CORA Cork Open Research Archive
Cartlann Taighde Oscailte Chorcaí

Title	Lactobacillus paracasei: an evaluation of the flavour diversification potential of the species through genomics, metabolomics and applications
Authors	Stefanovic, Ewelina
Publication date	2017
Original Citation	Stefanovic, E. 2017. Lactobacillus paracasei: an evaluation of the flavour diversification potential of the species through genomics, metabolomics and applications. PhD Thesis, University College Cork.
Type of publication	Doctoral thesis
Rights	© 2017, Ewelina Stefanovic. - http://creativecommons.org/licenses/by-nc-nd/3.0/
Download date	2026-11-15 18:19:50
Item downloaded from	https://hdl.handle.net/10468/6556



UCC
University College Cork, Ireland
Coláiste na hOllscoile Corcaigh

Lactobacillus paracasei: an evaluation of the flavour diversification potential of the species through genomics, metabolomics and applications

A Thesis Presented to the National University of Ireland, Cork

for the Degree of Doctor of Philosophy

by

Ewelina Stefanovic

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

School of Microbiology, University College Cork, Cork, Ireland

August 2017

Head of the Department: Prof. Gerald Fitzgerald

Research Supervisors:

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

Table of contents

Table of contents ..... **i**

Declaration ..... **vi**

Abstract ..... **vii**

Chapter 1

Literature Review

Advances in the genomics and metabolomics of dairy lactobacilli: A review **1**

1.1 Abstract 2
1.2 Introduction 3
1.3 Genomics of the *Lactobacillus* genus 6
1.3.1 *Lactobacillus delbrueckii* 8
1.3.2 *Lactobacillus helveticus* 10
1.3.3 The *Lactobacillus casei/paracasei* group 10
1.3.4 *Lactobacillus acidophilus* 12
1.3.5 *Lactobacillus rhamnosus* 13
1.3.6 *Lactobacillus plantarum* 15
1.4 Niche adaptability of lactobacilli 17
1.4.1 Horizontal gene transfer (HGT) is the main pathway of niche adaptability in lactobacilli 17
1.4.2 Niche adaptation studies reveal lactobacilli as niche specialists 19
1.5 A genomic perspective on key dairy traits: flavour formation and phage resistance 25
1.5.1 Diverse proteolytic and flavour formation abilities of dairy lactobacilli 25
1.5.2 CRISPR regions of dairy-related lactobacilli 31
1.5.2.1 Applications of CRISPR systems 35
1.6 Genome scale metabolic models and metabolic engineering of *Lactobacillus* species 37
1.6.1 Metabolic engineering as a future application of lactobacilli 42
1.7 Concluding remarks 52
1.8 Acknowledgments 53
1.9 References 54

Chapter 2

Genetic, enzymatic and metabolite profiling of the *Lactobacillus casei* group reveals strain biodiversity and potential applications for flavour diversification **74**

2.1 Abstract 75
2.2 Introduction 76
2.3 Materials and methods 80
2.3.1 Bacterial strains and culture conditions 80
2.3.2 DNA isolation, PCR and sequence analysis of the 16S rRNA gene 80
2.3.3 Pulsed-Field Gel Electrophoresis (PFGE) 81

2.3.4 Phylogenetic analysis 82
2.3.5 Determination of cell envelope proteinase activity 82
2.3.6 Determination of aminopeptidase activities 83
2.3.7 Determination of aromatic aminotransferase activity 84
2.3.8 Determination of glutamate dehydrogenase activity 85
2.3.9 Production of volatile compounds from single amino acid metabolism 85
2.3.10 Head-Space Solid Phase Microextraction Gas Chromatography-Mass Spectrometry analysis (HS-SPME GC-MS) 87
2.3.11 Data processing and compound identification 88
2.3.12 Statistical analysis 89
2.4 Results 90
2.4.1 Origins of *L. casei* group strain bank and confirmation of species 90
2.4.2. Comparative phylogenetic analysis reveals extensive genomic diversity in *L. casei* group strain bank 90
2.4.3 Key enzymatic assays show diverse activities among selected strains 91
2.4.4 HS-SPME GC-MS volatile analysis confirms metabolic diversity of selected strains 92
2.4.5 Two strains with the same genomic fingerprint show different phenotypic characteristics 95
2.5 Discussion 97
2.6 Acknowledgments 107
2.7 References 108

Chapter 3

Strains of the *Lactobacillus casei* group show diverse abilities for the production of flavour compounds in two model systems 140

