. (2011). Fungal arabinan and L-arabinose metabolism. *Appl. Microbiol Biotechnol.* 89**,** 1665-1673.

Siezen, R. J., Starrenburg, M. J., Boekhorst, J., Renckens, B., Molenaar, D., and Van Hylckama Vlieg, J. E. (2008). Genome-scale genotype-phenotype matching of two *Lactococcus lactis* isolates from plants identifies mechanisms of adaptation to the plant niche. *Appl. Environ. Microbiol.* 74**,** 424-436.

Singh, T., Drake, M., and Cadwallader, K. (2003). Flavor of Cheddar cheese: a chemical and sensory perspective. *Comprehensive Reviews in Food Science and Food Safety.* 2**,** 166-189.

Smit, G., Smit, B. A., and Engels, W. J. (2005). Flavour formation by lactic acid bacteria and biochemical flavour profiling of cheese products. *FEMS Microbiol. Rev.* 29**,** 591-610. DOI: 10.1016/j.fmrre.2005.04.002.

Stern, A., and Sorek, R. (2011). The phage-host arms race: Shaping the evolution of microbes. *Bioessays.* 33**,** 43-51.

Tiffany, M., Spears, J., Xi, L., and Horton, J. (2003). Influence of dietary cobalt source and concentration on performance, vitamin B12 status, and ruminal and plasma metabolites in growing and finishing steers. *J. Anim. Sci.* 81**,** 3151-3159.

Van De Guchte, M., Serror, P., Chervaux, C., Smokvina, T., Ehrlich, S. D., and Maguin, E. (2002). Stress responses in lactic acid bacteria. *Antonie Van Leeuwenhoek.* 82**,** 187-216.

Van Den Brink, J., and De Vries, R. P. (2011). Fungal enzyme sets for plant polysaccharide degradation. *Appl. Microbiol. Biotech*. 91**,** 1477-1492.

Van Kranenburg, R., Kleerebezem, M., Van Hylckama Vlieg, J., Ursing, B. M., Boekhorst, J., Smit, B. A., Ayad, E. H. E., Smit, G., and Siezen, R. J. (2002). Flavour formation from amino acids by lactic acid bacteria: predictions from genome sequence analysis. *Int. Dairy J.* 12**,** 111-121.

Wang, G., Li, X. and Wang, Z. (2009) APD2: the updated antimicrobial peptide database and its application in peptide design. *Nucleic Acids Res*. 37, D933-D937.

Wang, Z. and Wang, G. (2004) APD: the antimicrobial peptide database. *Nucleic Acids Res*. 32, D590-D592

Wegmann, U., O'connell-Motherway, M., Zomer, A., Buist, G., Shearman, C., Canchaya, C., Ventura, M., Goesmann, A., Gasson, M. J., and Kuipers, O. P. (2007). Complete genome sequence of the prototype lactic acid bacterium *Lactococcus lactis* subsp. *cremoris* MG1363. *J. Bacteriol.* 189**,** 3256-3270. DOI: 10.1128/JB.01768-06.

Whistler, R. L., and Richards, E. L. (1970). "Hemicelluloses," in *The Carbohydrates-Chemistry and Biochemistry,* eds. W. Pigman and D. Horton. (New York: Academic press), 447-469.

Wolin, M., Miller, T., and Stewart, C. (1997). "Microbe-microbe interactions," in *The rumen microbial ecosystem,* eds. P. M. Hobson and C. S. Stewart (Netherlands, Springer), 467-491.

Wouters, J. A., Frenkiel, H., De Vos, W. M., Kuipers, O. P., and Abee, T. (2001). Cold shock proteins of *Lactococcus lactis* MG1363 are involved in cryoprotection and in the production of cold-induced proteins. *Appl. Environ. Microbiol.* 67**,** 5171-5178.

Yvon, M., and Rijnen, L. (2001). Cheese flavour formation by amino acid catabolism. *Int. Dairy J.* 11**,** 185-201.

Table 1. General features and sequencing statistics of DPC genomes.

	L. lactis DPC6853	*L. lactis* DPC6856	*L. lactis* DPC6860
Approximate size (Mb)	2.5	2.9	2.62
% G+C	35%	35.3%	35.6%
Total assembled reads	240887	223535	210608
Total no. of contigs	116	163	104
Avg. length of assembled reads	414	422	416
Avg. quality of assembled reads	29	29	28
Contig N_{50} score	40kb	39kb	57kb
Fold coverage	40x	32x	33x
No. of non-alignment contigs[a]	31	57	30
No. of plasmids	2[b]	7[c]	4[d]

[a] Non-alignment contigs were mapped against the reference strains IL1403 and MG1363

[b] plasmid sizes: ~42 kb, ~75 kb.

[c] plasmid sizes: ~2 kb, ~3 kb, ~5 kb, ~25 kb, ~48 kb, ~50 kb, ~51 kb.

[d] plasmid sizes ~25 kb, ~40 kb, ~45 kb, ~51 kb.

Table 2. Carbohydrate fermentation profiles of DPC strains and the type strains MG1363 (subspecies *cremoris*) and IL1403 (subspecies *lactis*).

	DPC6853	DPC6856	DPC6860	MG1363	IL1403
Monosaccharides					
D -arabinose	−	−	−	−	−
L-arabinose	−	+	−	−	−
D -xylose	−	−	++++	−	−
D -galactose	+++	++++	++++	++++	+++
D -mannose	++++	+++	++++	++++	++
D -ribose	+++	−	+++	−	++
rhamnose	−	−	−	−	−
lyxose	−	−	−	−	−
D -glucose	+++	+++	++++	++++	+++
D -fructose	++++	++++	++++	++++	+++
Disaccharides					
D -maltose	+++	++++	+++	+++	++
cellobiose	+++	++++	+++	++++	++
lactose	++++	++++	++++	−	−
trehalose	++++	++++	++++	++++	++
melibiose	−	+	−	−	−
sucrose	++++	−	−	−	−
lactulose	++++	+++	++++	−	−
Trisaccharides					
raffinose	−	−	−	−	−
melezitose	−	−	−	−	−
Polyol					
D -mannitol	−	−	+++	+++	−
xylitol	−	−	−	+	−
Oligosaccharide					
Inulin	++	++	+	−	+

Table 3. Putative cold shock proteins (CSPs) identified in DPC lactococcal strains.

Best Blast hit	Contig	E-value	Amino acid identity (%)	Size (aa)	MW (Da)
DPC6860					
cold shock protein C [*L. lactis* ssp. *cremoris* GE214]	G37	1e-42	69/ 69 (100)	69	8135
cold shock protein CspD [*L. lactis* ssp. *lactis* IO-1]	G37	2e-39	66/ 66 (100)	66	7198
cold shock domain protein CspD4 [*L. lactis* ssp. *cremoris* KW2]	G40	2e-39	66/ 66 (100)	66	7198
cold shock domain protein CspD3 [*L. lactis* ssp. *cremoris* KW2]	G51	3e-39	66/ 66 (100)	66	7268
cold shock domain protein CspD2 [*L. lactis* ssp. *cremoris* KW2]	G51	2e-40	66/ 66 (100)	66	7137
cold shock domain protein CspD1 [*L. lactis* ssp. *cremoris* KW2]	G7	1e-38	65/ 65 (100)	65	7137
DPC6856					
cold-shock protein [*L. lactis* ssp. *cremoris* GE214]	R66	9e-39	66/ 66 (100)	66	7216
cold shock domain protein CspD4 [*L. lactis* ssp. *cremoris* KW2]	R52	1e-39	65/ 65 (100)	66	7198
cold shock domain protein CspD1 [*L. lactis* ssp. *cremoris* KW2]	R6	1e-38	65/ 65 (100)	65	7137
cold shock protein CspC [*L. lactis* ssp. *lactis* CV56]	R53	2e-40	66/ 66 (100)	66	7677
cold shock protein CspC [*L. lactis* ssp. *lactis* CV56]	R54	2e-40	66/ 66 (100)	66	7677
cold-shock protein [*L. lactis* ssp. *lactis* KF147]	R54	2e-38	64/ 66 (97)	66	7242
cold shock-like protein cspA [*L. lactis* ssp. *cremoris* MG1363]	R72	2e-40	66/ 66 (100)	66	7589
cold shock domain protein CspD2 [*L. lactis* ssp. *cremoris* KW2]	R72	2e-40	66/ 66 (100)	66	7589
cold shock protein D [*L. lactis ssp. lactis*]	R115	2e-51	84/ 91 (92)	90	10382
DPC6853					
cold shock protein CspC [*L. lactis* ssp. *lactis* CV56]	C48	2e-40	66/ 66 (100)	66	7677
cold shock-like protein cspD2 [*L. lactis* ssp. *cremoris* MG1363]	C48	2e-39	66/ 66 (100)	66	7198
cold shock protein CspD [*L. lactis* ssp. *lactis* IO-1]	C67	8e-39	66/ 66 (100)	66	7200
cold-shock protein [*L. lactis* ssp. *lactis* KF147]	C7	1e-38	65/ 65 (100)	65	7123

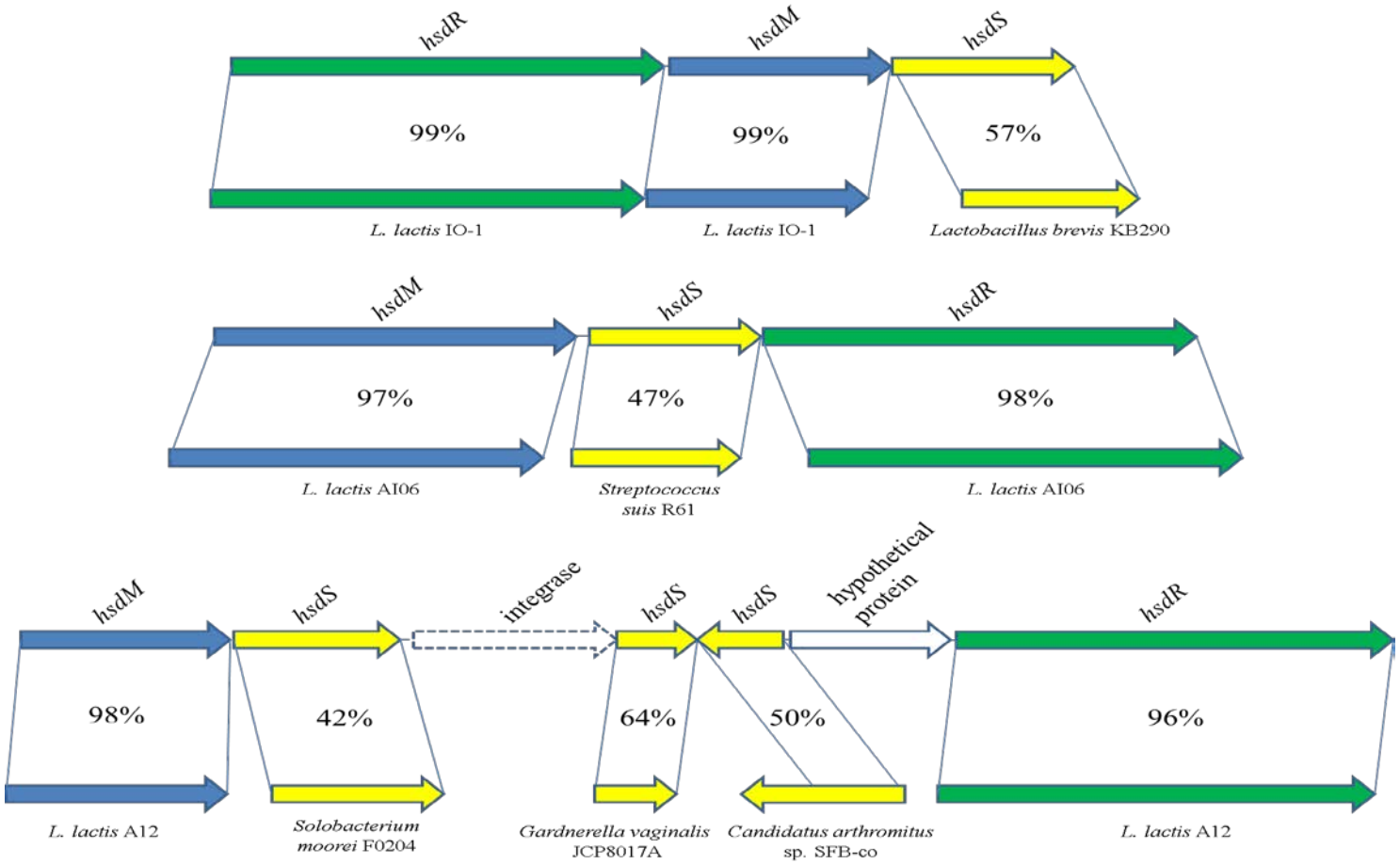


Fig. 1. Type I restriction modification (R-M) systems identified in *L. lactis* DPC6853. Values between ORFs indicate % amino acid identity.

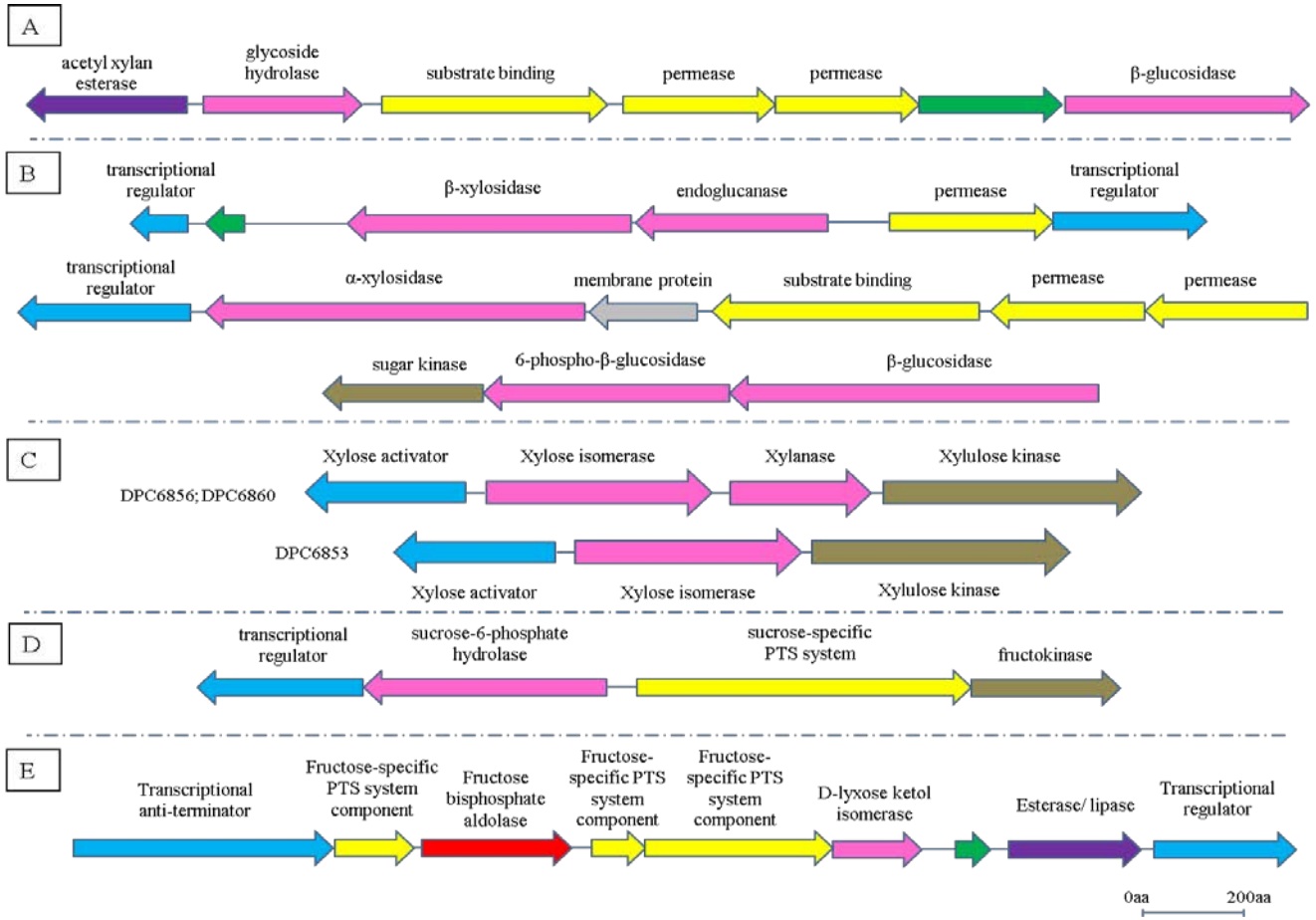


Fig. 2. Putative operons encoding for carbohydrate/ sugar utilisation identified in (A) DPC6853 and DPC6860, (B) DPC6860 and DPC6856, (C) DPC6856 and DPC6860, (D) DPC6853, (E) DPC6856. lyase; transcription regulator; PTS component; esterase; kinase; hypothetical; membrane associated.

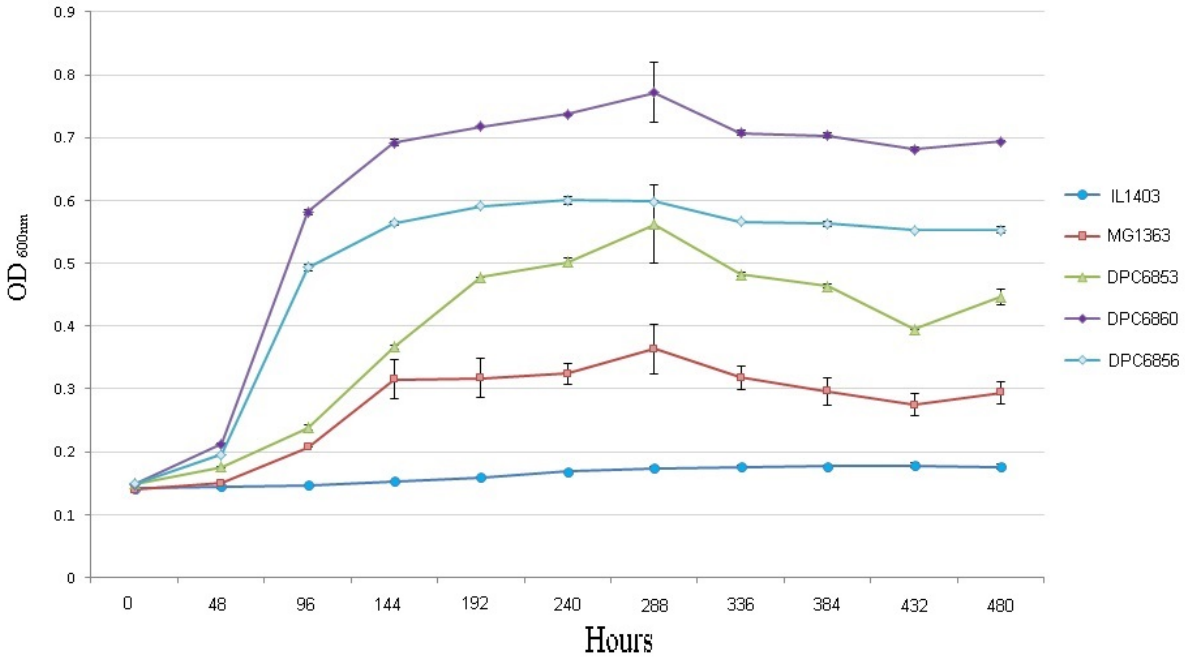


Fig. 3. Growth of dairy strains (MG1363, IL1403) and non-dairy strains (DPC6853, DPC6856, DPC6860) at 6 ºC over 20 days.

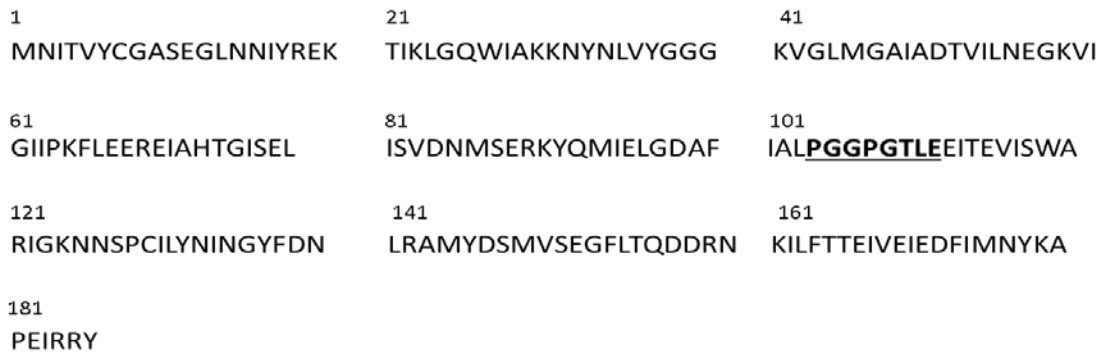


Fig. 4. PGGXGTXXE motif identified in the putative lysine decarboxylase gene of *L. lactis* DPC6853.