3.1 Abstract 141
3.2 Introduction 142
3.3 Materials and methods ..... 145+
3.3.1 Bacterial strains 145
3.3.2 Model system 1 (MS1): resting cells in media containing free amino acids 145
3.3.3 Model system 2 (MS2): growing cells in processed curd 146
3.3.4 Gas Chromatography-Mass Spectrometry (GC-MS) analysis of volatiles 147
3.3.5 Chemometric data processing and identification of compounds 148
3.3.6 Statistical analysis 148
3.4 Results 150
3.4.1 Resting cells in amino acid-rich medium show diverse metabolic activities 150
3.4.2 Growing cells in processed curd show limited diversity in volatiles production 152
3.5 Discussion 154
3.6 Conclusion 161
3.7 Acknowledgments 162
3.8 References 163

Chapter 4

Variation of volatile profiles resulting from the choice of extraction and separation techniques - examples from a cheese model system 182

4.1 Abstract 183
4.2 Introduction 184
4.3 Materials and methods 188
4.3.1 Standards for LOD and LOQ determination 188
4.3.2 Cheese model system and sample preparation 188
4.3.3 Method A: HS-SPME GC-MS 189
4.3.4 Method B: HS-Trap GC-MS 190
4.3.5 Data processing, identification of volatile compounds and statistical analysis 190
4.4 Results 192
4.4.1 The analysed methods differ in linearity ranges, LOD and LOQ 192
4.4.2 Method A and Method B generated different volatile profiles from the same set of samples 192
4.4.3 Quantitative comparison of detected compounds in Methods A and B 193
4.4.4 Global variations among the samples differed after analysis by the two methods 195
4.5 Discussion 197
4.6 Conclusion 202
4.7 Acknowledgments 203
4.8 References 204

Chapter 5

Comparative genomic and metabolic analysis of three *Lactobacillus paracasei* cheese isolates reveals considerable genomic differences in strains from the same niche 216

5.1 Abstract 217
5.2 Introduction 218
5.3 Materials and methods 221
5.3.1 Bacterial strains for comparative analysis 221
5.3.2 Identification of strain-specific genes in each of the input genomes 221
5.3.3 Plasmid profiles 222
5.3.4 Minimal media to assess carbohydrate fermentation 222
5.3.5 Carbohydrate fermentation 222
5.3.6 Growth in the presence of heavy metal salts 223
5.3.7 Putrescine production 223
5.3.8 Exopolysaccharide (EPS) production 224
5.4 Results and discussion 225
5.4.1 Genome characteristics of DPC2071, DPC4206 and DPC4536 225
5.4.2 Plasmid-encoded markers suggest a more complex evolutionary route for DPC2071 225
5.4.3 Specificities of carbohydrate utilisation of three cheese isolates 227
5.4.3.1 Diverse carbohydrate utilisation profiles 227
5.4.3.2 Higher numbers of BglG transcriptional regulators were present in the genomes of DPC4206 and DPC4536 229

5.4.3.3 Specific genes for the fermentation of plant derived carbohydrates 230
5.4.4 Genomic content as an indicator of the flavour development potential of the cheese isolates 232
5.4.4.1 Higher number of cysthationine lysase encoding genes explains higher potential for volatile sulfur compounds production in DPC4206 233
5.4.4.2 DPC4206 and DPC4536 have additional genes encoding pyridoxine-5'-phosphate oxidases 235
5.4.5 DPC2071 shows resistance to the toxic effect of cadmium and arsine salts 236
5.4.6 DPC2071 possess unusual arginine metabolism gene 237
5.4.7 CRISPR array content provides evidence of the independent evolution of DPC4206 and DPC4536 238
5.4.8 Cell surface molecules and secreted components 239
5.4.8.1 Exopolysaccharide production was detected in anaerobic conditions 239
5.4.8.2 Specific collage-adhesion encoding genes in DPC2071 suggest the potential for gut colonisation 241
5.5 Conclusion 242
5.6 Acknowledgments 243
5.7 References 244

Chapter 6

Evaluation of the potential of *Lactobacillus paracasei* adjuncts for cheese flavour development and diversification 260

6.1 Abstract 261
6.2 Introduction 262
6.3 Materials and methods 266
6.3.1 Bacterial strains used in cheese manufacture 266
6.3.2 Cheese manufacture and ripening 266
6.3.3 Enumeration of starter and adjunct bacteria in cheeses 267
6.3.4 Cheese compositional and biochemical analysis 268
6.3.5 Free fatty acids (FFA) analysis of cheese lipid extracts by Gas Chromatography with Flame Ionisation Detector (GC-FID) 268
6.3.6 Cheese volatile analysis by Gas Chromatography-Mass Spectrometry (GC-MS) 270
6.3.7 Sensory Affective Evaluation and Ranking Descriptive Analysis of cheeses 271
6.3.8 Statistical analysis 272
6.4 Results 273
6.4.1 Starter and adjunct enumeration showed typical evolution of cheese microbiota 273
6.4.2 The presence or choice of adjunct cultures did not influence gross cheese composition 273
6.4.3 Free amino acid content in cheeses significantly differed in first weeks of ripening 274
6.4.4 The content of FFA in cheese lipid extracts did not significantly differ 275
6.4.5 Month 3 cheeses show higher differentiation in volatile profiles 275
6.4.6 Sensory analysis showed minimal differences in the sensory attributes of cheeses 277

6.5 Discussion 278
6.6 Conclusion 286
6.7 Acknowledgments 287
6.8 References 288

Chapter 7

General Discussion 308

7.1 General discussion 309
7.2 References 319

Appendix 322

Appendix 1 323

Draft genome sequence of *Lactobacillus casei* DPC6800, an isolate with potential to diversify flavour in cheese 323

8.1 Abstract 324
8.2 Genome Announcement 325
8.3 Acknowledgments 328
8.4 References 329

Appendix 2 330

Draft genome sequences of three *Lactobacillus paracasei* strains, members of the non-starter microbiota of mature Cheddar cheese 330

9.1 Abstract 331
9.2 Genome Announcement 332
9.3 Acknowledgments 334
9.4 References 335

Acknowledgments 336

Declaration

This is to certify that the work I am submitting is my own and has not been submitted for another degree, either at University College Cork or elsewhere. All external references and sources are clearly acknowledged and identified within the contents. I have read and understood the regulations of University College Cork concerning plagiarism.

_____

Ewelina Stefanovic

Student number: 113222506

Date: 1.8.2017

Abstract

Strains of *Lactobacillus paracasei* are commonly isolated from numerous and diverse niches, such as dairy products, plant materials and reproductive and gastrointestinal tracts of humans and animals. In cheese, strains of *L. paracasei* belong predominantly to the non-starter microbiota, which is often considered to be immensely important for the development of flavour. In this project, the genetic and phenotypic diversity of *L. paracasei* strains and their application as adjunct cultures to support the development of flavour compounds was assessed. The bank of 310 strains investigated in this study consisted of isolates originating from cheese, yoghurt and sourdough. After genomic profiling, 99 strains were selected for future examination. The phenotypic characterisation included *in vitro* assessment of the key proteolytic enzyme activities, the most important factor contributing to the flavour compound development. The activities of the examined enzymes significantly differed among the analysed strains. Ten strains showing different enzyme activities were selected to compare their ability for flavour compounds production in two cheese model systems. The volatile profiles of the strains differed in both model systems, and according to the all generated results, three strains (DPC2071, DPC4206 and DPC4536) were selected as adjunct cultures for Cheddar cheese manufacture. The cheese analysis showed that although some differences existed, they were minimal and cheeses were of similar flavour characteristics. Finally, to identify and characterise specific genes that may contribute to the overall differentiation of the selected strains, genome sequencing and assembly and comparative genome analysis were performed on the three strains used in the cheese production, and a considerable level of genetic heterogeneity was observed.

Chapter 1

Literature Review

Advances in the genomics and metabolomics of dairy lactobacilli: A review

All of this chapter has been published in Food Microbiology (2016). 61, 33-49
DOI: 10.1016/j.fm.2016.08.009

Ewelina Stefanovic chapter contributions:

- Literature investigation, data collection, major contributor to manuscript preparation

1.1 Abstract

The *Lactobacillus* genus represents the largest and most diverse genera of all the lactic acid bacteria (LAB), encompassing species with applications in industrial, biotechnological and medical fields. The increasing number of available *Lactobacillus* genome sequences has allowed understanding of genetic and metabolic potential of this LAB group. Pangenome and core genome studies are available for numerous species, demonstrating the plasticity of the *Lactobacillus* genomes and providing the evidence of niche adaptability. Advancements in the application of lactobacilli in the dairy industry lie in exploring the genetic background of their commercially important characteristics, such as flavour development potential or resistance to the phage attack. The integration of available genomic and metabolomic data through the generation of genome scale metabolic models has enabled the development of computational models that predict the behaviour of organisms under specific conditions and present a route to metabolic engineering. Lactobacilli are recognised as potential cell factories, confirmed by the successful production of many compounds. In this review, we discuss the current knowledge of genomics, metabolomics and metabolic engineering of the prevalent *Lactobacillus* species associated with the production of fermented dairy foods. In-depth understanding of their characteristics opens the possibilities for their future knowledge-based applications.