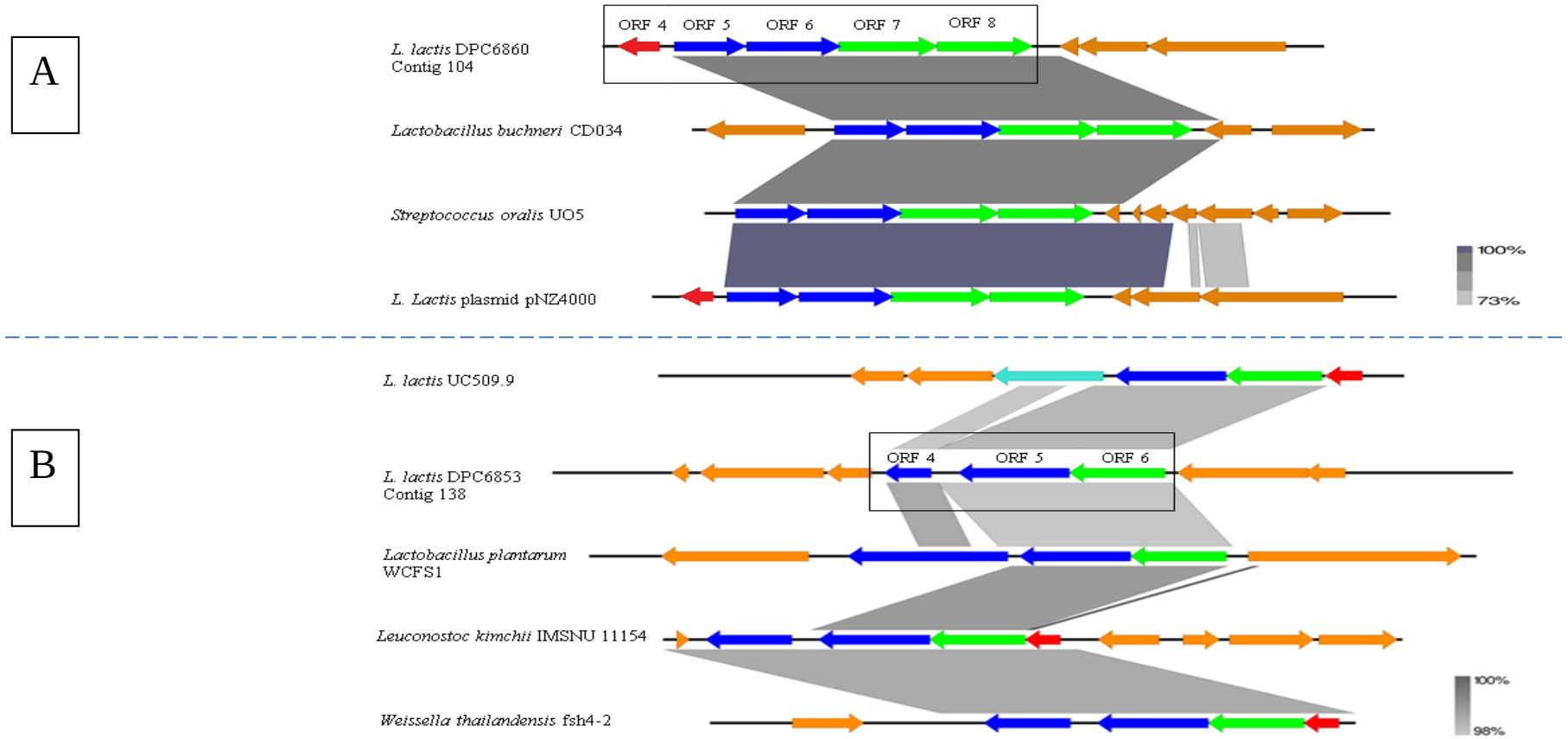


Fig. 5. Cobalt (Co^{2+}) transport systems identified in (A) DPC6853 and (B) DPC6860. No putative Co^{2+} transport systems were identified in DPC6856. ATP binding, membrane associated, pseudogene, possible CbiN protein, non-transport associated proteins.

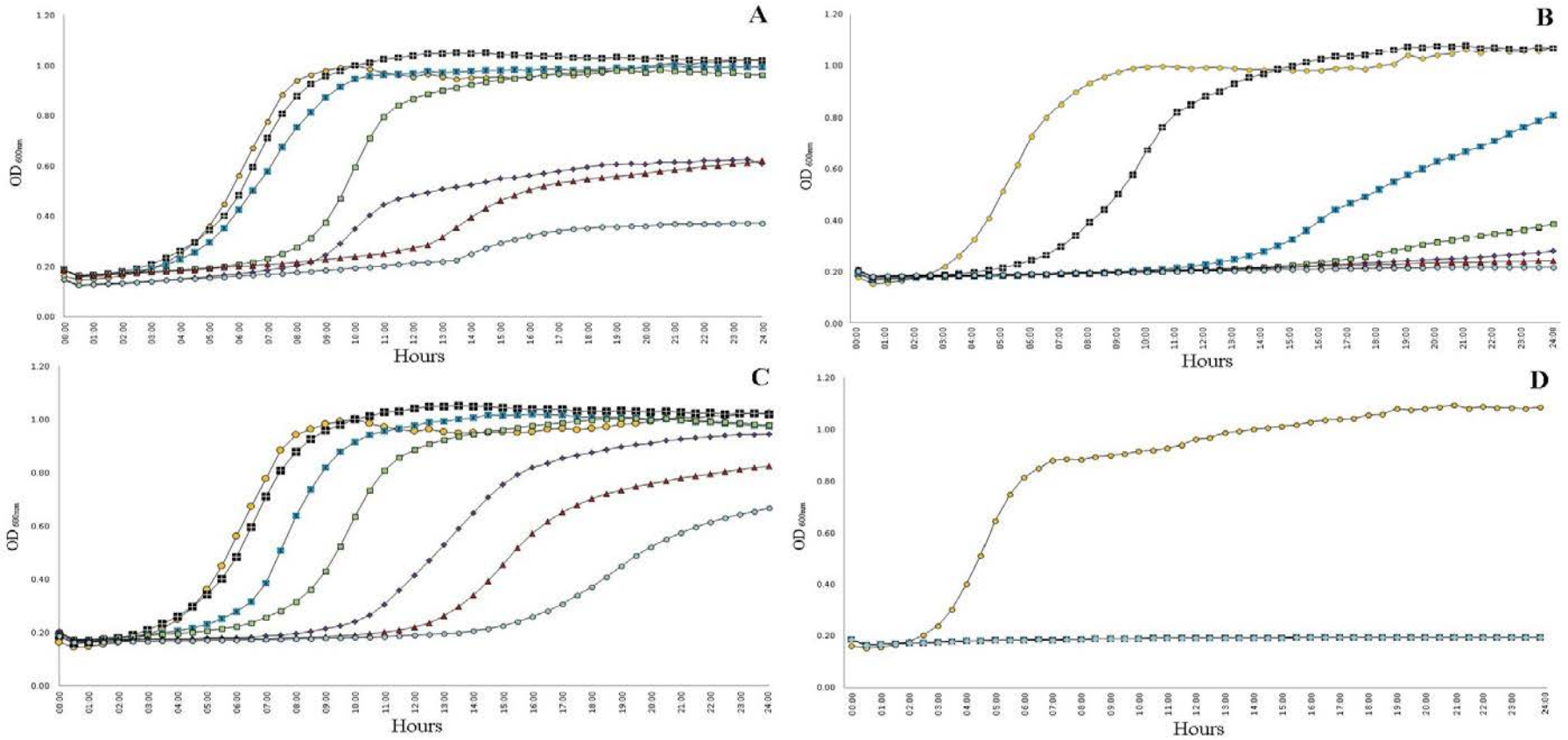


Fig. 6. Growth of non-dairy lactococci at increasing concentrations of Co^{2+} at 30 ºC. (A) DPC6860, (B) DPC6856, (C) DPC6853, (D) IL1403. 0 mM ( ), 1 mM ( ), 2 mM ( ), 3 mM ( ), 4 mM ( ), 5 mM ( ), 6mM ( )The average of duplicate experiments is presented.

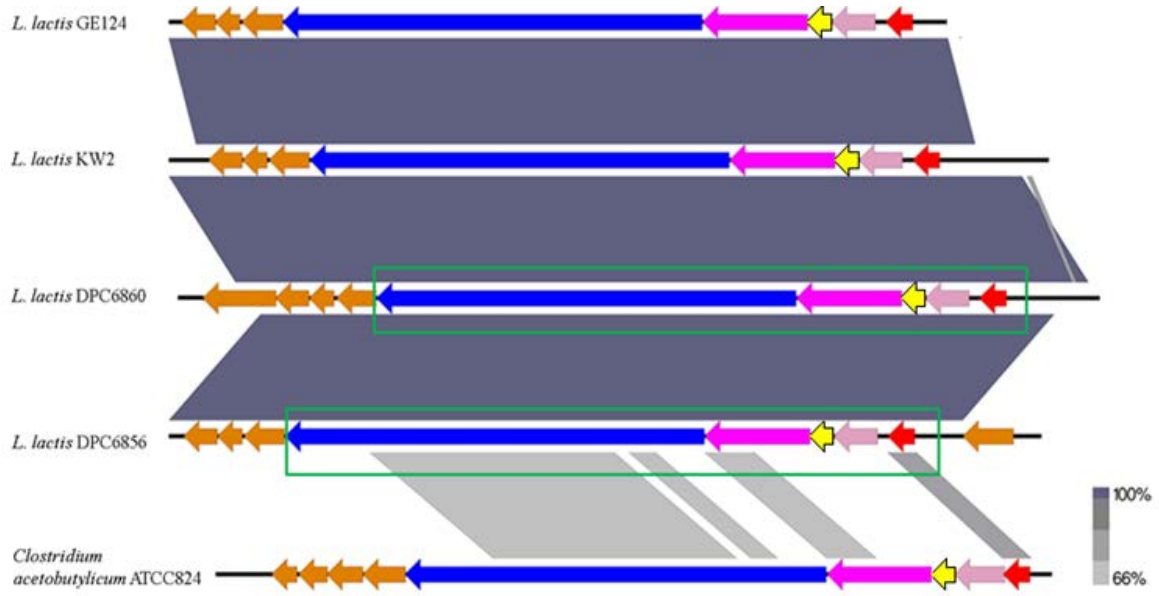


Fig. 7. Alignment of WXG secretion system (Wss) identified in *L. lactis* DPC6860 and DPC6856. Non-secretion system associated ORFs are shown in orange.

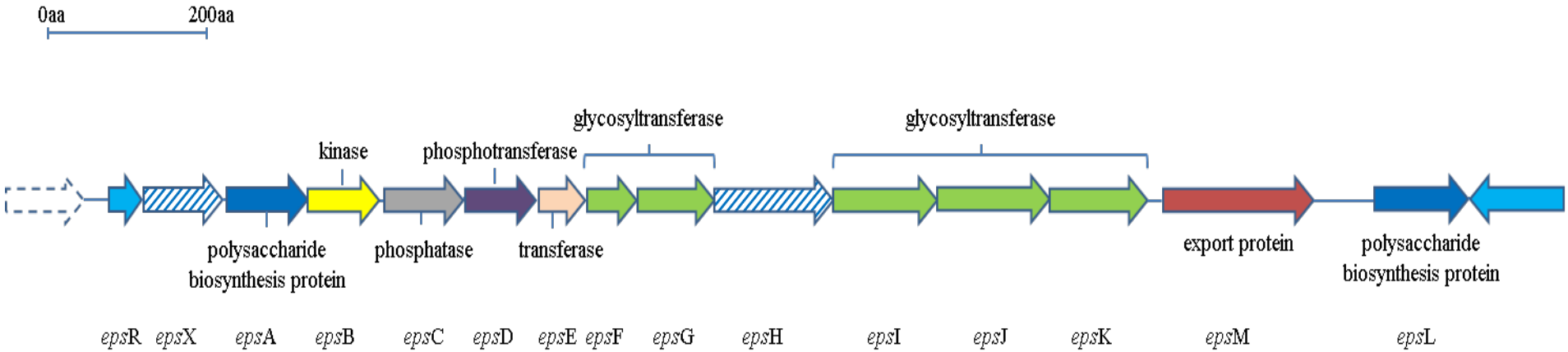


Fig. 8. Exopolysaccharide (EPS) biosynthesis gene cluster identified in strain DPC6860. hypothetical protein, transcription regulator, mobile element protein, glycosyltransferase, polysaccharide biosynthesis protein, tyrosine kinase, tyrosine phosphatase, undecaprenyl phosphategalactose phosphotransferase, acetylglucosamine transferase.

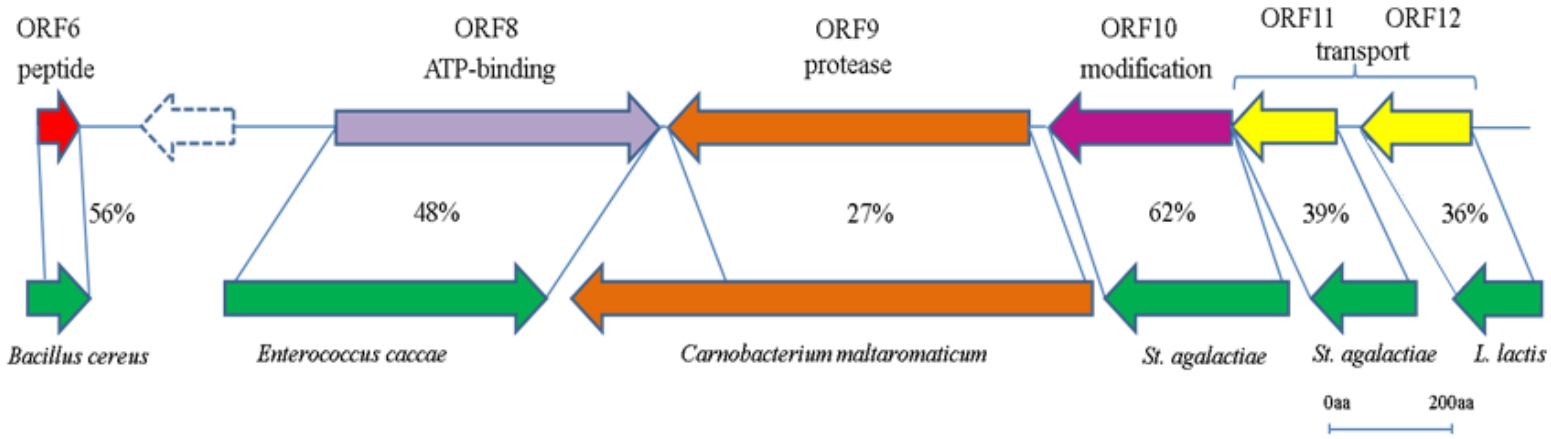


Fig. 9.Putative *yydFGHIJ* like gene cluster identified on contig R163 of DPC6856. hypothetical, transport, mobile element protein, protease, ATP binding, modification.

Name/ Class	Source	Similarity percentage	APD[a] ID
Cyto-insectotoxin 1a	venom, *Lachesana tarabaevi*	30.76 %	AP02163
Query sequence:	+GFFGNTWKKIKGKADKIMLKKAVKIMVKKEGISKE++EAQ++AKVDAMSK+KQIRLYLLKYYGKKALQKASEKL+++		
Reference sequence:	MANF+N+++ +AFEKA+KFETKNNYKI+PKK++I+KERKE+QMSIMLTPTNKEK+IR++SL++ +ADEANMSVSE+LIAY		
Latarcin 6b	*Lachesana tarabaevi*	29.03 %	AP01020
Query sequence:	++++QAF+K+TF+TPDWN+KIRNDAK+R+++MQDNL+EQMK+K+RFNL++++N+++++L+++		
Reference sequence:	MANFNAFEKAKFET+KNNYKIPKKIKERKEQMSIMLTPTNKEKIR+SLADEANMSVSELIAY		
Im-1	venom, *Isometrus maculatus*	28.98 %	AP00482
Query sequence:	+++F+S F+KRLKGFAKKLWNSKLARKI++RTK+++++GLKYVKNFAKDMLSEGEEA+PPAAEPPVEAPQ		
Reference sequence:	MANFNAFEK+AK+FETK+NNYKIPKKIKER+KEQMSIML+TPTNKEK+I R SLADEANMSVSE++LIA+Y		
SPP-3	*Caenorhabditis elegans*	28.73 %	AP02227
Query sequence:	++NGIECEMCKMSVKIVVPMLGEDT+++ESI+KKAVDAECKKEFHSIPFGTQECK+KFIDTKL+DPIIHELENGTAPSDVCTKLGMC		
Reference sequence:	MANFNAFE++ K ++AK+ +++ ++FE+TKNNYKIPKK++IKE+RKEQMSI+MLTPTNKEK+I+RSLAD++++E+AN+ ++MS+V+SELIAY		
Enterocin L50	*Enterococcus faecium* L50	28.57 %	AP01190
Query sequence:	MGA++IAKLVAKF++++GWPI+VK+KYYK+Q++IM++QFIGE+++GWA++IN+KIIEWIKKHI		
Reference sequence:	M+ANFNAFEKAKFETKNNYKIPKKIKERKEQMSIMLTPTNKEKIRSLADEANMSVSELI++AY		

Fig. 10. Alignment of the putative antimicrobial compound identified in DPC6856 using the Antimicrobial Peptide Database[a] (Wang, 2010, Wang and Wang, 2004).

5.8. Appendices

Appendix 5-1

ORF table of non-alignment contigs to reference genomes for DPC6856 (Rumen)

ORF	Start	Stop	Size (aa)	Predicted function	Identity	Accession no.	E-value	Size
Contig_R104								
1	3	464	153	Gls24 family general stress protein [*Lactococcus lactis* ssp. *lactis* KF147]	139/ 153 (91)	YP_003354674.1	2e-75	183
2	1851	655	398	plasmid replication initiation protein [*Lactococcus lactis*]	394/ 398 (99)	WP_023349300.1	0.0	398
3	2284	3132	282	replication-associated protein [*Lactococcus lactis* ssp. *lactis* biovar. *diacetylactis* LD61]	282/ 282 (100)	YP_001966493.1	0.0	282
4	3151	3450	99	hypothetical protein [*Lactococcus lactis*]	98/ 99 (99)	WP_023349301.1	8e-63	99
5	3865	4545	226	insertion sequence IS946 transposase [*Lactococcus lactis*]	223/ 226 (99)	NP_047328.1	8e-162	226
6	4580	5178	208	transposase element IS1297 [*Streptococcus thermophilus* MN-ZLW-002]	177/ 204 (87)	YP_006339954.1	7e-117	226
7	5318	6288	323	oxalate/ formate antiporter [*Lactococcus lactis* ssp. *cremoris* MG1363]	321/ 324 (99)	YP_001032006.1	0.0	418
Contig_R115								
1	542	303	79	membrane protein [*Lactococcus lactis*]	77/ 79 (97)	WP_017864515.1	5e-41	79
2	895	623	90	transposase, partial [*Lactococcus lactis* ssp. *cremoris* TIFN5]	70/ 78 (90)	WP_021213978.1	4e-41	97
3	3021	1672	449	di-hydrolipoamide dehydrogenase [*Lactococcus lactis* ssp. *cremoris*]	449/ 449 (100)	YP_006965841.1	0.0	449
4	3344	3039	101	hypothetical protein pLP712_46 [*Lactococcus lactis* ssp. *cremoris*]	101/ 128 (79)	YP_006965842.1	2e-60	128
5	3993	3370	207	isochorismatase family protein [*Lactococcus lactis* ssp. *cremoris*]	207/ 207 (100)	YP_006965843.1	1e-152	207
6	4184	4020	54	hypothetical protein EF62_0834 [*Enterococcus faecalis* 62]	50/ 54 (93)	YP_005704819.1	9e-27	54
7	4848	4258	196	putative HTH-type transcriptional regulator [*Lactococcus lactis* ssp. *cremoris*]	196/ 196 (100)	YP_006965845.1	1e-139	196
8	5056	5736	226	transposase [*Enterococcus durans*]	221/ 226 (98)	WP_016177626.1	6e-162	226
9	6286	5849	145	universal stress protein UspA-like nucleotide-binding protein [*Lactococcus lactis* ssp. *cremoris* SK11]	145/ 145 (100)	YP_796557.1	1e-98	145
10	7902	6295	535	Mn^{2+}/ Fe^{2+} transporter [*Lactococcus lactis* ssp. *cremoris* A76]	532/ 535 (99)	YP_005877251.1	0.0	535
Contig_R103								
1	33	2036	667	Phage replicative DNA helicase, repA [*Lactococcus garvieae*]	526/ 664 (79)	WP_004260205.1	0.0	764
2	2363	2584	73	hypothetical protein [*Lactococcus garvieae*]	59/ 73 (81)	WP_004260212.1	1e-35	74
3	2577	3209	210	hypothetical protein CVCAS_0996 [*Lactococcus lactis* ssp. *lactis* CV56]	126. 221 (57)	YP_005868379.1	2e-64	224
4	3708	3848	118	hypothetical protein [*Lactococcus garvieae*]	106/ 118 (90)	WP_004260223.1	8e-70	118
5	3209	3565	46	peptidase M56 [*Desulfitobacterium hafniense*]	13/ 33 (39)	WP_005813145.1	9.2	617
6	3857	4411	184	hypothetical protein [*Lactococcus garvieae*]	135/ 182 (74)	WP_017369939.1	5e-96	182
7	4415	5215	266	hypothetical protein [*Lactococcus garvieae*]	139/ 242 (57)	WP_017369938.1	9e-86	269

8	5225	5704	159	hypothetical protein Q33_0030 [*Lactococcus* phage ΦQ33]	143/ 152 (94)	AFV51059.1	2e-101	152
Contig_R105								
1	189	16	57	hypothetical protein kw2_0196 [*Lactococcus lactis* ssp. *cremoris* KW2]	39/ 43 (91)	YP_008567868.1	3e-07	458
2	376	831	151	transposase for insertion sequence element IS904 [*Lactococcus lactis* ssp. *cremoris* A76]	139/ 146 (95)	YP_005875645.1	3e-100	152
3	1924	974	316	ribonucleoside-diphosphate reductase [*Lactococcus lactis* ssp. *lactis* A12]	315/ 316 (99)	YP_001174719.1	0.0	316
4	2193	1999	64	lipid hydroperoxide peroxidase [*Streptococcus infantarius*]	51/ 55 (93)	WP_006532146.1	5e-26	163
5	2706	2332	124	transposase [*Lactococcus lactis* ssp. *cremoris* TIFN3]	123/ 124 (99)	WP_021215593.1	6e-85	126
Contig_R110								
1	111	791	226	transposase [*Enterococcus durans*]	220/ 226 (97)	WP_016177626.1	9e-160	226
2	885	1460	191	glycopeptide antibiotics resistance protein [*Lactococcus lactis* ssp. *cremoris* UC509.9]	190/ 191 (99)	YP_007000490.1	2e-130	191
3	240	1725	251	oxalate/ formate antiporter [*Lactococcus lactis* ssp. *cremoris* UC509.9]	213/ 213 (100)	YP_007000511.1	2e-145	418
Contig_R51								
1	79	465	128	hypothetical protein, partial [*Lactococcus lactis*]	127/ 128 (99)	WP_021215402.1	3e-86	128
2	1232	597	211	PIL6_6 [*Lactococcus lactis* ssp. *lactis*]	206/ 211 (98)	YP_006962314.1	7e-144	211
3	1318	1665	115	transposase [*Streptococcus thermophilus* M17PTZA496]	115/ 115 (100)	ETW90402.1	5e-75	148
4	1683	1964	93	hypothetical protein, partial [*Lactococcus garvieae*]	87/ 90 (97)	WP_017370935.1	1e-54	151
Contig_R114								
1	167	1165	332	hypothetical protein kw2_1451 [*Lactococcus lactis* ssp. *cremoris* KW2]	280/ 332 (84)	YP_008569094.1	0.0	417
2	1363	2739	458	D-xylose-proton symporter [*Lactococcus lactis* ssp. *cremoris* KW2]	452/ 457 (99)	YP_008569093.1	0.0	457
3	3882	2797	361	GntR family transcriptional regulator [*Lactococcus lactis* ssp. *lactis* KF147]	356/ 361 (99)	YP_003354048.1	0.0	361
4	4465	4031	144	disaccharide permease [*Lactococcus lactis* ssp. *lactis* KF147]	138/ 144 (96)	YP_003354047	4e-89	421
5	4552	4812	86	transposase [*Lactococcus lactis* ssp. *lactis* IL1403]	86/ 86 (100)	NP_266244	4e-53	86
Contig_R150								
1	637	89	182	repB family protein [*Lactococcus lactis* ssp. *cremoris* HP]	169/ 182 (93)	EUN33389.1	1e-111	170
2	1796	630	388	*repA* [*Lactococcus lactis*]	387/ 388 (99)	CAA42983.1	0.0	388
3	2653	3333	226	transposase [*Listeria inocua* Clip11262]	222/ 226 (98)	NP_569202.1	2e-161	226
Contig_R145								
1	474	332	47	transposase [*Lactobacillus helveticus* R0052]	35/ 35 (100)	YP_006656720.1	8e-18	79
2	535	1358	94	replication initiation protein [*Leuconostoc lactis*]	89/ 91 (98)	NP_862729.1	5e-59	208
3	948	1358	136	GNAT family acetyltransferase [*Lactococcus lactis* ssp. *lactis* CV56]	135/ 136 (99)	YP_005869715.1	5e-90	136
4	1552	1406	48	hypothetical protein [*Lactococcus lactis* ssp. *cremoris* A76]	47/ 48 (98)	YP_005875627	7e-25	48
5	1784	1632	50	hypothetical protein llmg_0669 [*Lactococcus lactis* ssp. *cremoris* MG1363]	49/ 50 (98)	YP_001032008	1e-25	50
6	2137	1835	100	hypothetical protein llmg_0670 [*Lactococcus lactis* ssp. *cremoris* MG1363]	99/ 100 (99)	YP_001032009.1	9e-65	100
7	2310	2134	58	hypothetical protein llmg_0671 [*Lactococcus lactis* ssp. *cremoris* MG1363]	56/ 58 (97)	YP_001032010.1	3e-31	58
8	2713	2447	88	hypothetical protein LLNZ_03480 [*Lactococcus lactis* ssp. *cremoris* NZ9000]	87/ 88 (99)	YP_006356107	6e-57	88

9	3394	2873	173	hypothetical protein llmg_0673 [*Lactococcus lactis* ssp. *cremoris* MG1363]	173/ 173 (100)	YP_001032012	5e-124	205
10	3457	3717	86	transposase [*Lactococcus lactis* ssp. *lactis* IL1403]	85/ 86 (99)	NP_267352	3e-52	86
Contig_R112								
1	2217	64	717	glycoside hydrolase GH92 family [*Lactococcus lactis* ssp. *cremoris* KW2]	700/ 717 (98)	YP_008569091.1	0.0	717
2	3251	2214	345	transcriptional regulator GntR family [*Lactococcus lactis* ssp. *cremoris* KW2]	345/ 345 (100)	YP_008569090	0.0	345
3	3468	4208	246	glycoside hydrolase GH125 family [*Lactococcus lactis* ssp. *cremoris* KW2]	229/ 234 (98)	YP_008569089.1	6e-171	431
4	4574	4209	121	hypothetical protein [*Lactococcus lactis*]	121/ 121 (100)	WP_021211150.1	2e-81	121
Contig_R108								
1	91	447	118	IS*431* transposase, partial [*Listeria inocua*]	115/ 118 (97)	WP_003769216.1	2e-78	118
2	1408	509	299	*lacX* [*Lactococcus lactis* ssp. *cremoris*]	297/ 299 (99)	YP_006965846.1	0.0	299
3	3133	1700	477	phospho-beta-D-galactosidase [*Lactococcus lactis* ssp. *lactis*]	475/ 477 (99)	YP_004761515.1	0.0	477
4	4918	3212	568	PTS lactose transporter subunit IIBC [*Lactococcus lactis*]	567/ 568 (99)	WP_003131734.1	0.0	568
5	5239	4922	105	PTS lactose transporter subunit IIA [*Enterococcus* faecalis]	102/ 105 (97)	WP_002411590.1	7e-68	105
6	6247	5267	326	tagatose 1,6-diphosphate aldolase [*Lactococcus lactis* ssp. *cremoris* SK11]	326/ 326 (100)	YP_796532.1	0.0	326
7	7182	6250	310	tagatose 6-P kinase [*Lactococcus lactis* ssp. *lactis*]	310/ 310 (100)	YP_004761518	0.0	310
8	7708	7193	171	galactose-6-phosphate isomerase subunit LacB [*Lactococcus lactis* ssp. *cremoris* SK11]	171/ 171 (100)	YP_796534.1	4e-122	171
9	8150	7725	141	galactose 6-P isomerase subunit LacA [*Lactococcus lactis* ssp. *lactis*]	141/ 141 (100)	YP_004761520.1	3e-95	141
10	8547	8687	46	hypothetical protein [*Lactococcus lactis*]	45/ 46 (98)	WP_003131724.1	7e-22	46
11	8638	9405	255	lactose transport regulator [*Lactococcus lactis* ssp. *cremoris* SK11]	255/ 255 (100)	YP_796536.1	0.0	255
12	10470	9904	188	site-specific recombinase, DNA invertase Pin related protein [*Lactococcus lactis* ssp. *cremoris* SK11]	186/ 188 (99)	YP_796537.1	4e-130	193
Contig_R109								
1	2392	38	784	cell surface protein, partial [*Streptococcus thermophilus*]	429/ 803 (53)	WP_023909488.1	0.0	1016
2	2657	3022	121	hypothetical protein [*Lactococcus lactis*]	121/ 121 (100)	WP_017864521.1	2e-81	121
3	3025	3909	294	hypothetical protein [*Lactococcus lactis*]	292/ 294 (99)	WP_021214709.1	0.0	294
4	3896	5182	428	hypothetical protein [*Lactococcus lactis*]	427/ 428 (99)	WP_021214708.1	0.0	428
5	5873	5322	183	transposon Tn552 DNA-invertase *binR* [*Lactococcus lactis*]	181/ 183 (99)	WP_021214707.1	3e-130	183
6	6963	6379	194	dTDP-4-dehydrorhamnose 3,5-epimerase [*Lactococcus lactis* ssp. *lactis* CV56]	193/ 194 (99)	YP_005867336.1	3e-140	194
7	7526	7684	52	hypothetical protein [*Lactococcus lactis* ssp. *cremoris* A76]	50/ 52 (96)	YP_005875632.1	2e-28	52
8	7693	8193	166	ADP-ribose pyrophosphatase [*Lactococcus lactis* ssp. *cremoris* A76]	166/ 166 (100)	YP_005875631.1	3e-118	166
9	8437	9132	231	hypothetical protein [*Enterococcus durans*]	223/ 231 (97)	WP_016176879.1	2e-147	231
10	9330	9476	48	hypothetical protein [*Lactococcus lactis* ssp. *cremoris* A76]	45/ 48 (94)	YP_005875627.1	2e-23	48
11	9747	10247	166	transposase, partial [*Lactococcus lactis*]	106/ 107 (99)	WP_021211177.1	8e-70	108
Contig_R147								
1	1	465	154	transposase, partial [*Lactococcus lactis*]	153/ 154 (99)	WP_021165891.1	8e-108	157
2	1232	598	211	PIL6_6 [*Lactococcus lactis* ssp. *lactis*]	206/ 211 (98)	YP_006962314.1	7e-144	211

3	1375	1656	94	hypothetical protein L137335 [*Lactococcus lactis* ssp. *lactis* IL1403]	94/ 94 (100)	NP_266292.1	1e-59	96
4	1623	1946	108	unnamed protein product [*Lactococcus lactis*]	107/ 108 (99)	CAA55220.1	3e-68	299
Contig_R121								
1	691	23	222	putative replication protein [*Lactococcus lactis*]	214/ 222 (96)	AAF27320.1	1e-154	383
2	752	636	38	hypothetical protein [*Lactococcus lactis*]	34/ 38 (89)	WP_003132681.1	9e-16	38
3	1125	981	47	unknown protein				
4	1417	1283	44	unknown protein				
Contig_R106								
1	363	44	106	hypothetical protein [*Lactococcus lactis*]	75/ 76 (99)	WP_021214051.1	4e-44	90
2	507	1256	249	nicotinamide mononucleotide transporter [*Lactococcus lactis* ssp. *lactis* KF147]	249/ 249 (100)	YP_001174683.1	1e-172	249
3	1593	1751	53	sugar ABC transporter permease [*Streptococcus dysgalactiae* ssp. *equisimilis* 167]	39/ 53 (74)	YP_008630542.1	3e-17	296
Contig_R107								
1	1	423	140	transposase, partial [*Lactococcus lactis*]	140/ 140 (100)	WP_021165254.1	5e-99	142
2	1282	456	277	disaccharide permease [*Lactococcus lactis* ssp. *lactis* KF147]	257/ 270 (95)	YP_003354047.1	3e-155	421
3	2274	1304	323	alpha-N-arabinofuranosidase [*Lactococcus lactis* ssp. *lactis* KF147]	305/ 324 (94)	YP_003354046.1	0.0	323
4	3665	2298	455	arabinose-proton symporter [*Lactococcus lactis* ssp. *lactis* Dephy 1]	451/ 455 (99)	WP_023188764.1	0.0	455
5	3865	5448	527	L-Ribulokinase [*Lactococcus lactis* ssp. *lactis* KF147]	521/ 527 (99)	YP_003354044.1	0.0	527
6	5463	6164	233	L-ribulose-5-phosphate 4-epimerase [*Lactococcus lactis* ssp. *lactis* A12]	228/ 233 (98)	WP_021723025.1	7e-168	233
7	6166	7611	481	L-arabinose isomerase [*Lactococcus lactis* ssp. *lactis* A12]	473/ 481 (98)	WP_021723026.1	0.0	481
8	7760	10228	822	phosphoketolase [*Lactococcus lactis* ssp. *cremoris* KW2]	813/ 822 (99)	YP_008569092.1	0.0	822
Contig_R149								
1	1	191	63	hypothetical protein [*Lactococcus lactis*]	58/ 60 (97)	WP_021166198.1	7e-33	75
2	257	476	72	hypothetical protein [*Lactococcus lactis*]	72/ 72 (100)	WP_021166197.1	1e-43	93
3	1219	575	214	hypothetical protein [*Lactococcus lactis*]	213/ 214 (99)	WP_021166196.1	1e-147	218
Contig_R120								
1	403	93	101	transposase [*Lactococcus lactis* ssp. *lactis* IO-1]	100/ 101 (99)	YP_007507882.1	3e-64	101
2	1243	893	116	RepB family protein [*Lactococcus lactis* ssp. *cremoris* HP]	113/ 116 (97)	EUN33389.1	2e-72	170
Contig_R158								
1	617	120	165	acetyltransferase [*Lactococcus lactis* ssp. *lactis* biovar. *diacetylactis*]	164/ 165 (99)	WP_023349305.1	4e-117	165
2	1473	778	231	peptidase E [*Lactococcus lactis* ssp. *cremoris* SK11]	231/ 231 (100)	YP_809513.1	7e-166	231
3	1949	1473	158	phosphoribosylanthranilate isomerase [*Lactococcus lactis*]	158/ 158 (100)	WP_023349307.1	8e-106	158
4	2346	1963	127	XRE family transcriptional regulator [*Lactococcus lactis*]	117/ 118 (99)	WP_023349308.1	3e-74	122
Contig_R111								
1	1	426	141	transposase, partial [*Lactococcus lactis*]	141/ 141 (100)	WP_021165254.1	9e-100	142
2	1235	459	258	abortive phage resistance protein [*Streptococcus parasanguinis*]	136/ 242 (56)	WP_003014061.1	4e-87	295
3	2854	1829	341	RecAlp [*Lactococcus lactis*]	335/ 341 (98)	YP_001966451.1	0.0	341
4	3261	2869	130	hypothetical protein CVCAS_pA0014 [*Lactococcus lactis* ssp. *lactis* CV56]	125/ 130 (96)	YP_005867340.1	1e-86	131

5	4724	3258	488	impB/ MucB/ SamB family protein [*Lactococcus lactis* ssp. *lactis* CV56]	484/ 488 (98)	YP_005867341.1	0.0	488
6	5503	4865	212	hypothetical protein LLDT4_12715 [*Lactococcus lactis* ssp. *lactis* biovar. *diacetylactis* TIFN4]	208/ 212 (98)	WP_021214678.1	2e-148	212
7	5862	6752	296	transposase [*Lactococcus lactis* ssp. *cremoris* SK11]	293/ 296 (99)	YP_808666.1	0.0	296
Contig_R151								
1	30	443	137	integrase, partial [*Lactococcus lactis*]	137/ 137 (100)	WP_021166222.1	9e-95	137
2	1475	624	283	nisin leader peptide-processing serine protease nisP [*Lactococcus lactis* ssp. *lactis* CV56]	280/ 282 (99)	YP_005867988.1	0.0	682
Contig_R152								
1	765	88	225	putative replication initiator protein [*Lactococcus garvieae*]	224/ 224 (100)	YP_005351612.1	8e-163	383
2	2122	1799	108	putative cytochrome B [*Lactococcus lactis* ssp. *cremoris* UC509.9]	93/ 108 (86)	YP_006998540.1	1e-58	107
Contig_R146								
1	431	4	142	putative resolvase [*Lactococcus lactis* ssp. *lactis* biovar. *diacetylactis*]	142/ 142 (100)	YP_001098358.1	3e-95	199
2	890	435	151	hypothetical protein [*Lactococcus lactis*]	148/ 151 (98)	WP_003132677.1	7e-101	151
3	1217	921	98	hypothetical protein pIL4_48 [*Lactococcus lactis* ssp. *lactis*]	81/ 98 (83)	YP_004761488.1	5e-50	98
Contig_R118								
1	664	4	220	transposase [*Lactococcus lactis* ssp. *lactis* IO-1]	197/ 211 (93)	YP_007508154.1	6e-135	370
2	757	1208	150	site-specific recombinase, DNA invertase Pin related protein [*Lactococcus lactis*]	128/ 129 (99)	WP_023189736.1	2e-84	193
Contig_R122								
1	15	377	120	replication protein RepB [*Lactococcus lactis* ssp. *lactis*]	120/ 120 (100)	YP_004761555.1	2e-76	383
2	374	997	207	putative replication associated protein [*Lactococcus garvieae*]	166/ 201 (83)	YP_005351613.1	4e-103	207
Contig_R126								
1	509	60	149	putative transporter protein [*Lactococcus lactis* ssp. *cremoris* UC509.9]	146/ 148 (99)	YP_006998507.1	1e-96	150
2	876	712	54	hypothetical protein [*Lactococcus lactis*]	54/ 54 (100)	WP_021166195.1	3e-29	54
3	1367	1242	41	hypothetical protein [*Lactococcus lactis*]	40/ 41 (98)	WP_021166194.1	2e-18	53
4	1369	1539	57	hypothetical protein [*Lactococcus lactis*]	56/ 57 (98)	WP_021214051.1	2e-30	90
Contig_R127								
1	5	433	142	transposase, partial [*Lactococcus lactis*]	133/ 140 (95)	WP_021165254.1	5e-95	142
2	723	1268	181	hypothetical protein LLKF_0283 [*Lactococcus lactis* ssp. *lactis* KF147]	51/ 143 (36)	YP_003352755.1	6e-09	143
3	1366	1704	112	hypothetical protein [*Streptococcus thoraltensis*]	38/ 100 (38)	WP_018380804.1	9e-13	172
Contig_R155								
1	92	1259	388	cell surface protein precursor, partial [*Lactobacillus coryniformis*]	371/ 400 (93)	WP_010014574.1	0.0	992
2	1356	1709	117	MarR family transcriptional regulator [*Lactobacillus coryniformis*]	115/ 117 (98)	WP_010014573.1	4e-75	145
3	1892	2398	168	acyltransferase [*Tetragenococcus halophilus* HO]	129/ 153 (84)	WP_021730194.1	3e-84	614
4	2558	3115	185	DNA invertase Pin-like site-specific recombinase [*Lactococcus lactis* ssp. *lactis* KF147]	177/ 184 (96)	YP_003353150.1	3e-123	184
5	4360	3173	395	*abiGii* [*Lactococcus lactis* ssp. *cremoris*]	236/ 296 (80)	AAB38313.1	1e-166	397