Keywords: *Lactobacillus,* dairy, genomic, metabolic engineering

1.2 Introduction

The lactic acid bacteria (LAB) are a group of Gram-positive, non-sporulating, aerotolerant bacteria, with a fermentative metabolism that has lactic acid as the principal final product. The LAB group comprises seven main genera: *Lactococcus, Lactobacillus, Enterococcus, Pediococcus, Streptococcus, Leuconostoc* and *Oenococcus* (O'Sullivan et al., 2009). The practical importance of the organisms within this group is unquestionable as they find application in industry, food and health-related fields. In the food industry, LAB are widely used in the production of fermented dairy, meat and vegetable products as well as in wine and sourdough production (Pfeiler and Klaenhammer, 2007, O'Sullivan et al., 2009). In addition, the production of antimicrobials or bacteriocins by certain species of the LAB has prompted their use as biopreservative agents in foods (Cleveland et al., 2001, Cotter et al., 2005, De Vuyst and Leroy, 2007). Other members of the LAB group exhibit health benefits and are often used as probiotics in the treatment of intestinal infections, inflammatory bowel disease and allergy development (Ljungh and Wadstrom, 2006). Members of the LAB group have also been suggested for use in mucosal vaccines as delivery vehicles for vaccine antigens (Bermudez-Humaran et al., 2011, Villena et al., 2011, Wyszynska et al., 2015). The wide variety and number of applications of the LAB raises the need to correlate industrially and clinically important features with genomic information to examine the possibilities for exploitation of their metabolic potential, thus improving their use in biotechnological and health-related applications. The complete and draft genomes of many LAB species are available in online databases (Genome Online Database, www.gold.jgi.doe.gov, NCBI database www.ncbi.nlm.nih.gov/genome, Ensemble

Genomes database www.ensemblgenomes.org, etc.) and they present valuable sources of information regarding genetic diversity and the metabolic potential of strains. In addition, state-of-the-art developments in genomics and metabolomics provide the tools for a more 'knowledge-based' approach to selection of desirable cultures for application in industry (McAuliffe, 2017).

LAB are phylogenetically closely related, but the number of predicted protein-coding genes in the LAB varies between 1700 and 2800 (Makarova et al., 2006). Genomic studies of members of the LAB have confirmed the overall trend of minimisation of genomes, which is in close agreement with the transition to nutritionally rich environments. Nevertheless, some gene families were expanded by gene duplication or acquisition of paralogous genes *via* horizontal gene transfer (HGT) (Makarova et al., 2006). Based on the analysis of the genomes of 12 LAB species it was concluded that the core LAB genome, comprising orthologous genes conserved in all analysed genomes (Collins and Higgs, 2012), consists of 567 genes, mostly encoding translation, transcription and replication processes, but 41 of the genes were uncharacterised and 50 had only general functions predicted. This study also identified two core genes exclusive for LAB, the products of which are LysM (peptidoglycan-binding) domain and the highly conserved LaCOG01237 with no known domains, but based on its localisation, it is probably involved in modification of tRNA (Makarova et al., 2006).

The genus *Lactobacillus* comprises a diverse group of bacteria currently consisting of more than 200 species and subspecies (Sun et al., 2015a) that share the common features of other LAB, including low GC content, acid tolerance and conversion of sugars to lactic acid as one of the main end products of metabolism. Species of lactobacilli are present in various environments such as plants, fermented food

products (dairy, meat, wine), and both the human and animal gastrointestinal tracts. Their ability to ferment milk, meat and plant material presents the basis for their artisanal and industrial usage (Sun et al., 2015a). Apart from this, strains of *Lactobacillus* are well known for their probiotic properties (Lebeer et al., 2008).