6	4536	4796	86	transposase [*Lactococcus lactis* ssp. *lactis* IL1403]	86/ 86 (100)	NP_266244.1	2e-43	86
7	4808	4879	24	truncated transposase B [*Lactococcus lactis*]	24/ 24 (100)	YP_005877382.1	5e-08	107
Contig_R142								
1	97	684	195	hypothetical protein [*Lactobacillus fermentum*]	193/ 195 (99)	WP_021349473.1	2e-137	404
2	1316	747	189	transposase [*Pediococcus claussenii* ATCC BAA-344]	139/ 162 (86)	YP_005004471.1	1e-94	205
3	3445	1424	673	nickase [*Lactococcus lactis* ssp. *cremoris* TIFN7]	662/ 673 (98)	WP_021211359.1	0.0	673
Contig_R154								
1	47	538	163	transposase IS1216 [*Lactococcus lactis*]	158/ 163 (97)	WP_021166207.1	1e-111	226
2	610	1197	195	hypothetical protein [*Lactococcus garvieae*]	186/ 195 (95)	WP_017370784.1	2e-136	195
3	2039	1828	70	unknown				
Contig_R141								
1	9	323	104	hypothetical protein LLT6_13815, partial [*Lactococcus lactis* ssp. *cremoris* TIFN6]	104/ 104 (100)	EQC55638.1	2e-68	128
2	345	572	75	hypothetical protein [*Lactococcus garvieae*]	36/ 70 (51)	WP_019293275.1	1e-13	133
3	1725	1264	153	hypothetical protein [*Mycoplasma yeatsii*]	51/ 150 (34)	WP_004426892.1	1e-20	180
Contig_R133								
1	53	355	100	integrase/recombinase plasmid associated, putative [*Lactococcus lactis*]	96/ 98 (98)	WP_023189051.1	5e-64	154
2	559	882	107	cytochrome B [*Lactococcus lactis* ssp. *lactis* CV56]	102/ 107 (95)	YP_005867402.1	9e-68	107
3	854	1013	52	hypothetical protein [*Lactobacillus helveticus*]	46/ 52 (88)	WP_003627881.1	5e-25	52
Contig_R130								
1	808	158	216	replication-associated protein [*Lactococcus lactis* ssp. *cremoris*]	175/ 213 (82)	YP_006964733.1	3e-110	239
2	1164	805	119	replication protein RepB [*Lactococcus lactis* ssp. *cremoris* HP]	104/ 118 (88)	EUN33281.1	2e-64	388
Contig_R129								
1	200	763	187	methionine ABC transporter substrate-binding protein [*Lactococcus lactis* ssp. *cremoris* KW2]	176/ 177 (99)	YP_008567989.1	6e-120	286
Contig_R131								
1	729	1	243	replication protein [*Lactococcus lactis* ssp. *lactis* biovar. *diacetylactis*]	229/ 242 (95)	NP_690629.1	4e-164	423
Contig_R163								
1	2	181	59	universal stress protein UspA-like protein [*Lactococcus lactis* ssp. *lactis*]	59/ 59 (100)	YP_004761547.1	5e-34	145
2	752	294	152	transposase [*Enterococcus faecium* EnGen0015]	150/ 152 (99)	WP_002329974.1	8e-106	152
3	1637	927	236	IS*946* transposase [*Lactococcus lactis* ssp. *cremoris*]	236/ 236 (100)	WP_023349245.1	5e-173	236
4	1668	2573	301	*ltrC* [*Lactococcus lactis*]	66/ 144 (46)	NP_047307.1	5e-28	355
5	2772	3557	261	cobyric acid synthase *cobQ* [*Bacillus cereus*]	173/ 260 (67)	WP_000692380.1	5e-118	266
6	3591	3797	68	hypothetical protein [*Bacillus cereus*]	30/ 54 (56)	WP_002016733.1	2e-09	101
7	4548	4103	149	transposase [*Lactococcus lactis* ssp. *lactis* IO-1]	104/ 124 (84)	YP_007508154.1	1e-61	370
8	5055	6629	524	hypothetical protein [*Enterococcus caccae*]	252/ 524 (48)	WP_010770548.1	1e-164	525

9	8438	6669	589	papain cysteine protease family protein [*Carnobacterium maltaromaticum* LMA28]	141/ 524 (27)	YP_006992557.2	5e-30	850
10	9560	8526	344	hypothetical protein [*Streptococcus* sp. HSISS3]	209/ 336 (62)	WP_021151537.1	2e-159	340
11	10065	9553	170	hypothetical protein [*Streptococcus thoraltensis*]	69/ 175 (39)	WP_018380804.1	2e-32	172
12	10705	10163	180	putative DNA-binding protein, XRE family [*Lactococcus lactis* ssp. *lactis* IO-1]	51/ 143 (36)	YP_007507912.1	4e-09	143
13	11846	10998	282	transposase [*Lactococcus lactis*]	252/ 272 (93)	WP_017865390.1	0.0	284
14	11810	12274	154	transposase, partial [*Lactococcus lactis*]	146/ 152 (96)	WP_021165590.1	1e-102	154
15	12296	12523	75	hypothetical protein [*Lactococcus garvieae*]	36/ 70 (51)	WP_019293275.1	1e-13	133
16	13775	13215	186	antibiotic ABC transporter [*Oribacterium sinus*]	62/ 179 (35)	WP_007158144.1	3e-26	185
17	14156	14866	236	transposase IS*1216* [*Lactococcus lactis*]	232/ 236 (98)	WP_023349245.1	5e-170	236
18	15845	15672	57	hypothetical protein [*Lactococcus lactis*]	53/ 57 (93)	WP_021166204.1	3e-28	57
19	16073	15873	66	Putative uncharacterized protein [*Lactococcus lactis*]	31/ 64 (48)	WP_021722479.1	1e-04	72
20	17624	16203	473	lactococcin A secretion protein LcnD [*Lactococcus lactis* ssp. *cremoris* UC509.9]	464/ 472 (98)	YP_006998503.1	0.0	474
21	19786	17639	715	lactococcin A ABC transporter ATP-binding protein/ permease [*Lactococcus lactis* ssp. *lactis* IL1403]	674/ 715 (94)	NP_266235.1	0.0	715
22	20944	20045	299	transposase [*Lactococcus lactis* ssp. *cremoris* A76]	298/ 299 (99)	YP_005876626.1	0.0	300
23	21156	20896	86	transposase, partial [*Lactococcus lactis*]	86/ 86 (100)	WP_021211123.1	2e-53	115
24	21339	21193	48	GNAT family acetyltransferase [*Lactococcus lactis* ssp. *lactis* CV56]	45/ 46 (98)		4e-20	136
25	21753	21439	104	replication initiation protein [*Lactococcus lactis* ssp. *cremoris* A76]	100/ 101 (99)	YP_005875625.1	2e-68	105
26	22209	23165	318	nisin resistance protein [*Lactococcus lactis*]	316/ 318 (99)	ACE80201.1	0.0	318
27	23902	23234	222	CAAX amino terminal protease [*Lactococcus lactis*]	221/ 222 (99)	WP_021214621.1	3e-148	222
28	24563	24003	186	hypothetical protein [*Lactococcus lactis*]	179/ 183 (98)	WP_021214620.1	3e-121	185
29	25172	24801	123	HTH domain-containing protein [*Lactococcus lactis* ssp. *cremoris* HP]	98/ 116 (84)	EUN33279.1	1e-61	171
30	25180	25305	41	hypothetical protein [*Lactococcus lactis*]	40/ 41 (98)	WP_021166194.1	2e-18	53
31	25671	25835	54	hypothetical protein [*Lactococcus lactis*]	54/ 54 (100)	WP_021166195.1	3e-29	54
32	26038	26490	150	putative transporter protein [*Lactococcus lactis* ssp. *cremoris* UC509.9]	149/ 150 (99)	YP_006998507.1	9e-99	150
33	26529	27185	218	hypothetical protein [*Lactococcus lactis*]	216/ 218 (99)	WP_021166196.1	5e-150	218
34	27503	27285	72	hypothetical protein [*Lactococcus lactis*]	72/ 72 (100)	WP_021166197.1	1e-43	93
35	27795	27568	75	hypothetical protein [*Lactococcus lactis*]	73/ 75 (97)	WP_021166198.1	7e-45	75
36	28887	28300	195	hypothetical protein [*Lactococcus garvieae*]	186/ 195 (95)	WP_017370784.1	2e-136	195
37	29639	28959	226	transposase IS*1216* [*Lactococcus lactis*]	222/ 226 (98)	WP_021166207.1	7e-162	226
38	29754	30224	156	restriction endonuclease subunit R [*Enterococcus faecium*]	154/ 156 (99)	WP_002307466.1	1e-108	156
39	30404	31249	281	aldo/keto reductase [*Enterococcus faecium*]	280/ 281 (99)	WP_002314393.1	0.0	281
40	31451	31600	49	hypothetical protein [*Lactococcus lactis*]	48/ 49 (98)	WP_023349359.1	6e-26	49
41	32504	32133	123	hypothetical protein CAR_c24020 [*Carnobacterium* sp. 17-4]	61/ 121 (50)	YP_004376073.1	1e-23	123
42	32994	32650	114	hypothetical protein LSEI_0032 [*Lactobacillus casei* ATCC 334]	61/ 107 (57)	YP_805340.1	2e-34	119
43	34105	33647	152	integrase [*Lactococcus lactis*]	152/ 152 (100)	WP_021215110.1	1e-106	152

44	34134	34568	144	transposase, partial [*Lactococcus lactis*]	143/ 144 (99)	WP_021165382.1	8e-99	145
45	34678	35934	418	oxalate/ formate antiporter [*Lactococcus lactis* ssp. *cremoris* MG1363]	415/ 418 (99)	YP_001032006.1	0.0	418
46	36249	36052	65	putative transposase InsK for insertion sequence element IS150 [*Lactobacillus brevis* KB290]	30/ 61 (49)	YP_007655304.1	1e-09	298
47	36256	36564	102	replication initiation protein [*Lactococcus lactis* ssp. *cremoris* A76]	101/ 105 (96)	YP_005875625.1	8e-67	105
48	36638	36937	99	GNAT family acetyltransferase [*Lactococcus lactis* ssp. *lactis* bv. *diacetylactis*]	97/ 98 (99)	YP_004770029.1	6e-60	136
49	37832	36942	296	transposase [*Lactococcus lactis* ssp. *cremoris* A76]	291/ 296 (98)	YP_005876203.1	0.0	296
Contig_R164								
1	1	399	132	integrase, partial [*Lactococcus lactis*	132/ 132 (100)	WP_021166222.1	7e-91	137
2	441	1316	291	hypothetical protein Y016_10255 [*Streptococcus thermophilus* TH985]	263/ 301 (87)	EWM56040.1	7e-66	678
3	2039	1329	236	transposase IS*1216* [*Lactococcus lactis*]	236/ 236 (100)	WP_023349245.1	5e-173	236
4	2103	2933	276	hypothetical protein L37749 [*Lactococcus lactis* ssp. *lactis* IL1403]	163/ 235 (69)	NP_266805.1	2e-102	389
5	3101	2892	39	Putative uncharacterized protein, phage protein [*Lactococcus lactis*]	32/ 39 (82)	WP_021722049.1	7e-11	39
6	3983	3123	286	Putative uncharacterized protein [*Lactococcus lactis* ssp. *lactis* A12]	240/ 286 (84)	CDG03865.1	6e-161	279
7	4266	4126	46	transposase IS*431*, partial [*Lactococcus lactis*]	14/ 15 (93)	WP_021214232.1	5.2	72
8	5492	4272	406	cell wall surface anchor protein [*Lactococcus lactis*]	338/ 406 (83)	WP_014734938.1	0.0	405
9	6550	5516	344	peptidase C39 family [*Lactococcus lactis* ssp. *cremoris* KW2]	277/ 344 (81)	YP_008568127.1	2e-164	344
10	6750	8702	650	hypothetical protein [*Lactococcus lactis* ssp. *cremoris* A76]	192/ 467 (41)	WP_021164893.1	2e-88	1077
11	9943	9299	214	putative replication initiator protein [*Lactococcus garvieae*]	382/ 383 (99)	YP_005351612.1	0.0	383
12	11094	9943	383	putative replication associated protein [*Lactococcus garvieae*]	166/ 214 (78)	YP_005351613.1	2e-104	207
13	12204	12494	97	hypothetical protein, partial [*Lactococcus lactis*]	55/ 67 (82)	WP_021164893.1	9e-23	110
14	12788	13234	148	IS*LL6* transposase/hypothetical protein fusion [*Lactococcus lactis*]	119/ 136 (88)	YP_005877591.1	2e-73	409
15	13252	14091	279	transposase [*Lactococcus lactis* ssp. *cremoris* SK11]	278/ 279 (99)	YP_808650.1	0.0	279
Contig_R165								
1	249	1	83	nisin resistance protein [*Lactococcus lactis*]	82/ 83 (99)	ACE80201.1	7e-51	318
2	554	246	102	enolase [*Mycoplasma mycoides*]	18/ 59 (31)	WP_017697988.1	1.3	451
3	777	1775	332	*abiI* [*Lactococcus lactis*]	332/ 332 (100)	AAB66882.1	0.0	332
4	2091	3542	483	SOS response UmuC protein [*Lactococcus lactis* ssp. *lactis* IO-1]	448/ 483 (93)	YP_007508452.1	0.0	483
5	3539	3934	131	hypothetical protein CVCAS_pA0014 [*Lactococcus lactis* ssp. *lactis* CV56]	130/ 131 (99)	YP_005867340.1	1e-89	131
6	3946	4971	341	recombinase A [*Lactococcus lactis* ssp. *lactis* CV56]	341/ 341 (100)	YP_005867339.1	0.0	341
7	5302	5135	55	*abiEii* [*Lactococcus lactis*]	47/ 50 (94)	YP_001966452.1	9e-23	298
8	5617	5408	69	putative orotidine-5-phosphate decarboxylase [*Lactococcus lactis*]	69/ 69 (100)	YP_002332310.1	1e-38	317
Contig_R166								
1	336	1	112	transposase IS*1216* [*Lactococcus lactis*]	112/ 112 (100)	WP_021216434.1	1e-75	226
2	522	818	98	hypothetical protein pNZ4000_15 [*Lactococcus lactis* ssp. *cremoris*]	98/ 98 (100)	NP_053020.1	8e-64	98
3	847	1311	154	hypothetical protein pNZ4000_14 [*Lactococcus lactis* ssp. *cremoris*]	154/ 154 (100)	NP_053019.1	8e-110	154
4	1515	1847	110	cytochrome B [*Lactococcus lactis* ssp. *lactis* CV56]	102/ 107 (95)	YP_005867402.1	1e-67	107

5	2854	2375	159	hypothetical protein LLKF_0688 [*Lactococcus lactis* ssp. *lactis* KF147]	154/ 159 (97)	YP_003353152.1	2e-106	159
6	3122	4201	359	Abi-like protein [*Veillonella* sp. oral taxon 780]	167/ 368 (45)	WP_009352868.1	2e-97	371
7	4904	4446	152	transposase [*Enterococcus faecium*]	150/ 152 (99)	WP_010728590.1	8e-106	152
8	5088	4894	65	hypothetical protein [*Lactococcus lactis*]	40/ 42 (95)	WP_017865304.1	9e-20	57
9	5842	7020	392	replication protein [*Lactococcus lactis* ssp. *lactis*]	340/ 391 (87)	YP_004761466.1	0.0	394
10	7017	7733	238	replication-associated protein [*Lactococcus lactis* ssp. *cremoris*]	184/ 243 (76)	YP_006964733.1	3e-115	239
11	7742	8242	166	hypothetical protein pMRC01_043 [*Lactococcus lactis*]	151/ 151 (100)	NP_047327.1	2e-107	286
12	8247	8561	105	hypothetical protein pMRC01_042 [*Lactococcus lactis*]	104/ 105 (99)	NP_047326.1	3e-63	107
13	9874	9188	228	IS*1216* transposase [*Enterococcus faecalis*]	228/ 228 (100)	YP_003896017.1	2e-166	228
14	10056	9922	44	hypothetical protein [*Lactococcus lactis*]	38/ 38 (100)	CAA73217.1	5e-19	221
15	10299	10129	56	hypothetical protein LLCHP_2335 [*Lactococcus lactis* ssp. *cremoris* HP]	56/ 56 (100)	WP_021212881.1	5e-31	56
Contig_R157								
1	270	50	73	hypothetical protein [*Streptococcus* sp. oral taxon 056]	32/ 68 (47)	WP_009754589.1	2e-09	79
2	1040	417	207	hypothetical protein, partial [*Streptococcus mutans*]	75/ 223 (34)	WP_019319068.1	4e-16	212
3	1240	1596	118	IS*LL6* transposase/ hypothetical protein [*Lactococcus lactis* ssp. *lactis*]	117/ 117 (100)	YP_004761559.1	5e-77	120
Contig_R156								
1	2	307	101	hypothetical protein, partial [*Lactococcus lactis*]	101/ 101 (100)	WP_021216519.1	9e-69	101
2	766	368	132	transposase IS1077, partial [*Lactococcus lactis*]	131/ 132 (99)	WP_021164987.1	2e-92	132
3	1647	777	289	MucBP domain-containing cell surface protein [*Lactococcus lactis* ssp. *cremoris* KW2]	273/ 286 (95)	YP_008569303.1	0.0	523
Contig_R160								
1	194	3	64	transposase, partial [*Lactococcus lactis*]	64/ 64 (100)	WP_021211111.1	3e-38	116
2	419	231	62	hypothetical protein kw2_1771 [*Lactococcus lactis* ssp. *cremoris* KW2]	39/ 39 (100)	YP_008569411.1	6e-19	279
3	735	427	102	hypothetical protein kw2_1772 [*Lactococcus lactis* ssp. *cremoris* KW2]	100/ 102 (98)	YP_008569412.1	1e-66	102
4	1132	818	104	EsaC-like protein [*Lactococcus lactis* ssp. *cremoris* KW2]	104/ 104 (100)	YP_008569413.1	7e-68	104
5	1447	1260	63	putative acetyltransferase [*Lactococcus lactis* ssp. *cremoris* NZ9000]	57/ 62 (92)	YP_006355503.1	3e-30	150
Contig_R138								
1	398	96	100	transposase [*Lactococcus lactis* ssp. *lactis* CV56]	88/ 92 (96)	YP_005867944.1	2e-54	96
2	1037	486	183	general stress protein, Gls24 family [*Lactococcus lactis* ssp. *lactis* A12]	182/ 183 (99)	YP_007507909.1	1e-121	183
3	1246	1058	62	hypothetical protein llmg_1259 [*Lactococcus lactis* ssp. *cremoris* MG1363]	62/ 62 (100)	YP_001032567.1	1e-33	62
4	1816	1256	186	hypothetical protein LLCRE1631_01277 [*Lactococcus lactis* ssp. *lactis* CNCM I-1631]	175/ 186 (94)	WP_003131013.1	2e-111	186
5	2076	1846	76	hypothetical protein llmg_1257 [*Lactococcus lactis* ssp. *cremoris* MG1363]	75/ 76 (99)	YP_001032565.1	4e-39	79
6	2869	2588	90	cold shock protein D [*Lactococcus lactis* ssp. *lactis*]	84/ 91 (92)	YP_006962316.1	1e-51	91
Contig_R167								
1	3	145	47	putative nicotinamide mononucleotide transporter [*Lactococcus lactis*]	47/ 47 (100)	YP_001174683.1	2e-21	249
2	1009	1449	146	hypothetical protein [*Lactococcus garvieae*]	143/ 146 (98)	YP_005352404.1	1e-96	146