This review aims to present recent findings related to the genus *Lactobacillus*, with a particular emphasis on strains commonly used in the production of fermented dairy foods. Genomic features of the main dairy species will be discussed, including their remarkable niche specialisation. Advancements in our knowledge through genomic analysis of key attributes of dairy species will also be reviewed. Finally, innovations in the applications of genome scale metabolic models and metabolic engineering, highlighting new possibilities in exploitation of strains of *Lactobacillus,* are also discussed.

1.3 Genomics of the *Lactobacillus* genus

Due to their importance in various biotechnological and health-related applications, there has been a growing interest in exploring the genomic features of the genus *Lactobacillus,* which is the largest and most diverse genus of LAB (Broadbent et al., 2012). *Lactobacillus* genomes range in size from 1.23 Mbp (*L. sanfranciscensis)* to 4.91 Mbp (*L. parakefiri)* (Sun et al., 2015a). Species of this genus are present in dairy products *(L. delbrueckii* ssp. *bulgaricus, L. helveticus*), human and animal gastrointestinal tracts (*L. acidophilus* and *L. gasseri*) or in a variety of niches (*L. plantarum, L. pentosus, L. brevis,* and *L. paracasei*) (Smokvina et al., 2013). The first genome of the *Lactobacillus* genus sequenced was *L. plantarum* WCFS1 (Kleerebezem et al., 2003) followed by *L. johnsonii* NC533 (Pridmore et al., 2004) and *L. acidophilus* NCFM (Altermann et al., 2005). These studies revealed some interesting genomic features of the *Lactobacillus* genus, such as lifestyle adaptation islands in *L. plantarum* WCFS1, lack of general biosynthetic pathways in the probiotic strain *L. johnsonii* NC553 and unique structures called potential autonomous units (PAU) in *L. acidophilus* NCFM, all of which triggered further investigation and comparison with newly sequenced strains of the same species. Currently (July 2016), there are 214 *Lactobacillus* genome sequencing projects available in public databases (www.ncbi.nlm.nih.gov).

The pangenome (or supragenome) is considered as the full set of all genes within a selected genome set (species, genera or higher taxonomic groups) (Medini et al., 2005, Collins and Higgs, 2012). The size of the pangenome generated for *Lactobacillus* and associated genera of LAB reaches almost 45000 gene families, while 73 genes mainly responsible for cell growth and replication make up the core

genome (Sun et al., 2015a). In a study based on the features of 20 complete *Lactobacillus* genomes representing 14 species whose genomes ranged from 1.8 to 3.3 Mbp, the number of proteins within these genomes was between 1721 and 3100 (Kant et al., 2011). The estimated size of the pangenome of the *Lactobacillus* genus consists of almost 14000 proteins, while the core genome consists of 383 orthologs (Kant et al., 2011). This number is higher than the 141 core genes reported in the study of Claesson et al. (2008), who used more strict criteria and took into account only 12 completely sequenced *Lactobacillus* genomes. Over 100 out of 383 genes of the *Lactobacillus* core genome were organised in operon-like clusters that are conserved in other related Gram-positive bacteria (Kant et al., 2011). Among 41 genes specific for *Lactobacillus*, 13 were predicted to code for ribosomal proteins, and 13 were annotated as hypothetical (Kant et al., 2011). Taken together, comparative genomic studies of lactobacilli confirmed the overall trend observed in other LAB, which is loss of ancestral genes and minimisation of genomes, as well as acquisition of genes by HGT as a response to adaptation to the primary habitat of these bacteria (Makarova et al., 2006).

The main species of *Lactobacillus* used as starter cultures for the production of fermented dairy products are *L. delbrueckii* and *L. helveticus*, but more recently, a group of non-starter lactobacilli has attracted growing attention due to their contribution to the quality and characteristics of the final products. This group includes *L. casei*, *L. paracasei*, *L. rhamnosus* and less often *L. plantarum*. Additionally, dairy products can be used as "carriers" of probiotic strains, such as *L. acidophilus* and *L. rhamnosus*. Therefore, general information regarding genomics of these most important dairy-related lactobacilli is presented in Table 1, and specific genomic features of these species will be discussed in more detail.

1.3.1 *Lactobacillus delbrueckii*

From the perspective of the dairy industry, *Lactobacillus delbrueckii* contains two industrially important subspecies: subspecies *bulgaricus* and subspecies *lactis*. Of the 22 genome sequences available for these two subspecies, five are complete sequences. While *L. delbrueckii* ssp. *bulgaricus* is widely used in the production of yoghurt, subspecies *lactis* is used primarily as a starter in the manufacture of cheeses like Emmental, Grana Padano and Parmigiano Reggiano (El Kafsi et al., 2014). The core genome of the three *L. delbrueckii* ssp. *bulgaricus* strains (2038, ATCC 11842 and ATCC BAA-365) consists of 1276 genes, with the genomes of strains 2038, ATCC 11842 and ATCC BAA-365 consists of 211, 150 and 166 unique genes, respectively (Hao et al., 2011). An alignment of the three genomes revealed two duplicated segments flanking the predicted replication terminus, but strain 2038 has a unique 8.5 kbp region between the duplication regions, which could be the reason for the bigger genome size (1.87 Mbp compared to 1.86 Mbp ATCC 11842 and ATCC BAA-365). This region is most likely inherited from an ancestor, but lost in the other two strains, probably due to their independent evolution from strain 2038 (Hao et al., 2011).

A genome analysis of sequenced *L. delbrueckii* strains showed that the average GC and GC3 content (GC at codon position 3) in coding sequences (CDSs) is approximately 52 % and 65 %, respectively (El Kafsi et al., 2014), which is in agreement with a previously reported higher GC content in *L. delbrueckii* ssp. *bulgaricus* compared to other lactobacilli (van de Guchte et al., 2006). Higher GC content is a sign of rapid ongoing evolution in these species (O'Sullivan et al., 2009). In both subspecies, decay and inactivation of superfluous genes was evident, indicating an evolutionary trend towards adaptation to the dairy environment. A

deeper insight into the genomics of these subspecies revealed some interesting genetic differences. Firstly, it was shown that the size of the ssp. *bulgaricus* genomes is smaller compared to ssp. *lactis* (1.8 Mbp and 2 Mbp, respectively). However, the number of CDS did not differ considerably between the two subspecies, as it varied in range from 1333-1783 for subspecies *bulgaricus* to 1593-1721 for subspecies *lactis*. Comparison of the core proteomes of five ssp. *lactis* and five ssp. *bulgaricus* strains surprisingly revealed quite similar sizes of core proteomes and significant overlapping of these. The overall core proteome consists of 989 proteins, with 65 proteins specific for ssp. *lactis* and 25 proteins specific for ssp. *bulgaricus.* The majority of the 65 specific ssp. *bulgaricus* proteins have unknown functions, while those of known function are mainly membrane transporter-associated proteins. The 25 specific ssp. *lactis* proteins have mainly known functions, involved in carbohydrate and amino acid metabolism. For both subspecies, fragments of other subspecies-specific genes could be found as pseudogenes, implying that differential loss of genes caused subspecies divergence. Another important finding of the extensive genomic analysis is re-classification of strain ND02, which was designated as ssp. *bulgaricus* but confirmed to be ssp. *lactis*, not only due to the larger genome but also due to the higher number of insertion sequences (IS). Besides that, it was previously shown that *L. delbrueckii* subspecies can be distinguished based on the number of *Eco*RI sites in their 16S rRNA sequences, where ssp. *lactis* possesses one, and ssp. *bulgaricus* has two restriction sites (Giraffa et al., 1998). The detailed analysis of 16S rRNA of strain ND02 showed it did not contain two specific restrictions sites, adding an argument to its re-classification as ssp. *lactis* (El Kafsi et al., 2014).

1.3.2 *Lactobacillus helveticus*

Lactobacillus helveticus represents an important starter for the production of Swiss-type and long-ripened Italian cheeses (Broadbent et al., 2011, Giraffa, 2014). Apart from the dairy environment, *L. helveticus* strains are present in fermented plant and meat materials as well as the gastrointestinal and urogenital tracts of humans and animals and their probiotic activity is confirmed (Taverniti and Guglielmetti, 2012, Strahinic et al., 2013). While the complete genome sequences of eight strains are currently available, a total comparative genomic study of this species has not been performed to date, and information regarding the core, pan and specific genomes is not currently available, to the best of our knowledge. Strains sequenced to date originate from various fermented dairy products, such as koumis, sour milk, kurut, or they were used as industrial starters. Genome sizes vary from 1.87 to 2.38 Mbp, with a GC content of 37 %, and the number of genes ranges between 1743 and 2540.