3	1530	1845	104	hypothetical protein [*Lactococcus garvieae*]	68/ 95 (72)	YP_005352405.1	9e-33	104
4	1892	2794	300	ABC transporter like protein [*Lactococcus lactis* ssp. *cremoris* A76]	298/ 300 (99)	YP_005875640.1	0.0	300
5	2787	3524	245	ABC transporter permease [*Lactococcus lactis* ssp. *cremoris* A76]	245/ 245 (100)	YP_005875639.1	2e-171	245
6	3645	3941	98	DNA-directed DNA polymerase [*Lactococcus lactis* ssp. *lactis* biovar. *diacetylactis* str. LD61]	95/ 98 (97)	YP_004761488.1	1e-61	98
7	3972	4427	151	hypothetical protein LLCRE1631_02432 [*Lactococcus lactis* ssp. *lactis* CNCM I-1631]	148/ 151 (98)	WP_003132677.1	8e-101	151
8	4431	5030	199	putative resolvase [*Lactococcus lactis* ssp. *lactis* biovar. *diacetylactis*]	199/ 199 (100)	YP_001098358.1	7e-140	199
9	5560	5676	38	hypothetical protein LLCRE1631_02433 [*Lactococcus lactis* ssp. *lactis* CNCM I-1631]	34/ 38 (89)	WP_003132681.1	9e-16	38
10	5621	6772	383	RepB protein [*Lactococcus lactis*]	371/ 383 (97)	WP_023349383.1	0.0	383
11	6773	6937	54	hypothetical protein LLDT2_03950 [*Lactococcus lactis* ssp. *lactis* biovar. *diacetylactis* TIFN2]	36/ 54 (67)	WP_021216441.1	9e-10	53
12	7329	6946	127	hypothetical protein llh_13965 [*Lactococcus lactis* ssp. *cremoris* A76]	125/ 127 (98)	YP_796552.1	2e-78	127
13	7616	7335	93	hypothetical protein LLCHP_2200 [*Lactococcus lactis* ssp. *cremoris* HP]	93/ 93 (100)	EUN33409.1	2e-56	93
14	8729	8312	138	cytidine deaminase [*Staphylococcus lentus*]	131/ 138 (95)	WP_017000533.1	6e-93	138
15	8851	9273	141	IS6 family transposase [*Lactococcus lactis* ssp. *cremoris* UC509.9]	141/ 141 (100)	YP_006998524.1	6e-99	236
Contig_R168								
1	2343	295	682	*nisP* [*Lactococcus lactis* ssp. *lactis*]	676/ 682 (99)	ADJ56357.1	0.0	682
2	3082	2345	245	nisin immunity protein nisI [*Lactococcus lactis* ssp. *lactis* CV56]	241/ 245 (98)	YP_005867987.1	1e-173	245
3	4104	3418	228	transposase [*Lactococcus garvieae*]	226/ 228 (99)	WP_003134290.1	2e-163	228
4	4343	4645	100	hypothetical protein uc509_p8064 [*Lactococcus lactis* ssp. *cremoris* UC509.9]	100/ 100 (100)	YP_006998461.1	6e-66	100
5	4989	4789	66	hypothetical protein LLT5_13880 [*Lactococcus lactis* ssp. *cremoris* TIFN5]	66/ 66 (100)	WP_021166208.1	4e-40	66
6	5235	5116	39	hypothetical protein [*Lactococcus lactis*]	39/ 39 (100)	WP_021166209.1	5e-29	39
7	5827	5432	131	transposase IS*904*, partial [*Lactococcus lactis*]	131/ 131 (100)	WP_021216565.1	4e-94	134
Contig_R169								
1	1	1398	466	mucus binding protein [*Leuconostoc citreum* KM20]	426/ 457 (93)	YP_001727229.1	0.0	1977
Contig_R171								
1	339	7	111	oxalate:formate antiporter, MFS family [*Lactococcus lactis* ssp. *cremoris* A76]	111/ 111 (100)	YP_005877273.1	5e-70	418
2	1159	479	226	transposase element IS*1297* [*Streptococcus thermophilus* MN-ZLW-002]	221/ 226 (98)	YP_006339954.1	2e-160	226
Contig_R173								
1	193	3	63	putative replication initiator protein [*Lactococcus garvieae*]	63/ 63 (100)	YP_005351620.1	3e-34	386
2	1335	1198	45	hypothetical protein [*Lactococcus lactis*]	39/ 40 (98)	WP_017865385.1	1e-17	62
3	2004	1774	76	*repC* [*Lactococcus lactis* ssp. *cremoris*]	76/ 76 (100)	NP_858116.1	7e-49	76
4	2639	2013	208	hypothetical protein pND324p2 [*Lactococcus lactis*]	208/ 208 (100)	YP_783821.1	2e-141	208
5	3790	2639	383	replication protein [*Lactococcus lactis*]	382/ 383 (99)	YP_783820.1	0.0	383
Contig_R174								

1	2	193	63	Gluconate permease, Bsu4004 homolog [*Lactococcus lactis*]	62/ 63 (98)	WP_023190130.1	1e-35	234
Contig_R175								
1	499	194	101	transposase [*Lactococcus lactis* ssp. *lactis* IO-1]	100/ 101 (99)	YP_007507882.1	4e-64	101
2	1537	989	182	RepB family protein [*Lactococcus lactis* ssp. *cremoris* HP]	169/ 182 (93)	EUN33389.1	1e-111	170
3	2477	1530	315	*repA* [*Lactococcus lactis*]	314/ 315 (99)	CAA42983.1	0.0	338
Contig_R176								
1	212	388	58	hypothetical protein llmg_0671 [*Lactococcus lactis* ssp. *cremoris* MG1363]	57/ 58 (98)	YP_001032010.1	8e-32	58
2	385	687	100	hypothetical protein llmg_0670 [*Lactococcus lactis* ssp. *cremoris* MG1363]	99/ 100 (99)	YP_001032009.1	9e-65	100
3	738	890	50	hypothetical protein LLT6_04635 [*Lactococcus lactis* ssp. *cremoris* TIFN6]	49/ 50 (98)	WP_021213338.1	1e-25	50
4	1577	1167	136	GNAT family acetyltransferase [*Lactococcus lactis* ssp. *lactis* CV56]	135/ 136 (99)	YP_005869715.1	5e-90	136
5	2164	2568	134	hypothetical protein LLNZ_03495 [*Lactococcus lactis* ssp. *cremoris* NZ9000]	132/ 134 (99)	YP_006356110.1	1e-89	134
6	3084	2728	118	acetyltransferase [*Lactococcus lactis* ssp. *lactis* biovar. *diacetylactis* str. LD61]	117/ 118 (99)	WP_023349305.1	3e-81	165
Contig_R177								
1	3	1188	395	transposase [*Lactococcus lactis* ssp. *cremoris* MG1363]	321/ 358 (90)	AAB01067.1	0.0	385
2	1318	1188	66	hypothetical protein, partial [*Lactococcus lactis*]	66/ 66 (100)	WP_021216559.1	2e-41	94
Contig_R178								
1	2	496	164	hypothetical protein LLT6_14725 [*Lactococcus lactis* ssp. *cremoris* TIFN6]	164/ 164 (100)	EQC54512.1	6e-114	431
2	604	1173	189	transposase [*Pediococcus claussenii* ATCC BAA-344]	139/ 162 (86)	YP_005004471.1	1e-94	205
3	2297	1314	327	hypothetical protein [*Lactobacillus fermentum*]	324/ 327 (99)	WP_021349473.1	0.0	404
Contig_R179								
1	138	326	62	hypothetical protein llmg_0689 [*Lactococcus lactis* ssp. *cremoris* MG1363]	33/ 39 (85)	YP_001032028.1	2e-15	42
2	855	1037	61	hypothetical protein LLT7_14225 [*Lactococcus lactis* ssp. *cremoris* TIFN7]	28/ 29 (97)	WP_021211588.1	3e-10	84
3	1157	1312	51	hypothetical protein LLT6_14610 [*Lactococcus lactis* ssp. *cremoris* TIFN6]	51/ 51 (100)	WP_021213379.1	7e-27	51
4	1369	1962	197	*repB* [*Lactobacillus acidophilus*]	129/ 186 (69)	NP_862699.1	7e-96	192
Contig_R180								
1	630	1	210	hypothetical protein LLNZ_06175 [*Lactococcus lactis* ssp. *cremoris* NZ9000]	208/ 210 (99)	YP_001032512.1	3e-144	275
2	780	649	43	hypothetical protein LLNZ_06170 [*Lactococcus lactis* ssp. *cremoris* NZ9000]	43/ 43 (100)	YP_006356621.1	4e-23	43
3	1237	749	162	hypothetical protein llmg_1198 [*Lactococcus lactis* ssp. *cremoris* MG1363]	162/ 162 (100)	YP_001032511.1	9e-115	162

Appendix 5-2

ORF table of non-alignment contigs to reference genomes for DPC6853 (Corn)

ORF	Start	Stop	Size (aa)	Predicted function	Identity	Accession no.	E-value	Size
contig_C30								
1	283	5	92	hypothetical protein P620_09490 [*Lactococcus lactis* ssp. *lactis* KLDS 4.0325]	80/ 81 (99)	YP_008703326.1	1e-46	331
2	498	971	157	hypothetical protein N036_57170 [*Enterococcus gallinarum* EGD-AAK12]	156/ 157 (99)	ERE52201.1	5e-103	157
3	988	1194	68	Transcriptional regulator, Cro/CI family [*Lactococcus lactis*]	68/ 68 (100)	WP_023189708.1	5e-41	68
4	1345	1542	65	hypothetical protein N036_62475 [*Enterococcus gallinarum* EGD-AAK12]	64/ 64 (100)	ERE50927.1	9e-37	80
5	1926	3314	462	hypothetical protein N036_56625, partial [*Enterococcus gallinarum* EGD-AAK12]	188/ 191 (98)	ERE52372.1	8e-119	191
6	3462	4115	217	CAAX amino terminal protease family protein [*Lactococcus lactis*]	206/ 217 (95)	WP_023189704.1	2e-142	217
contig_C99								
1	595	1230	141	hypothetical protein N036_63560, partial [*Enterococcus gallinarum* EGD-AAK12]	140/ 141 (99)	ERE50473.1	1e-91	141
contig_C91								
1	9	861	291	peptidase M13, partial [*Lactococcus lactis*]	279/ 291 (96)	WP_023349133.1	0.0	293
contig_C96								
1	60	266	68	family 2 glycosyltransferase [*Lactococcus lactis* ssp. *lactis* KF147]	68/ 68 (100)	YP_003352612.1	1e-37	295
2	267	1343	358	group 1 glycosyltransferase [*Lactococcus lactis* ssp. *lactis* KF147]	350/ 358 (98)	YP_003352613.1	0.0	358
3	1379	2326	315	glycosyltransferase [*Lactococcus lactis* ssp. *lactis* KF147]	310/ 315 (98)	YP_003352614.1	0.0	315
4	2319	3744	475	polysaccharide biosynthesis export protein [*Lactococcus lactis* ssp. *lactis* KF147]	447/ 474 (99)	YP_003352615.1	0.0	474
contig_C97								
1	10	1947	645	DNA-directed RNA polymerase subunit beta [*Lactococcus lactis* ssp. *lactis* IL1403]	641/ 642 (99)	WP_017864927.1	0.0	1196
contig_C101								
1	142	399	85	transposase for insertion sequence-like element IS431mec [*Enterococcus faecalis*]	85/ 85 (100)	WP_010730026.1	2e-52	85
2	483	686	67	hypothetical protein [*Lactococcus garvieae*]	67/ 67 (100)	WP_017370349.1	3e-38	67
3	1518	3458	646	tetracycline resistance protein [*Lactococcus garvieae*]	646/ 646 (100)	YP_001798653.1	0.0	646
contig_C24								
1	983	864	39	hypothetical protein [*Lactococcus lactis*]	38/ 39 (97)	WP_003131874.1	4e-18	39
2	1817	1056	253	transposase of IS*904B* [*Lactococcus lactis* ssp. *lactis* IL1403]	249/ 253 (98)	NP_266293.1	0.0	253
3	2281	1913	122	IS-LL6 transposase/hypothetical protein [*Lactococcus lactis* ssp. *lactis*]	103/ 105 (98)	YP_004761559.1	8e-66	120
4	2623	2303	106	membrane protein [*Lactococcus lactis*]	106/ 106 (100)	YP_001174687.1	3e-69	106
5	3558	2803	251	ABC-type uncharacterized transport system, permease component [*Weissella thailandensis* fsh4-2]	249/ 251 (99)	CCC56315.1	4e-176	251
6	4206	3555	216	ABC transporter ATP-binding protein [*Lactobacillus plantarum* JDM1]	215/ 216 (99)	YP_003064165.1	7e-157	216

7	5178	4288	296	transposase AB of IS*LL6* [*Lactococcus lactis*]	288/ 296 (98)	WP_021723386.1	0.0	296
8	5425	5150	91	transposase [*Lactococcus lactis* ssp. *cremoris* MG1363]	90/ 91 (99)	YP_001031998.1	4e-57	91
9	5835	6977	380	Na+/ H+ antiporter [*Lactococcus lactis* ssp. *lactis*]	376/ 380 (99)	YP_004761490.1	0.0	380
10	6994	8331	445	putative K+ transporter [*Lactococcus lactis* ssp. *lactis* biovar. *diacetylactis*]	445/ 445 (100)	YP_004770067.1	0.0	445
11	8351	9019	222	putative K+ transport system [*Lactococcus lactis*]	222/ 222 (100)	YP_001174694.1	6e-158	222
12	9852	9395	153	mobile element protein [*Lactococcus lactis*]	149/ 151 (99)	WP_023189783.1	2e-104	152
contig_C102								
1	11	427	138	transposase [*Enterococcus faecalis*]	136/ 138 (99)	WP_010711741.1	1e-93	214
2	620	2648	676	potassium transport system protein kup 1 [*Lactococcus lactis* ssp. *lactis* KF147]	651/ 675 (95)	YP_003354939.1	0.0	675
3	3526	2846	226	transposase [*Streptococcus macedonicus* ACA-DC 198]	225/ 226 (99)	YP_005094019.1	3e-163	226
4	3677	4258	193	transposase [*Lactococcus lactis* ssp. *cremoris* SK11]	192/ 193 (99)	YP_811975.1	2e-137	193
5	4207	4611	134	transposase [*Lactococcus lactis*]	122/ 123 (99)	WP_021723058.1	3e-38	287
contig_C104								
1	1241	459	260	gluconate transport protein [*Lactococcus lactis* ssp. *lactis* IO-1]	253/ 253 (100)	YP_007509476.1	1e-172	471
2	1363	1247	38	ABC transporter permease protein [*Lactococcus lactis* ssp. *cremoris* KW2]	37/ 38 (98)	YP_008567945.1	2e-15	266
contig_C72								
1	2380	578	600	oligopeptide transporter protein A [*Lactococcus lactis* ssp. *lactis* biovar. *diacetylactis*]	582/ 600 (97)	YP_004770063.1	0.0	600
2	2755	2486	89	hypothetical protein [*Lactococcus lactis*]	88/ 89 (99)	WP_021214635.1	2e-54	89
contig_C49								
1	140	367	75	hypothetical protein [*Lactococcus lactis*]	75/ 75 (100)	WP_017864513.1	9e-44	107
2	377	565	62	hypothetical protein llmg_1259 [*Lactococcus lactis* ssp. *cremoris* MG1363]	62/ 62 (100)	YP_001032567.1	9e-34	62
3	586	1137	183	putative Gls24 family general stress protein [*Lactococcus lactis* ssp. *lactis* biovar. *diacetylactis*]	183/ 183 (100)	YP_004770016.1	2e-122	183
4	1263	2213	316	Mg^{2+} and Co^{2+} transporter, CorA family [*Lactococcus lactis*]	314/ 316 (99)	WP_021721703.1	0.0	316
5	2252	3193	313	N(5)-(carboxyethyl)ornithine synthase [*Lactococcus garvieae*]	302/ 313 (97)	WP_017371283.1	0.0	313
6	3597	4462	102	hypothetical protein [*Streptococcus criceti*]	74/ 100 (74)	WP_004225830.1	6e-48	105
7	3902	4462	186	lysine decarboxylase [*Streptococcus agalactiae*]	122/ 186 (66)	WP_000704750.1	3e-90	188
8	4626	4432	64	hypothetical protein ACD_45C00259G0003 [uncultured bacterium]	24/ 60 (40)	EKD73576.1	4.3	103
9	5577	4672	301	hypothetical protein CVCAS_0617 [*Lactococcus lactis* ssp. *lactis* CV56]	277/ 301 (92)	YP_005868013.1	0.0	301
10	5793	5587	68	Cro/CI family transcriptional regulator [*Lactococcus lactis* ssp. *lactis* CV56]	68/ 68 (100)	YP_005868014.1	8e-41	120
11	5960	6589	209	hypothetical protein [*Lactococcus garvieae*]	41/ 49 (84)	WP_017371284.1	1e-16	52
contig_C95								
1	144	227	27	protein-tyrosine phosphatase [*Lactococcus lactis* ssp. *lactis* KF147]	25/ 26 (96)	YP_003354428.1	3e-07	248
2	266	415	49	50S ribosomal protein L33 [*Lactococcus lactis* ssp. *lactis* IL1403]	47/ 49 (96)	NP_266249.1	1e-23	49
3	745	542	71	peptide deformylase [*Gracilibacillus halophilus*]	13/ 46 (28)	WP_003467263.1	4.8	184
4	1136	990	48	transposase [*Lactococcus lactis* ssp. *cremoris* TIFN7]	35/ 53 (62)	WP_021211370.1	3e-13	96

5	1244	2146	300	LytR family transcriptional regulator [*Lactococcus lactis* ssp. *lactis* biovar. *diacetylactis* LD61]	298/ 300 (99)	WP_023349368.1	0.0	300
6	3073	2171	300	polysaccharide biosynthesis protein [*Lactococcus lactis* ssp. *lactis* KF147]	294/ 300 (98)	YP_003352616.1	0.0	300
contig_C90								
1	1074	28	348	integrase [*Lactococcus lactis* ssp. *lactis* CV56]	330/ 332 (99)	YP_005868036.1	0.0	379
2	1363	1157	68	excisionase [*Lactococcus lactis* ssp. *lactis* CV56]	67/ 68 (99)	YP_005868035.1	7e-41	68
3	1438	1773	111	IS3 family transposase [*Lactococcus lactis* ssp. *cremoris* UC509.9]	101/ 111 (91)	YP_006998519.1	1e-66	111
contig_C93								
1	78	536	152	mobile element protein [*Lactococcus lactis*]	152/ 152 (100)	WP_023189783.1	5e-107	152
2	696	1094	132	transposase IS1077, partial [*Lactococcus lactis*]	130/ 132 (98)	WP_021164987.1	1e-91	132
3	3298	1180	705	cadmium efflux ATPase *cadA* [*Lactococcus lactis* ssp. *lactis*]	705/ 705 (100)	NP_862600.1	0.0	705
4	3654	3299	119	cadmium efflux system accessory protein [*Enterococcus faecium*]	72/ 118 (61)	WP_016612416.1	2e-43	119
5	3893	4443	184	transposon Tn552 DNA-invertase *binR* [*Lactococcus lactis* ssp. *lactis* biovar. *diacetylactis* LD61]	182/ 184 (99)	WP_023349229.1	3e-129	184
6	5869	4583	428	hypothetical protein LLDT4_03920 [*Lactococcus lactis* ssp. *lactis* biovar. *diacetylactis* TIFN4	416/ 428 (97)	WP_021214708.1	0.0	428
7	6741	5856	294	hypothetical protein [*Lactococcus lactis*]	288/ 294 (88)	WP_017864520.1	0.0	294
8	7109	6746	121	hypothetical protein LLDT4_03930 [*Lactococcus lactis* ssp. *lactis* biovar. *diacetylactis* TIFN4]	121/ 121 (100)	WP_017864521.1	2e-81	121
9	7721	7984	87	transposase of IS*1297* [*Lactococcus lactis*]	72/ 76 (95)	YP_005863074.1	1e-42	226
10	8903	8154	249	nicotinamide mononucleotide transporter [*Lactococcus lactis* ssp. *lactis* KF147]	249/ 249 (100)	YP_001174683.1	9e-173	249
11	9199	9879	226	transposase of IS*946V* [*Lactococcus lactis*]	221/ 226 (98)	YP_001174682.1	2e-162	226
12	10749	10020	242	partition protein B [*Lactococcus lactis*]	242/ 242 (100)	YP_001174681.1	1e-169	242
13	11511	10753	252	partition protein A [*Lactococcus lactis*]	252/ 252 (100)	YP_001174680.1	5e-178	252
14	11754	11521	77	hypothetical protein llh_13910 [*Lactococcus lactis* ssp. *cremoris* A76]	77/ 77 (100)	YP_005877367.1	4e-47	77
15	13618	12266	450	replication protein [*Lactococcus lactis* ssp. *lactis* KF147]	450/ 450 (100)	YP_001174679.1	0.0	450
contig_C98								
1	1184	336	282	family 2 glycosyltransferase [*Lactococcus lactis* ssp. *lactis* KF147]	273/ 278 (98)	YP_003352612.1	0.0	295
2	2262	1198	354	polysaccharide biosynthesis protein [*Lactococcus lactis* ssp. *lactis* KF147]	348/ 354 (98)	YP_003352611.1	0.0	354
3	3390	2320	356	group 1 glycosyltransferase [*Lactococcus lactis* ssp. *lactis* KF147]	347/ 356 (97)	YP_003352610.1	0.0	356
4	4144	3460	228	putative glucose-1-phosphate transferase [*Lactococcus lactis* ssp. *lactis*]	216/ 228 (95)	BAN70288.1	2e-157	228
5	4930	4166	254	tyrosine protein phosphatase [*Lactococcus lactis* ssp. *lactis* KLDS 4.0325]	252/ 254 (99)	YP_008701604.1	0.0	254
6	5680	4985	231	tyrosine protein kinase [*Lactococcus lactis* ssp. *lactis* KLDS 4.0325]	229/ 231 (99)	YP_008701603.1	1e-163	231
7	6469	5690	259	tyrosine-protein kinase trans-membrane modulator *epsC* [*Lactococcus lactis* ssp. *cremoris* A76]	258/ 259 (99)	YP_005874683.1	0.0	259
8	7278	6565	237	polysaccharide biosynthesis protein [*Lactococcus raffinolactis*]	233/ 236 (99)	WP_003137843.1	2e-164	255
9	7530	7312	72	transcription regulator *epsR* [*Lactococcus garvieae* Lg2]	105/ 105 (100)	YP_005870176.1	7e-70	105