1.3.3 The *Lactobacillus casei/paracasei* group

The taxonomic status of *L. casei* is still a matter of much debate (Smokvina et al., 2013) as molecular studies have implied that the majority of *L. casei* strains are more related to *L. casei* ATCC 334 (also named *L. paracasei*) than to the official type strain *L. casei* ATCC 393 (Dellaglio et al., 2002). Because of this uncertainty, the information available for both *L. casei* and *L. paracasei* will be reviewed together here. The members of this group have been isolated from dairy and plant materials (cheese, wine, pickle, silage) (Toh et al., 2013) and reproductive and gastrointestinal tracts of humans and animals (Cai et al., 2009). In the cheese industry, they are used as adjunct cultures for development of desired flavour (Milesi et al., 2010, Van Hoorde et al., 2010). Besides application in fermented food production, members of

this group are well known for their probiotic characteristics (Herias et al., 2005, Ya et al., 2008). Such a diverse range of sources and broad ranging possible applications makes this group one of the best explored species within the *Lactobacillus* genus with eight and seven genome sequences completed for *L. casei* and *L. paracasei,* respectively, and 27 and 46 draft genome sequences available for *L. casei* and *L. paracasei,* respectively. Genome sizes range from 2.38 Mbp for *L. paracasei* ssp. *tolerans* DMS20258 and 3.27 Mbp for *L. casei* Lbs2, with an average GC content of 46.5 %. Analysis of the draft sequences of 12 strains of *L. casei* of different origins (dairy, plant and human) along with five fully sequenced genomes have determined that the size of the *L. casei* pangenome is 3.2 × the average genome size, consisting of 1715 core and 4220 accessory genes (Broadbent et al., 2012). Another comparative study (Yu et al., 2015) performed on 12 draft *L. casei* genomes revealed 806 novel regions larger than 500 kbp harbouring both hypothetical proteins and mobile genetic elements in these strains compared to the seven complete genomes. This suggested that the *L. casei* pangenome expands with every new sequenced genome and potential for environmental adaptation within the species increases (Yu et al., 2015). Similarly, when 37 genomes of *L. paracasei* were analysed, 1800 core and 4200 accessory genes were detected (Smokvina et al., 2013). A common feature of all 37 analysed genomes of *L. paracasei* is a cluster involved in the conversion of branched-chain α-keto acids into branched-chain fatty acids important for maintenance of the colonic epithelium. This gene cluster is unique for *L. paracasei*, implying its acquisition through HGT (Smokvina et al., 2013). Pangenome analysis revealed the ability of *L. paracasei* to utilise a broad range of carbohydrates. In total, 74 sugar utilisation cassettes were detected 15 of which belonged to the core genome. These cassettes were localised on two genomic islands (Smokvina et al.,

2013), structures usually connected with the environmental adaptation (described in details below).

1.3.4 *Lactobacillus acidophilus*

Taxonomically, *Lactobacillus acidophilus* is part of a larger complex comprising several species: *L. acidophilus*, *L. amylovorus*, *L. crispatus*, *L. gallinarum*, *L. gasseri*, and *L. johnsonii* (Berger et al., 2007, Ramachandran et al., 2013). Strains of *L. acidophilus* are often used in dairy products as probiotics and as flavour contributing strain in certain dairy products, such as yoghurt, sweet acidophilus milk and cheese (Buriti et al., 2005, Ong et al., 2007, Ejtahed et al., 2011). The genome of *Lactobacillus acidophilus* NCFM was the first *L. acidophilus* to be sequenced (Altermann et al., 2005). Presently, 16 strains of this species have been sequenced, with three complete genomes available. Genomes range in size from 1.25 to 2.05 Mbp, with GC content of 34.7 %. Although phenotypic and biochemical characterisation of strains show a certain level of diversity, genotypic analysis indicates less variation within genomes of this species (Ramachandran et al., 2013, Stahl and Barrangou, 2013, Bull et al., 2014). In a recent study reporting the genome sequences of *L. acidophilus* strains isolated from yoghurt (Iartchouk et al., 2015), the alignment of the three sequenced genomes (FSI4, NCFM, and La-14) confirmed a high level of genome similarity for these strains at the DNA level. Similarly, alignment of La14 and NCFM showed extremely high similarity between these two strains and synteny with ATCC 4769 (Stahl and Barrangou, 2013). Strain 30SC was initially designated as *L. acidophilus*, but unlike other strains of this species, it possesses 2 plasmids and has higher GC content (38 %) (Stahl and Barrangou, 2013). After detailed phylogenetic analysis of its genome, it was re-classified as *L. amylovorus* (Bull et al., 2014).