10	8030	8611	193	transposase [*Lactococcus lactis* ssp. *cremoris* SK11]	193/ 193 (100)	YP_811975.1	1e-135	193
11	8560	9423	287	transposase [*Lactococcus lactis* ssp. *cremoris* A76]	283/287 (99)	YP_005874679.1	0.0	287
12	10250	9555	231	hypothetical protein [*Lactococcus lactis* ssp. *cremoris* A76]	223/ 231 (97)	YP_005874678.1	5e-158	231
contig_C65								
1	521	964	147	hypothetical protein LLKF_0688 [*Lactococcus lactis* ssp. *lactis* KF147]	81/ 143 (57)	YP_003353152.1	4e-45	159
2	1655	2443	262	hypothetical protein LLKF_0291 [*Lactococcus lactis* ssp. *lactis* KF147]	254/ 262 (97)	YP_003352763.1	9e-178	261
3	2462	2937	158	hypothetical protein lilo_0593 [*Lactococcus lactis* ssp. *lactis* IO-1]	155/ 158 (98)	YP_007507920.1	3e-103	158
4	3805	3050	251	hypothetical protein LLKF_0292 [*Lactococcus lactis* ssp. *lactis* KF147]	242/ 251 (96)	YP_003352765.1	8e-163	251
5	4638	3780	285	ABC transporter ATP-binding protein [*Lactococcus lactis* ssp. *lactis* CV56]	276/ 285 (97)	YP_005869663.1	0.0	285
6	5003	4635	122	R family transcriptional regulator [*Lactococcus lactis* ssp. *lactis* CV56]	121/ 122 (99)	YP_005869662.1	1e-79	122
7	6074	5169	301	hypothetical protein LLKF_0296 [*Lactococcus lactis* ssp. *lactis* KF147]	297/ 301 (99)	YP_003352768.1	0.0	301
8	6440	6048	118	Cro/ CI family transcriptional regulator [*Lactococcus lactis* ssp. *lactis* KF147]	117/ 118 (99)	YP_003352769.1	5e-78	118
9	7341	7478	45	Cro/ Cl family transcriptional regulator [*Lactococcus* garvieae]	41/45 (91)	WP_017371265.1	2e-19	106
10	7600	8689	359	hypothetical protein CVCAS_2301 [*Lactococcus lactis* ssp. *lactis* CV56]	307/ 363 (85)	YP_005869652.1	1e-169	362
11	8812	9081	89	hypothetical protein [*Flavobacterium* sp. SCGC AAA160-P02]	21/ 49 (43)	WP_020081317.1	0.93	818
12	9068	9335	95	testis-specific gene 13 protein [*Condylura cristata*]	23/ 57 (40)	XP_004677245.1	0.25	228
13	9558	9676	70	hypothetical protein PPL_04202 [*Polysphondylium pallidum* PN500]	21/ 61 (34)	EFA82514.1	1.2	419
14	9852	10151	99	hypothetical protein CVCAS_0622 [*Lactococcus lactis* ssp. *lactis* CV56]	98/ 99 (99)	YP_005868017.1	3e-62	99
15	10169	10645	158	hypothetical protein [*Lactococcus garvieae*]	153/ 158 (97)	WP_017371130.1	1e-106	158
16	10649	12335	561	cell division protein *ftsK* [*Lactococcus lactis*]	544/ 561 (97)	WP_003129470.1	0.0	561
17	12813	13967	384	phosphoglucomutase [*Lactococcus lactis* ssp. *lactis* IO-1]	378/ 384 (98)	YP_007507932.1	0.0	384
18	13960	14076	38	hypothetical protein [*Lactococcus lactis*]	36/38 (95)	WP_021214793.1	4e-05	38
19	14104	14433	109	hypothetical protein LLKF_0308 [*Lactococcus lactis* ssp. *lactis* KF147]	107/109 (98)	YP_003352780.1	2e-71	109
20	14520	14831	103	hypothetical protein [*Lactococcus lactis*]	100/ 103 (97)	WP_003129461.1	2e-64	103
21	14960	15220	86	hypothetical protein LLKF_0310 [*Lactococcus lactis* ssp. *lactis* KF147]	83/ 86 (97)	YP_003352782.1	7e-55	86
22	15387	15599	70	hypothetical protein [*Lactococcus lactis*]	69/ 70 (99)	WP_021214796.1	2e-41	70
23	15688	15906	72	hypothetical protein [*Lactococcus lactis*]	66/ 72 (92)	WP_003131791.1	3e-40	72
24	15854	17143	429	hypothetical protein [*Lactococcus lactis*]	425/ 429 (99)	WP_023349079.1	0.0	429
25	17152	17415	87	hypothetical protein LLKF_0314 [*Lactococcus lactis* ssp. *lactis* KF147]	86/ 87 (99)	YP_003352786.1	2e-50	87
26	17429	17839	136	hypothetical protein [*Lactococcus lactis*]	134/ 136 (99)	WP_021214799.1	1e-89	136
27	17843	18112	83	conjugative transposon protein, ATPase [*Lactococcus lactis* ssp. *lactis* KF147]	81/ 83 (98)	YP_003352788.1	4e-46	818
28	18532	18125	135	hypothetical protein, partial [*Lactococcus lactis*]	124/ 130 (95)	WP_021214740.1	8e-88	143
contig_C103								
1	124	441	105	hypothetical protein N036_33270, partial [*Enterococcus gallinarum* EGD-AAK12]	97/ 97 (100)	ERE47138.1	4e-56	555
2	419	961	180	amino acid amidohydrolase [*Lactococcus lactis* ssp. *lactis* CV56]	142/ 161 (89)	YP_005867664.1	6e-94	384
contig_C101								
1	201	428	75	copper chaperone [*Lactococcus lactis* ssp. *lactis* KF147]	75/ 75 (100)	YP_003354950.1	3e-44	75

2	477	1022	181	DNA binding protein [*Lactococcus lactis*]	180/ 181 (99)	YP_001174737.1	2e-127	181
3	1108	1797	229	transcriptional regulator type FNR [*Lactococcus lactis*]	228/ 229 (99)	WP_021723416.1	5e-166	229
4	1933	2097	54	putative secreted protein [*Lactococcus lactis* ssp. *lactis* A12]	54/ 54 (100)	WP_021723415.1	8-27	54
5	2160	2339	59	hypothetical protein LLT5_01685 [*Lactococcus lactis* ssp. *cremoris* TIFN5]	56/ 59 (95)	WP_021214056.1	4e-30	62
6	2403	2579	58	hypothetical protein llmg_0680 [*Lactococcus lactis* ssp. *cremoris* MG1363]	57/ 58 (99)	YP_001032019.1	1e-32	58
7	2678	2538	46	hypothetical protein LLNZ_03620 [*Lactococcus lactis* ssp. *cremoris* NZ9000]	42/ 46 (91)	YP_006356135.1	1e-17	47
8	2771	3704	310	oligopeptide ABC transporter ATP binding protein [*Lactococcus lactis* ssp. *lactis* IL1403]	292/ 309 (94)	NP_267998.1	0.0	338
contig_C105								
1	13	327	104	gluconate permease, Bsu4004 homolog [*Lactococcus lactis*]	104/ 104 (100)	WP_023190130.1	5e-64	234
contig_C107								
1	213	548	111	IS3 family transposase [*Lactococcus lactis* ssp. *cremoris* UC509.9]	101/ 111 (91)	YP_006998519.1	1e-66	111
2	593	1414	273	transposase [*Lactococcus garvieae*]	263/ 273 (96)	WP_019298997.1	0.0	273
contig_C108								
1	1055	688	125	ISLL-6 transposase [*Lactococcus lactis* ssp. *cremoris*]	116/ 123 (94)	AFK83768.1	5e-77	125
2	1860	1042	272	transposase [*Enterococcus faecalis* 62]	269/ 272 (99)	YP_005704815.1	0.0	272
3	2414	1857	185	transposase [*Lactococcus lactis* ssp. *cremoris* SK11]	181/ 185 (98)	YP_808648.1	4e-130	185
Contig_C89								
1	30	146	38	lead, cadmium, zinc and mercury transporting ATPase [*Streptococcus macedonicus* ACA-DC 198]	24/ 24 (100)	YP_005094017.1	4e-07	634
2	382	684	100	hypothetical protein [*Lactococcus lactis*]	100/ 100 (100)	WP_017865389.1	9e-58	100
3	1284	994	96	MerR family transcriptional regulator, partial [*Streptococcus agalactiae*]	55/ 60 (92)	WP_000300474.1	2e-29	102
4	1826	1284	180	transposase [*Enterococcus faecium*]	175/ 180 (97)	WP_023043047.1	2e-123	180
5	1932	1648	94	transposase for insertion sequence-like element IS431mec [*Enterococcus faecium*]	91/ 94 (97)	WP_002330552.1	1e-59	94
6	2114	2338	74	hypothetical protein LLKF_1763 [*Lactococcus lactis* ssp. *lactis* KF147]	74/ 74 (100)	YP_003354168.1	8e-46	74
7	2353	3825	490	cation binding protein, hemerythrin HHE domain family [*Lactococcus lactis* ssp. *lactis* KF147]	483/ 490 (99)	YP_003354167.1	0.0	490
8	4261	3959	100	hypothetical protein uc509_p8064 [*Lactococcus lactis* ssp. *cremoris* UC509.9]	100/ 100 (100)	YP_006998461.1	6e-66	100
9	4535	4404	43	hypothetical protein [*Lactococcus lactis*]	43/ 43 (100)	WP_017865348.1	2e-21	43
10	4837	5133	98	conserved hypothetical protein [*Enterococcus faecalis* E1Sol]	98/ 98 (100)	WP_002375293.1	4e-64	98
11	5233	5472	79	hypothetical protein LLT6_00185 [*Lactococcus lactis* ssp. *cremoris* TIFN6]	64/ 79 (81)	EQC57506.1	5e-15	79
12	5626	6195	189	transposon Tn552 DNA-invertase *binR* [*Lactococcus lactis*]	172/ 188 (91)	WP_017865303.1	5e-119	193
13	7468	6701	255	DeoR family transcriptional regulator [*Lactococcus lactis*]	253/ 255 (99)	WP_017865301.1	0.0	255
14	7950	8375	141	galactose-6-phosphate isomerase subunit *lacA* [*Lactococcus lactis* ssp. *lactis* biovar. *diacetylactis*]	141/ 141 (100)	YP_004770089.1	3e-95	141
15	8392	8907	171	galactose-6-phosphate isomerase subunit *lacB* [*Lactococcus lactis* ssp. *lactis* biovar. *diacetylactis*]	171/ 171 (100)	YP_004770088.1	4e-122	171

16	8918	9850	310	tagatose-6-phosphate kinase [*Lactococcus* garvieae]	310/ 310 (100)	WP_004256005.1	0.0	310
17	9853	10833	326	tagatose 1,6-diphosphate aldolase [*Lactococcus garvieae* IPLA 31405]	325/ 326 (99)	EIT67213.1	0.0	326
18	10861	11178	105	PTS family lactose-N,N-diacetylchitobiose-beta-glucoside transporter subunit IIA [*Lactococcus lactis* ssp. *lactis* biovar. *diacetylactis*]	105/ 105 (100)	YP_004770085.1	2e-69	105
19	11178	12884	568	PTS system lactose-specific transporter subunit IIBC [*Lactococcus lactis* ssp. *lactis* biovar. *diacetylactis*]	566/ 568 (99)	YP_004770084.1	0.0	568
20	13022	14428	468	6-phospho-beta-galactosidase [*Streptococcus salivarius*]	467/ 468 (99)	WP_002889671.1	0.0	487
21	14720	15619	299	protein lacX [*Lactococcus lactis*]	295/ 299 (99)	WP_017865297.1	0.0	299
22	15868	15680	62	hypothetical protein LLT6_12330, partial [*Lactococcus lactis* ssp. *cremoris* TIFN6]	61/ 62 (98)	EQC57084.1	1e-35	81
23	15856	16377	173	gp5 [*Lactococcus lactis*]	169/ 172 (98)	AAC38348.1	2e-113	332
24	16707	17312	201	pyrrolidone-carboxylate peptidase [*Lactococcus lactis* ssp. *cremoris* UC509.9]	198/ 198 (100)	YP_005877563.1	4e-141	215
25	17575	17724	49	transposon resolvase [*Enterococcus faecium*]	40/ 40 (100)	WP_010722379.1	1e-18	115
26	17829	18674	281	transcriptional regulator, AraC family [*Enterococcus* sp. HSIEG1]	281/ 281 (100)	EQC79108.1	0.0	281
27	19340	18843	165	acetyltransferase [*Lactococcus lactis* ssp. *cremoris* SK11]	162/ 165 (98)	YP_796476.1	8e-116	165
28	20957	19755	400	oxalate:formate antiporter, MFS family [*Lactococcus lactis* ssp. *cremoris* A76]	399/ 400 (99)	YP_005877273.1	0.0	418
29	22017	21361	218	transposase IS*1216* [*Lactococcus lactis*]	210/ 218 (96)	WP_017865296.1	3e-153	226
30	22255	22845	196	putative HTH-type transcriptional regulator [*Lactococcus lactis* ssp. *cremoris*]	195/ 196 (99)	YP_006965845.1	2e-138	196
31	22919	23119	66	hypothetical protein [*Lactococcus lactis*]	66/ 66 (100)	WP_017865294.1	7e-41	66
32	23109	23732	207	isochorismatase [*Lactococcus lactis*]	206/ 207 (99)	WP_017865293.1	3e-152	207
33	23758	24165	135	hypothetical protein N036_51545 [*Enterococcus gallinarum* EGD-AAK12]	135/ 135 (100)	ERE60112.1	2e-92	135
34	24162	24506	114	dihydrolipoamide dehydrogenase [*Lactococcus lactis*]	115/ 115 (100)	WP_017865291.1	2e-72	449
Contig_C109								
1	834	154	226	transposase [*Streptococcus macedonicus* ACA-DC 198]	225/ 226 (99)	YP_005094019.1	3e-163	226
2	1153	872	93	transposase [*Lactococcus lactis* ssp. *cremoris* SK11]	81/ 82 (99)	YP_808825.1	3e-52	181
3	1806	1225	193	transposase [*Lactococcus lactis* ssp. *cremoris* SK11]	192/ 193 (99)	YP_811975.1	2e-137	193
4	1957	2241	94	transposase for insertion sequence-like element IS431mec [*Enterococcus faecium*]	93/ 94 (99)	WP_002330552.1	9e-61	94
5	2222	2261	129	transposase [*Streptococcus* sp. C150]	109/ 113 (96)	WP_008534710.1	1e-73	127
Contig_C110								
1	204	1136	310	Oligopeptide transporter D [*Lactococcus lactis* ssp. *lactis* A12]	285/ 303 (94)	WP_021722648.1	0.0	338
2	1218	2177	319	Oligopeptide transporter F [*Lactococcus lactis* ssp. *lactis* A12]	317/ 319 (99)	WP_021468831.1	0.0	319
3	2158	3117	319	Oligopeptide transporter B [*Lactococcus lactis* ssp. *lactis* A12]	317/ 319 (99)	WP_021722649.1	0.0	319
4	3127	3564	146	Oligopeptide transport system permease protein oppC [*Lactococcus lactis* ssp. *lactis* A12]	146/ 146 (100)	YP_004770062.1	1e-96	294
Contig_C111								
1	1331	2	443	endopeptidase *pepO2* [*Lactococcus lactis* ssp. *lactis* A12]	412/ 426 (97)	WP_021722256.1	0.0	627
2	1315	1458	45	unknown protein				
Contig_C112								

1	130	1445	439	histidine kinase, partial [*Lactococcus lactis*]	513/ 515 (99)	WP_017864819.1	0.0	475
Contig_C113								
1	1070	1	356	ABC transporter ATP-binding protein, partial [*Enterococcus gallinarum* EGD-AAK12]	110/ 112 (98)	ERE51620.1	0.0	421
Contig C114								
1	154	1	47	UDP-diphosphatase [*Lactococcus lactis* ssp. *lactis* KF147]	43/ 47 (91)	YP_003354828.1	2e-21	284
2	223	1014	264	1,4-alpha-glucan-branching protein [*Lactococcus lactis* ssp. *lactis* KF147]	262/ 264 (99)	YP_003352601.1	0.0	648
Contig_C116								
1	227	1	409	Predicted cell-wall-anchored protein sasA (LPXTG motif) [*Lactococcus lactis* ssp. *lactis* Dephy 1]	393/ 409 (96)	WP_023188733.1	0.0	809

Appendix 5-3

ORF table of non-alignment contigs to reference genomes for DPC6860 (Grass)

ORF	Start	Stop	Size (aa)	Predicted function	Identity	Accession no.	E-value	Size
Contig_G74								
1	318	1	106	transposase [*Enterococcus faecium* EnGen0264]	95/ 100 (95)	WP_010723073.1	4e-62	256
2	424	1059	211	hypothetical protein LLT7_02890 [*Lactococcus lactis* ssp. *cremoris* TIFN7]	200/ 211 (95)	WP_021211520.1	3e-62	206
3	1526	2450	310	hypothetical protein CVCAS_pA0018 [*Lactococcus lactis* ssp. *lactis* CV56]	267/ 309 (86)	YP_005867344.1	0.0	307
4	3501	2781	241	DeoR family transcriptional regulator [*Lactococcus lactis* ssp. *cremoris* TIFN5]	231/ 241 (96)	WP_021166432.1	2e-162	243
5	4025	4450	141	galactose 6-P isomerase lacA subunit [*Lactococcus lactis* ssp. *cremoris* A76]	141/ 141 (100)	YP_004761520.1	3e-95	141
6	4467	4982	171	galactose 6-P isomerase subunit lacB [*Lactococcus lactis* ssp. *lactis*]	171/ 171 (100)	YP_004761519.1	7e-122	171
7	4993	5925	310	tagatose 6-P kinase [*Lactococcus lactis* ssp. *lactis*]	310/ 310 (100)	YP_004761518.1	0.0	310
8	5928	6908	326	tagatose 1,6-diphosphate aldolase [*Lactococcus lactis* ssp. *cremoris* SK11]	326/ 326 (100)	YP_796532.1	0.0	326
9	6936	7253	105	PTS family lactose-N,N-diacetylchitobiose-beta-glucoside transporter subunit IIA [*Lactococcus lactis* ssp. *lactis* biovar. *diacetylactis*]	103/ 105 (98)	YP_004770085.1	1e-68	105
10	7258	8964	568	PTS lactose transporter subunit IIBC [*Lactococcus lactis* ssp. *lactis* biovar. *diacetylactis* str. LD61]	566/ 568 (99)	WP_003131734.1	0.0	568
11	9024	10475	477	6-phospho-beta-galactosidase [*Lactococcus lactis* ssp. *cremoris* TIFN7]	476/ 477 (99)	YP_004761515.1	0.0	477
12	10767	11666	299	*lacX* [*Lactococcus lactis* ssp. *cremoris*]	296/ 299 (99)	YP_006965846.1	0.0	299
13	12408	11728	226	transposase [*Listeria inocua* Clip11262]	222/ 226 (98)	NP_569202.1	2e-162	226
14	13457	12558	299	*prtM* [*Lactococcus lactis* ssp. *lactis* biovar. *diacetylactis*]	297/ 299 (99)	ACF37493.1	0.0	299
15	13792	19680	1962	*prtP* lactocepin I [*Lactococcus lactis* ssp. *cremoris* UC509.9]	1914/ 1962 (98)	YP_007000506.1	0.0	1974
16	20608	19928	226	transposase element IS*1297* [*Streptococcus thermophilus* MN-ZLW-002]	220/ 226 (97)	YP_006339954.1	2e-161	226
17	22485	21136	449	dihydrolipoamide dehydrogenase [*Lactococcus lactis* ssp. *cremoris*]	448/ 449 (99)	YP_006965841.1	0.0	449
18	22889	22503	128	hypothetical protein pLP712_46 [*Lactococcus lactis* ssp. *cremoris*]	128/ 128 (100)	YP_006965842.1	3e-87	128
19	23538	22915	207	isochorismatase family protein [*Lactococcus lactis* ssp. *cremoris*]	207/ 207 (100)	YP_006965843.1	1e-152	207
20	23729	23565	54	conserved hypothetical protein [*Enterococcus faecalis* T8]	51/ 54 (94)	WP_002388567.1	1e-27	56
21	24393	23803	196	putative HTH-type transcriptional regulator [*Lactococcus lactis* ssp. *cremoris*]	196/ 196 (100)	YP_006965845.1	1e-139	196
22	24601	25005	135	transposase [*Lactococcus lactis*]	134/ 135 (99)	CAA42988.1	8e-93	175
Contig_G75								
1	3	664	213	hypothetical protein [*Lactococcus lactis* ssp. *cremoris* A76]	203/ 207 (98)	YP_005874678.1	3e-143	231
2	1513	774	247	transposase [*Lactococcus lactis* ssp. *cremoris* A76]	239/ 247 (97)	YP_005874679.1	8e-174	287
3	2166	1585	193	transposase [*Lactococcus lactis* ssp. *cremoris* SK11]	191/ 193 (99)	YP_811975.1	5e-138	193
4	2222	2527	101	transposase [*Lactococcus lactis* ssp. *lactis* IO-1]	100/ 101 (99)	YP_007507882.1	4e-64	101
Contig_G76								
1	468	163	101	transposase [*Lactococcus lactis* ssp. *lactis* IO-1]	100/ 101 (99)	YP_007507882.1	4e-64	101
2	1497	1922	141	ArsC family protein [*Lactococcus lactis* ssp. *cremoris* KW2]	135/ 141 (96)	YP_008568786.1	2e-92	141

3	2697	2287	136	transcriptional regulator [*Lactococcus lactis* ssp. *cremoris* SK11]	135/ 136 (99)	YP_809305.1	3e-93	136
4	2870	3727	285	aldo/ keto reductase family protein [*Lactococcus lactis* ssp. *cremoris* KW2]	280/ 285 (98)	YP_008568788.1	0.0	285
5	4023	4268	81	hypothetical protein kw2_1146 [*Lactococcus lactis* ssp. *cremoris* KW2]	81/ 81 (100)	YP_008568789.1	1e-47	81
6	5012	4689	108	phage-associated peptidase [*Lactococcus lactis* ssp. *cremoris* SK11]	26/ 43 (60)	YP_809104.1	4e-06	979
Contig_G77								
1	536	2	178	hypothetical protein [*Lactococcus lactis* ssp. *cremoris* A76]	170/ 174 (98)	YP_005874678.1	3e-116	231
Contig_G78								
1	3	1247	414	cation transporter E1-E2 family ATPase [*Lactococcus lactis* ssp. *cremoris* MG1363]	403/ 406 (99)	YP_001031895.1	0.0	775
Contig_G79								
1	523	398	41	hypothetical protein LLT6_10665 [*Lactococcus lactis* ssp. *cremoris* TIFN6]	40/ 41 (98)	EQC58098.1	5e-19	41
2	692	964	91	unknown				
3	1041	889	50	integrase [*Lactococcus* phage r1t]	34/ 38 (89)	NP_695029.1	2e-13	374
Contig_G101								
1	61	1224	388	DNA adenine methylase [*Streptococcus macacae* NCTC 11558]	245/ 385 (64)	WP_003081814.1	4e-166	653
Contig_G92								
1	261	271	136	cyclic nucleotide-binding domain-containing protein [*Lactococcus lactis* ssp. *cremoris* MG1363]	124/ 133 (93)	YP_001032544.1	2e-79	218
2	677	874	65	transposase IS*431*, partial [*Lactococcus lactis*]	64/ 65 (98)	WP_021214046.1	1e-37	67
3	2532	1258	424	major facilitator superfamily permease [*Lactococcus lactis* ssp. *cremoris*]	424/ 424 (100)	YP_006965834.1	0.0	457
Contig_G88								
1	1074	37	345	multi-copper oxidase [*Lactobacillus mali* KCTC 3596 = DSM 20444]	345/ 345 (100)	WP_003690689.1	0.0	345
Contig_G89								
1	286	1	95	D-alanyl-D-alanine carboxypeptidase [*Lactococcus lactis* ssp. *cremoris* SK11]	79/ 96 (82)	YP_808983.1	3e-38	248
2	907	441	115	bactoprenol glucosyl transferase [*Lactococcus lactis* ssp. *cremoris* MG1363]	144/ 152 (95)	YP_001031496.1	3e-96	316
3	968	1172	68	transposase IS*431*, partial [*Lactococcus lactis* ssp. *cremoris* TIFN6]	59/ 68 (87)	EQC54564.1	2e-32	80
Contig_G96								
1	841	221	206	DNA methyltransferase [*Haemophilus influenzae*]	120/ 206 (58)	WP_005671623.1	8e-75	360
2	1555	2286	244	RepB protein [*Lactococcus lactis* ssp. *cremoris* TIFN7]	241/ 244 (99)	WP_021211532.1	2e-176	388
Contig_G98								
1	10	417	135	transposase IS*RSO8*, partial [Lactococcus garvieae]	134/ 135 (99)	WP_017371154.1	3e-96	135
2	496	771	91	transposase A of IS*LL6* [*Lactococcus lactis*]	90/ 91 (99)	YP_001174741.1	9e-58	91
3	816	1676	286	hypothetical protein [Lactococcus garvieae]	261/ 273 (96)	WP_016170872.1	0.0	273
4	2020	1673	115	transposase IS*RSO8*, partial [*Lactococcus garvieae*]	114/ 115 (99)	WP_017371154.1	3e-80	135
Contig_G91								
1	950	597	117	hypothetical protein SMA_2045 [*Streptococcus macedonicus* ACA-DC 198]	109/ 113 (96)	YP_005095608.1	9e-70	126
2	1074	1419	115	signal transduction histidine kinase *lytS* [*Streptococcus salivarius* JIM8777]	111/ 113 (98)	YP_006070303.1	4e-70	200
Contig_G95								

1	423	22	133	putative transposase for insertion sequence element [*Tetragenococcus halophilus* NBRC 12172]	129/ 131 (98)	YP_004887832.1	2e-88	226
2	757	1251	164	mobilization protein [*Enterococcus faecalis*]	162/ 164 (99)	WP_002395943.1	3e-112	164
3	1230	1745	172	relaxase/mobilization nuclease domain protein, partial [*Enterococcus faecalis*]	147/ 161 (91)	WP_002395941.1	8e-101	181
Contig_G90								
1	59	415	118	cadmium efflux ATPase *cadA* [*Lactococcus lactis* ssp. *lactis*]	118/ 118 (100)	YP_004761537.1	2e-70	705
2	1432	797	211	PIL6_6 [*Lactococcus lactis* ssp. *lactis*]	206/ 211 (98)	YP_006962314.1	2e-143	211
3	1977	2730	251	type II restriction endonuclease [*Streptococcus equi* ssp. *zooepidemicus*]	153/ 250 (61)	YP_002744129.1	1e-100	580
Contig_G99								
1	227	51	58	hypothetical protein LLT7_01885 [*Lactococcus lactis* ssp. *cremoris* TIFN7]	49/ 58 (84)	WP_021211526.1	4e-22	62
2	714	364	116	Replication-associated protein [*Lactococcus lactis*]	69/ 108 (64)	WP_023189083.1	2e-31	106
3	1138	707	143	*repB* [*Lactococcus lactis* ssp. *lactis* Dephy 1]	135/ 140 (96)	YP_001098359.1	3e-88	383
Contig_G84								
1	58	730	223	relaxase [*Lactococcus lactis*]	212/ 223 (95)	WP_023349257.1	4e-146	410
2	727	1350	207	putative mobilization protein [*Lactococcus lactis* ssp. *cremoris*]	186/ 207 (90)	NP_053038.1	2e-131	207
Contig_G102								
1	84	389	101	transposase [*Lactococcus lactis* ssp. *lactis* IO-1]	100/ 101 (99)	YP_007507882.1	4e-64	101
2	485	1246	253	transposase of IS*904B* [*Lactococcus lactis* ssp. *lactis* IL1403]	252/ 253 (99)	NP_266293.1	0.0	253
Contig_G81								
1	353	108	81	RepB protein [*Lactococcus garvieae*]	77/ 77 (100)	WP_017370412.1	2e-44	385
2	1243	728	171	putative mobilization protein [*Lactococcus lactis* ssp. *lactis*]	171/ 171 (100)	YP_004761554.1	9e-121	200
Contig_G92								
1	1052	11	346	restriction endonuclease [*Streptococcus equi* ssp. *equi* 4047]	189/ 334 (57)	YP_002746828.1	2e-126	541
Contig_G86								
1	1	531	176	RepB protein [*Lactococcus garvieae*]	175/ 176 (99)	WP_017370412.1	4e-120	385
2	528	1118	196	RepB family protein [*Lactococcus lactis*]	153/ 197 (78)	WP_023349254.1	3e-96	197
Contig_G104								
1	348	1028	226	transposase IS*1216* [*Aerococcus viridans*]	225/ 226 (99)	WP_016897246.1	2e-163	226
2	1353	1763	136	GNAT family acetyltransferase [*Lactobacillales*]	136/ 136 (100)	WP_017865380.1	2e-92	136
3	2625	2035	196	resolvase [*Lactobacillales*]	195/ 196 (99)	YP_004761469.1	2e-137	196
4	3365	3012	117	hypothetical protein SMA_2045 [*Streptococcus macedonicus* ACA-DC 198]	109/ 113 (96)	YP_005095608.1	9e-70	126
5	3489	4091	200	hypothetical protein LBUCD034_1995 [*Lactobacillus buchneri* CD034]	197/ 200 (99)	YP_006726566.1	6e-139	200
6	4091	4882	263	ABC transporter permease [*Streptococcus oralis*]	260/ 263 (99)	WP_000130447.1	0.0	263
7	4858	5694	278	ABC transporter [*Lactobacillus helveticus* DPC 4571]	273/ 278 (98)	YP_001578094.1	0.0	278
8	5681	6481	266	cobalt ABC transporter ATPase [*Streptococcus parasanguinis*]	263/ 266 (99)	WP_003005192.1	0.0	266
9	6869	6696	57	HsdD protein [*Streptococcus gordonii* str. challis substr. CH1]	56/ 57 (98)	YP_001450217.1	6e-33	57
10	7435	6845	196	RepB family protein [*Lactococcus lactis* ssp. *lactis* biovar. *diacetylactis* str. LD61]	153/ 197 (78)	WP_023349254.1	3e-96	197
11	8589	7432	385	RepB protein [*Lactococcus garvieae*]	384/ 385 (99)	WP_017370412.1	0.0	385

12	9566	8964	200	putative mobilization protein [*Lactococcus lactis* ssp. *lactis*]	198/ 200 (99)	YP_004761554.1	2e-140	200
13	10206	9583	207	mobilization protein [*Lactococcus raffinolactis*]	203/ 207 (98)	WP_003140517.1	1e-145	207
14	11435	10203	410	Relaxase/ Mobilisation nuclease domain [*Enterococcus* sp. 7L76]	403/ 410 (98)	YP_007809070.1	0.0	410
15	11908	11414	164	mobilization protein [*Enterococcus faecalis*]	162/ 164 (99)	WP_002395943.1	3e-112	164
Contig_G105								
1	284	132	50	hypothetical protein llmg_0669 [*Lactococcus lactis* ssp. *cremoris* MG1363]	46/ 50 (92)	YP_001032008.1	8e-24	50
2	637	335	100	hypothetical protein llmg_0670 [*Lactococcus lactis* ssp. *cremoris* MG1363]	99/ 100 (99)	YP_001032009.1	9e-65	100
3	810	634	58	hypothetical protein LLT6_04645 [*Lactococcus lactis* ssp. *cremoris* TIFN6]	55/ 58 (95)	WP_021213339.1	5e-31	58
4	1213	947	88	hypothetical protein LLNZ_03480 [*Lactococcus lactis* ssp. *cremoris* NZ9000]	87/ 88 (99)	YP_006356107.1	6e-57	88
5	1828	1358	156	transposase for insertion sequence element IS*904* [*Lactococcus lactis* ssp. *cremoris* A76]	143/ 148 (97)	YP_005875645.1	3e-103	152
Contig_G106								
1	2	910	302	RepB protein [*Lactococcus lactis* ssp. *cremoris* TIFN7]	286/ 302 (95)	WP_021211532.1	0.0	388
2	903	1577	224	OrfX replication associated protein [*Lactococcus lactis* ssp. *cremoris* UC509.9]	170/ 224 (76)	YP_006998518.1	3e-107	176
3	2104	1895	69	hypothetical protein LLT5_12605 [*Lactococcus lactis* ssp. *cremoris* TIFN5]	60/ 61 (98)	WP_021213809.1	9e-34	69
4	2220	2041	59	hypothetical protein LACR_0119 [*Lactococcus lactis* ssp. *cremoris* SK11]	47/ 47 (100)	YP_808164.1	4e-23	60
5	3602	2361	413	hypothetical protein LACR_0118 [*Lactococcus lactis* ssp .*cremoris* SK11]	407/ 411 (99)	YP_808163.1	0.0	411
6	4046	3581	154	bactoprenol glucosyl transferase [*Lactococcus lactis* ssp. *cremoris* MG1363]	143/ 152 (94)	YP_001031496.1	9e-94	316
7	4077	4787	236	transposase IS*1216 [Lactococcus lactis*]	235/ 236 (99)	WP_023349245.1	8e-172	236
8	4846	6411	521	putative multi-copper oxidase [*Lactococcus lactis* ssp. *lactis*]	519/ 520 (99)	YP_004761475.1	0.0	520
9	6451	7401	316	MFS transporter permease [*Lactococcus lactis* ssp. *cremoris* TIFN5]	312/ 314 (99)	WP_021166413.1	0.0	457
Contig_G107								
1	393	103	96	hypothetical protein [*Lactococcus lactis*]	48/ 109 (44)	WP_017864523.1	3e-19	109
2	1213	560	217	putative replication protein [*Lactococcus garvieae*]	192/ 216 (89)	YP_005352348.1	2e-142	217
3	1428	1312	38	hypothetical protein LLDT4_13465 [*Lactococcus lactis* ssp. *lactis* biovar. *diacetylactis* TIFN4]	23/ 25 (92)	WP_021214973.1	3e-08	138
4	1398	2108	236	transposase IS*1216* [*Lactococcus lactis*]	234/ 236 (99)	WP_023349245.1	3e-171	236
5	2791	2492	99	major facilitator superfamily permease, partial [*Lactobacillus paracasei*]	99/ 99 (100)	WP_016379903.1	5e-59	188
Contig_G108								
1	722	42	223	transposase [*Enterococcus durans*]	223/ 226 (99)	WP_016177626.1	3e-162	226
2	1117	2868	583	type II restriction endonuclease [*Streptococcus equi* ssp. *zooepidemicus*]	343/ 576 (60)	YP_002744129.1	0.0	580
3	4960	2932	676	Adenine-specific methyltransferase [*Enterococcus faecalis* MTUP9]	419/ 656 (64)	KAJ81323.1	0.0	670
Contig_G109								
1	436	776	112	abortive infection mechanism-related protein [*Lactococcus lactis* ssp. *cremoris* UC509.9]	81/ 83 (98)	YP_006998531.1	2e-47	220
2	1441	872	189	site-specific recombinase [*Lactococcus lactis* ssp. *cremoris*]	189/ 189 (100)	AFK83769.1	2e-136	189
3	1670	2029	119	cadmium resistance regulator *cadC* [*Lactococcus lactis* ssp. *lactis* biovar. *diacetylactis*]	119/ 119 (100)	YP_004770500.1	1e-78	119
4	2026	4143	705	cadmium efflux ATPase *cadA* [*Lactococcus lactis* ssp. *lactis*]	705/ 705 (100)	NP_862600.1	0.0	705

5	5159	4525	211	PIL6_6 [*Lactococcus lactis* ssp. *lactis*]	204/ 211 (97)	YP_006962314.1	3e-142	211
Contig_G110								
1	422	75	115	transposase, partial [*Lactococcus garvieae*]	115/ 115 (100)	WP_017370934.1	2e-75	115
2	532	1167	211	PIL6_6 [*Lactococcus lactis* ssp. *lactis*]	207/ 211 (98)	YP_006962314.1	1e-144	211
3	1348	1908	201	transposase IS*1216* [*Lactococcus lactis* ssp. *lactis* biovar. *diacetylactis* str. LD61]	201/ 205 (98)	WP_023349245.1	5e-144	236

Chapter 6

General discussion

6.1. General discussion

Lactococcus lactis is a lactic acid bacterium (LAB) used the world over in the manufacture of fermented dairy products, most notably cheese. It is widely held that dairy strains used in modern dairy fermentations are descended from plant associated ancestors (Kelly *et al.*, 2010). The adaptation of these organisms to milk is likely to have been a long and complex process involving numerous events of plasmid gain and loss, and genome reduction (Passerini *et al.*, 2010). The movement from plant matter to milk is associated with both gene loss and genome decay, and the acquisition of genes enabling rapid growth in milk. By examining non-dairy strains of *L. lactis* we are essentially turning back the clock to before these organisms were domesticated to milk, and lost many of the physiological traits which enabled them to thrive in harsher environments. With the abolishment of milk quotas in Ireland in 2015, an increased volume of milk will be available on the market and subsequently fermented dairy products, such as cheese, are foreseen as target end products. In terms of diversifying cheese flavour, the use of non-dairy cultures could allow companies to expand their product range without the added cost of purchasing new processing equipment. Previous studies have investigated the application of 'wild' LAB to dairy fermentations (Ayad *et al.*, 2000; Randazzo *et al.*, 2008; Ciocia *et al.*, 2013); however, it is important to highlight the potential drawbacks of the use of these cultures such as biogenic amine production, antibiotic resistance or off-flavour production.

With the increased isolation and characterisation of 'wild' LAB isolates, much interest surrounds their application in food fermentations and their capacity for flavour diversification. This thesis focuses on the characterisation of non-

dairy lactococci and their application to dairy fermentations, specifically cheese. It examines the genetic diversity of *L. lactis* subspecies and highlights the potential issues of *L. lactis* subspecies taxonomy. A bank of non-dairy lactococci was examined for traits important for use in dairy applications with emphasis on flavour diversification and phage insensitivity. Furthermore, this thesis examined non-conserved genes in non-dairy lactococci which may offer strains from outside the milk environment an advantage in their respective niche. Non-dairy habitats remain a rich source of *L. lactis* strains and could potentially solve some of the problems encountered with the use of lactococcal strains in dairy fermentations such as vat failure caused by bacteriophage.

Chapter 2 describes the phenotype-genotype mismatching of *L. lactis* from grass, vegetables and rumen samples with applications in dairy fermentations. *L. lactis* strains were isolated from non-dairy origins which were capable of growing in milk. The characterisation of subspecies is important in the development of cultures for cheese production as *cremoris* strains are more suited to Cheddar cheese production creating cheeses with better flavour and with little bitterness, while both are used in the production of certain French cheese varieties (Godon *et al.*, 1992). In all, eight strains were isolated, seven of which possessed a *cremoris* genotype while all strains possessed a *lactis* phenotype. This supports the hypothesis that *cremoris* phenotypes are a rarity outside of the milk niche (Kelly and Ward, 2002). To date much discussion surrounds the classification of *L. lactis* subspecies. Numerous studies have classified *L. lactis* on the basis of MLST but, to date, none have performed species-wide analysis using average nucleotide identity (ANI) (Rademaker *et al.*, 2007; Taïbi *et al.*, 2010; Fernández *et al.*, 2011). Based on ANI and tetra nucleotide (TETRA) analysis performed in

this study, *L. lactis* ssp. *lactis* strains from 'wild' environments, including strain IL1403, appear to form a new species. The grouping of IL1403 with non-dairy lactis strains may be due to the plasmid cured nature of the strain. Future analysis of *L. lactis* genomes using ANI and TETRA which includes the parental strain of IL1403 may shed further light on the results obtained from this analysis. Based on ANI values only subspecies *cremoris* and *lactis* appear to form different species. However, this analysis differentiated species based on genomic analysis only and does not take into account the physical traits of the strains. If the current taxonomy of *L. lactis* is revised, designating subspecies *cremoris* strains as *Lactococcus paralactis* and subspecies *lactis* strains *Lactococcus lactis* could be a feasible option.

A polyphasic approach to prokaryotic taxonomy takes into account both the phenotypic and genotypic traits of an organism, and with specific parameters in place, may be the most suitable means for prokaryote taxonomy in the future (Kämpfer and Glaeser, 2012). In an age where sequence data and software tools are readily available, it is crucial to describe the phenotype to a sufficient degree to allow for appropriate circumscription. Genome sequence data obtained from the NCBI database, included reports where no information on strain phenotype was available. Many of these reports corresponded to genome announcements, which offer little information on the physical characterisation of the strain. Therefore, the subspecies phenotype of each strain was not determined and, pending further analysis, we may only speculate as to whether a proposal to define a new *L. lactis* species is valid one. This in turn highlights the importance to the *L. lactis* community of stating both the phenotype and genotype for analysis such as that performed in this study.

Chapter 3 describes the application of non-dairy cultures isolated in Chapter 2 to a miniature Gouda-type cheese model. In milk fermentations, non-dairy lactococci produced aroma profiles unlike those of their dairy counterparts and, on this basis, five strains were selected as adjuncts in a mini cheese system. Although a slower acidification rate was observed for 2 cheeses made using non-dairy strains; overall, the application of non-dairy lactococci to cheese making in conjunction with a commercial lactococcal starter mix did not negatively impact on the cheese make. Similarily, the addition of non-dairy cultures did not adversely impact on cheese composition. During cheese ripening, bacteria are exposed to low temperature, differential nutrient availability and varying moisture content, at a pH of 4.9-5.3 (Cogan *et al.*, 2007; Settanni and Moschetti, 2010). From day 14 to day 42, viable cell counts on LM17 agar plates increased for cheeses made using DPC6860 as an adjunct. Recently transcriptomic analysis of *Lactobacillus rhamnosus*, a non-starter LAB, during cheese ripening, revealed that this organism utilises pyruvate and ribose released by cell lysis for growth (Lazzi *et al.*, 2014). The exploitation of cellular lysis products may be responsible for the increase in viable cell numbers of *L. lactis* DPC6860 during cheese ripening. With the elucidation of non-dairy *L. lactis* genomes (Chapter 5) transcriptome analysis of these strains in cheese, may prove a useful in understanding the growth of DPC6860 in cheese.

Flash sensory analysis was performed using 21 descriptors which represent the main attributes of Gouda cheese. In comparison to flash sensory analysis, descriptive sensory analysis uses a trained sensory panel to detect and describe the quantitative and qualitative constituents of a food product (Meilgaard *et al.*, 2006). Sensory analysis of Gouda-type cheeses showed that non-dairy strains

formed different flavour profiles in comparison to those made with dairy cultures only; with non-dairy cheeses more associated with a nutty flavour and aroma. However, a large difference was observed in moisture levels between cheeses and the experiment should be repeated in a model where similar moisture levels can be created in all cheeses to confirm these results. In aged Cheddar cheese Strecker aldehydes (2-methylpropanal, 2-methylbutanal, 3-methylbutanal) have been shown to contribute greatly to the formation of a nutty flavour; however, without volatile analysis of these cheeses, we may only speculate as to the compounds responsible for the nutty flavour and aroma (Avsar *et al.*, 2004). It is important to note that the volatile compounds detected by instruments such as GC-MS may not give a true account of the aroma of that compound in food (Drake and Civille, 2003). Gas chromatography olfactometry (GC-O) permits assessors to identify and quantify the compounds which form an odour (Delahunty *et al.*, 2006). Although not used in this study, tandem use of descriptive sensory analysis with GC-O coupled with GC-MS, may offer a valuable insight into the compounds produced by non-dairy strains and their odour forming capacity during cheese ripening in the future.

Bitterness is an undesirable trait amongst cheese producers and was found to be strongly associated with cheeses made using non-dairy cultures in this study. Microfluidisation is emerging as a means of altering cheese flavour which involves forcing bacterial cells at high pressure against a solid surface within a chamber, permeabilising the cells, which slowly leak their intra-cellular contents into the cheese matrix (Yarlagadda *et al.*, 2014). The use of microfluidisation has been successfully applied to adjunct cultures used in a Gouda-type cheese, whereby industrial dairy lactococci (with the exception of strain Z8) formed

flavour profiles different from cheeses made with a commercial dairy adjunct (Yarlagadda *et al.*, 2014). In Chapter 3, this pre-treatment of cells showed potential in the reduction of bitterness using non-dairy lactococci, and may be a viable route for the future application of these strains. Although non-dairy strains showed potential to diversify flavour in dairy fermentations, consumer acceptability of the use of such cultures remains to be established. Artisan and raw milk cheeses contain a multitude of bacteria from different environments; for example, analysis of the microbiota of artisanal cheeses identified bacteria belonging to the *Prevotella* genus which are associated with the gastrointestinal tract of hooved animals (Quigley *et al.*, 2012). Time will tell whether such strains have a future in the dairy industry; however, further analyses in terms of sensory analysis and examining specific aromas generated by these strains is warranted.

Attack by bacteriophage remains a significant problem in the dairy industry, leading to large economic losses each year. Thus, the isolation of strains with increased insensitivity/ resistance to dairy lactococcal phages could prove useful in minimising vat failure. Chapter 4 describes the phage insensitivity of non-dairy lactococci isolated in Chapter 2 to dairy phages. This in turn led to the isolation of ΦL47, which infects the non-dairy strain DPC6860. Overall, non-dairy lactococci were mostly insensitive to dairy lactococcal phages. In dairy fermentations such strains may be useful in starter culture rotations to reduce phage titres from reaching high enough numbers to affect fermentation. Careful use of these strains must be employed however, as phages are capable of rapidly overcoming host defences and may adapt to lyse non-dairy strains if they are not properly employed.

To further our knowledge of lactococcal phages, and particularly those capable of infecting non-dairy isolates, we set about isolating lactococcal phages from raw sewage. Due to the relatively limited sequence data available on large lactococcal phages, ΦL47 was selected for genome sequencing based on restriction digest profiles. Comparative genome analysis showed that ΦL47 was most similar to Φ949 a large lactococcal phage isolated almost 40 years ago. A multiple alignment of Φ949 with previously sequenced *L. lactis* phages showed that this phage formed a single group; however, with the isolation of ΦL47 it appears that Φ949 is not unique, but the first representative of the so-called '949 group' (Samson and Moineau, 2010). In contrast to Φ949, imaging of ΦL47 showed that this phage possesses a long tail fiber not previously reported in dairy lactococcal phages. Four ORFs sharing 60-94% aa identity to ΦbIL286 were putatively assigned to the formation of this tail fiber. Tail fibers of other Siphoviridae are held to be important in different aspects of host infection such as cell adsorption (Stockdale *et al.*, 2013). Mutational events occurring in the tail fiber gene have been shown to modify the host range of the phage as shown in *Bacillus anthracis* phages Wβ and γ (Schuch and Fischetti, 2006). Characterisation studies mapping the role the tail fiber plays in host specificity may prove useful in understanding the insensitivity of non-dairy strains to dairy phages and determine the importance of this extended tail fiber in phage infection.

The final study, Chapter 5, analysed three selected non-dairy lactococcal strains for unconserved ORFs which may offer a competitive advantage in their respective environments. This was performed in tandem with detailed analysis of contigs which did not align with reference genomes (IL1403 and MG1363) and may correspond to putative plasmids or genomic islands. All non-dairy isolates

possessed gene clusters associated with dairy organisms such as those encoding lactose utilisation. Only one strain, DPC6856, encoded the extra-cellular proteinase, *prtP*, and its maturation protein, *prtM*. It is of interest to note that only the strain associated with cattle encoded these genes which may hint at a possible route for the acquisition of genes enabling rapid growth in milk. Therefore, the question must be raised therefore: why do these non-dairy isolates encode genes for the utilisation of a milk-associated sugar and protein? One hypothesis involves the acquisition of plasmids encoding for lactose or casein utilisation which also encode for other important traits such as bacteriophage resistance, heavy metal resistance or bacteriocin production. Corresponding with Chapter 4, non-dairy strains appeared to possess diverse phage resistance mechanisms which may contribute to their insensitivity to dairy lactococcal phages. The possession of these systems may be of benefit to currently established dairy systems and may contribute to enhancing phage resistance in specific dairy strains.

Non-dairy environments are more varied in their composition in terms of their chemical make-up, and bacterial inhabitants are exposed to various stresses that dairy organisms are not. Previously, genome analyses of other non-dairy lactococci have revealed information on their ability to thrive in their respective environments (Siezen *et al.*, 2008; Passerini *et al.*, 2013). In the current study, this was most strongly evident for genes involved in carbohydrate utilisation. For example, it is presumable that DPC6856 came to inhabit the rumen through the animal's ingestion of grass; however, both DPC6860 and DPC6856 possess diverse genes involved in carbohydrate utilisation. In turn, this suggests that DPC6856 is not transient, but a common inhabitant of the rumen. This was further supported by lack of a Co^{2+} transport system in DPC6856 and absence of genes

encoding EPS production. Overall, this chapter highlights the diversity that exists amongst *L. lactis* from different environmental niches and the traits they possess which reflect the environment that they inhabit.

In conclusion, the results presented in this thesis demonstrate the diversity that exists within *L. lactis* from non-dairy environments, and the potential application of these non-dairy lactococci to cheese making. Phenotype-genotype disparity in *L. lactis* highlights the potential need for a revised means of subspecies description whereby both traits are explicitly understood. Non-dairy strains exhibit diverse amino acid transferase activity which in dairy fermentations may lead to the formation of diverse flavour profiles. In milk and cheese matrices, non-dairy strains can diversify volatile and flavour profiles as determined by GC-MS and flash sensory analysis respectively. Not only can such strains be used to alter flavour, but they also appear to possess enhanced insensitivity to dairy lactococcal phages which may be of added benefit when applying these strains to starter systems. Genome analysis of non-dairy strains revealed numerous genes/ gene clusters which may be important in niche specialisation and survival in their respective environments. Overall, this thesis demonstrates the genetic and metabolic diversity of non-dairy *L. lactis* isolates and the potential application of these strains to cheese diversification.

6.2. References

Avsar, Y., Karagul-Yuceer, Y., Drake, M., Singh, T., Yoon, Y., and Cadwallader, K. (2004). Characterization of nutty flavor in Cheddar cheese. *J. Dairy Sci.* 87**,** 1999-2010.

Ayad, E. H. E. , Verheul, A., Wouters, J., and Smit, G. (2000). Application of wild starter cultures for flavour development in pilot plant cheese making. *Int. Dairy J.* 10**,** 169-179. DOI: 10.1016/s0958-6946(00)00041-8.

Ciocia, F., Mcsweeney, P. L., Piraino, P., and Parente, E. (2013). Use of dairy and non-dairy *Lactobacillus plantarum, Lactobacillus paraplantarum* and *Lactobacillus pentosus* strains as adjuncts in Cheddar cheese. *Dairy Sci. Technol.* 93**,** 623-640.

Cogan, T., Beresford, T., Steele, J., Broadbent, J., Shah, N., and Ustunol, Z. (2007). *Invited Review:* Advances in Starter Cultures and Cultured Foods. *J. Dairy Sci.* 90**,** 4005-4021.

Delahunty, C. M., Eyres, G., and Dufour, J. P. (2006). Gas chromatography-olfactometry. *J. Sep. Sci.* 29**,** 2107-2125.

Drake, M. A., and Civille, G. (2003). Flavor lexicons. *Comprehensive reviews in food science and food safety* 2**,** 33-40.

Fernández, E., Alegría, Á., Delgado, S., Martín, M. C., and Mayo, B. (2011). Comparative phenotypic and molecular genetic profiling of wild *Lactococcus lactis* subsp. l*actis* strains of the *L. lactis* subsp. *lactis* and *L. lactis* subsp. *cremoris* genotypes, isolated from starter-free cheeses made of raw milk. *Appl. Environ. Microbiol.* 77**,** 5324-5335. DOI: 10.1128/AEM.02991-10.

Godon, J.-J., Delorme, C., Ehrlich, S. D., and Renault, P. (1992). Divergence of genomic sequences between *Lactococcus lactis* subsp. *lactis* and *Lactococcus lactis* subsp. *cremoris*. *Appl. Environ. Microbiol.* 58**,** 4045-4047.

Kelly, W., and Ward, L. (2002). Genotypic vs. phenotypic biodiversity in *Lactococcus lactis*. *Microbiol.* 148**,** 3332-3333.

Kelly, W. J., Ward, L. J. H., and Leahy, S. C. (2010). Chromosomal diversity in *Lactococcus lactis* and the origin of dairy starter cultures. *Genome Biol. Evol.* 2**,** 729-744. DOI: 10.1093/gbe/evq056.

Lazzi, C., Turroni, S., Mancini, A., Sgarbi, E., Neviani, E., Brigidi, P., and Gatti, M. (2014). Transcriptomic clues to understand the growth of *Lactobacillus rhamnosus* in cheese. *BMC Microbiol.* 14**,** 28.

Meilgaard, M. C., Carr, B. T., and Civille, G. V. (2006). *Sensory evaluation techniques.* CRC press.

Passerini, D., Beltramo, C., Coddeville, M., Quentin, Y., Ritzenthaler, P., Daveran-Mingot, M.-L., and Le Bourgeois, P. (2010). Genes but not genomes reveal bacterial domestication of *Lactococcus lactis*. *PLoS One.* 5**,** e15306. DOI: 10.1371/journal.pone.0015306.

Passerini, D., Coddeville, M., Le Bourgeois, P., Loubière, P., Ritzenthaler, P., Fontagné-Faucher, C., Daveran-Mingot, M.-L., and Cocaign-Bousquet, M. (2013). The carbohydrate metabolism signature of *Lactococcus lactis* strain A12 reveals its sourdough ecosystem origin. *Appl. Environ. Microbiol.* 79**,** 5844-5852. DOI: 10.1128/AEM.01560-13.

Quigley, L., O'Sullivan, O., Beresford, T. P., Ross, R. P., Fitzgerald, G. F., and Cotter, P. D. (2012). High-throughput sequencing for detection of subpopulations of bacteria not previously associated with artisanal cheeses. *Appl. Environ. Microbiol.* 78**,** 5717-5723.

Rademaker, J. L. W., Herbet, H., Starrenburg, M. J. C., Naser, S. M., Gevers, D., Kelly, W. J., Hugenholtz, J., Swings, J., and Van Hylckama Vlieg, J. E. T. (2007). Diversity analysis of dairy and nondairy *Lactococcus lactis* isolates, using a novel

multilocus sequence analysis scheme and (GTG) 5-PCR fingerprinting. *Appl. Environ. Microbiol.* 73**,** 7128-7137. DOI: 10.1128/AEM.01017-07.

Randazzo, C., Pitino, I., De Luca, S., Scifo, G., and Caggia, C. (2008). Effect of wild strains used as starter cultures and adjunct cultures on the volatile compounds of the Pecorino Siciliano cheese. *Int. J. Food Microbiol.* 122**,** 269-278.

Samson, J. E., and Moineau, S. (2010). Characterization of *Lactococcus lactis* phage 949 and comparison with other lactococcal phages. *Appl. Environ. Microbiol.* 76**,** 6843-6852. DOI: 10.1128/aem.00796-10.

Schuch, R., and Fischetti, V. A. (2006). Detailed genomic analysis of the Wβ and γ phages infecting *Bacillus anthracis*: implications for evolution of environmental fitness and antibiotic resistance. *J. Bacteriol.* 188**,** 3037-3051.

Settanni, L., and Moschetti, G. (2010). Non-starter lactic acid bacteria used to improve cheese quality and provide health benefits. *Food Microbiol.* 27**,** 691-697.

Siezen, R. J., Starrenburg, M. J., Boekhorst, J., Renckens, B., Molenaar, D., and Van Hylckama Vlieg, J. E. (2008). Genome-scale genotype-phenotype matching of two *Lactococcus lactis* isolates from plants identifies mechanisms of adaptation to the plant niche. *Appl. Environ. Microbiol.* 74**,** 424-436. DOI: 10.1128/AEM.01850-07.

Stockdale, S. R., Mahony, J., Courtin, P., Chapot-Chartier, M.-P., Van Pijkeren, J.-P., Britton, R. A., Neve, H., Heller, K. J., Aideh, B., and Vogensen, F. K. (2013). The lactococcal phages Tuc2009 and TP901-1 incorporate two alternate forms of their tail fiber into their virions for infection specialization. *J. Biol. Chem.* 288**,** 5581-5590.

Taïbi, A., Dabour, N., Lamoureux, M., Roy, D., and Lapointe, G. (2010). Evaluation of the genetic polymorphism among *Lactococcus lactis* subsp. *cremoris* strains using comparative genomic hybridization and multilocus

sequence analysis. *Int. J. Food Microbiol.* 144**,** 20-28. DOI: 10.1016/j.ijfoodmicro.2010.02.020.

Yarlagadda, A., Wilkinson, M., O'Sullivan, M., and Kilcawley, K. (2014). Utilisation of microfluidisation to enhance enzymatic and metabolic potential of lactococcal strains as adjuncts in Gouda type cheese. *Int. Dairy J.* 38, 124-132.

Acknowledgements

I wish to express my sincere gratitude to Dr. Olivia McAuliffe and Dr. Kieran Kilcawley for their invaluable and constant guidance, and support throughout the course of my PhD studies. I would also like to thank Prof. Gerald Fitzgerald for his encouragement and help over the last 4 years.

A few written words are not enough to convey my thanks and appreciation to so many people but here it goes…

Very special thanks have to go to my parents Charlie and Ellen without whom I would be in a far different place to where I am today. Thank you for keeping the faith and making me strive to be the 'favourite'. To Dermot and Niamh for keeping me grounded…I promise good Christmas presents from now on. To Christine for all of her help and support with the last 100 yards, which like any long race, are the toughest…roll-on November 2015.

To Aidan, your help and assistance has been greatly appreciated over the course of my PhD and for that I will always owe you a cup of coffee. To Rob K. many thanks for all the 'back in the day' stories, bond movies and being a fantastic housemate. To Dan O'S on making a tough 4 years move that bit easier with plenty of advice, 'decent' stories and (albeit more recently) great rock climbing. To Calum thanks for being a great gym buddy and assisting me in the pursuit of 'gains'. To Pete thanks for all the rugby memories (good and bad!) and all the great rugby chat.

My thanks to all of those in Moorepark who have been excellent through the years for sharing their time, expertise and jokes particularly Elaine P., Anil Y., Paul R., Ruairi R., Kiera M., Dan B., Rebecca W., Vincenzo F., Fiona F., Jia Hou (Ariel),., Ewelina S., Vincenzo P., Tom O'C., Andrea B.,

To those who have come and gone Devon, Charlotte, Federica, Viola, Dave R., Rob McC. thanks for all of the great memories.

Special thanks to Joanne Hayes, Fiona Fouhy, Mary Rea, Catriona Guinane, Michelle O'Donnell, Susan Mills, Brian Healy, Paula O'Connor and Vincenzo Fallico for sharing their expertise and knowledge, and always being willing to lend a helping hand.

This research thesis was made possible the Teagasc Walsh Fellowship programme for which I am eternally grateful.