Intraspecific diversity of 33 *L. acidophilus* strains was examined by whole genome multi locus sequence typing (wgMLST), at 1864 loci defined in the *L. acidophilus* NCFM genome sequence (Bull et al., 2014). It was found that the core genome comprised 1815 genes, which makes up to 97.4 % of *L. acidophilus* NCFM loci. A number of commercial strains analysed in this study showed a narrow window of variation, unlike the type strains analysed where a somewhat higher level of variation in loci was detected. When a pairwise comparison of selected isolate sequences was performed with the NCFM strain, it confirmed that the genetic variation in the core genome was predominantly the effect of single nucleotide polymorphism (SNP). Pairwise analysis also revealed partial evidence of gene decay, during which phage, mucus-binding and sugar metabolism genes were lost. Similar findings were observed at the phenotypic level where no significant differences between the commercial or culture collection strains was observed, following analysis by API 50CHL. An interesting finding of this study is that all investigated isolates showed no evidence of extrachromosomal DNA, such as plasmids, and no evidence of an active phage, again confirming the stability of *L. acidophilus* genomes. However, three prophage remnants termed Potentially Autonomic Units (PAU) discovered in NCFM genome (Altermann et al., 2005) and a novel region with phage related functions showed variable presence in other *L. acidophilus* isolates. While PAU1 was present in all analysed isolates, PAU2 and PAU3 were present in commercial isolates, but variably present in culture collection isolates (Bull et al., 2014).

1.3.5 *Lactobacillus rhamnosus*

Lactobacillus rhamnosus is present in various dairy products, such as cheese and yoghurt, but also in human cavities and gastrointestinal tract (GIT) (Douillard et al.,

2013, Kant et al., 2014). In dairy products, it is mainly present as part of the non-starter LAB (NSLAB) in Italian cheese varieties (Gobbetti et al., 2015), and there is evidence of its positive effect on flavour development in these products (Sgarbi et al., 2013, Innocente et al., 2016). However, its main application is as probiotic cultures (Tuo et al., 2013), often administered through fermented dairy products. To date, 102 genome sequences have been elucidated, with the completed sequences of six strains available. The size of the genomes range from 2.52 Mbp for strain MTCC 5462 up to 3.41 Mbp for strain CRL1505, and the average GC content is 46.7 %. General genomic features of this species were determined based on 100 sequenced strains of various origin (cheese, yoghurt, vaginal cavity, oral cavity, intestinal tract, abscess, blood, clinical isolates) mapped according to the reference strain *L. rhamnosus* GG. The number of shared genes between these 100 strains and strain GG ranged from 87 to 100 % (Douillard et al., 2013). The pangenome analysis based on the complete or draft genomes of 13 strains, originating from various environments (milk, human airways, feces, dairy starter, infected dental pulp, Cheddar cheese and gut biopsy), estimates a total of 4893 genes, 1.6 × the average size of a *L. rhamnosus* genome (Kant et al., 2014). Pangenome studies show that, in general, the rate of increase of the size of the pangenome slows down with every additional genome being sequenced (Kant et al., 2014). As the pangenome curve of *L. rhamnosus* reaches a plateau at about 5000 genes, it is predicted that with only a few more additional genomes of strains from different origins would be sufficient to reach total genome variability of the species (Kant et al., 2014). The core genome of *L. rhamnosus* is estimated to encode 2095 genes, or approximately 43 % of the pangenome. There are at least 75 genes present only in *L. rhamnosus* species, and the majority of these are hypothetical proteins followed by membrane transporters,

transcriptional regulators and glycosyl-transferases. The dispensable genome, which contains genes present in two or more strains (Medini et al., 2005), of *L. rhamnosus* is estimated to contain 2798 genes, and the number of unique (strain-specific) genes is 855, which is approximately 30 % of the dispensable genome. Most of the dispensable genes in the *L. rhamnosus* pangenome are annotated as hypothetical and it remains unknown what proportion of these would actually encode functional proteins (Kant et al., 2014).

1.3.6 *Lactobacillus plantarum*

Lactobacillus plantarum is present in many ecological niches ranging from vegetables, meat, dairy products and gastrointestinal tract. Apart from a prominent role in fermentations such as sourdough (Corsetti and Settanni, 2007), strains of this species are present in dairy fermentations and non-starter flora (Settanni and Moschetti, 2010, Gobbetti et al., 2015). Besides that, they are well known for their probiotic characteristics (Siezen and van Hylckama Vlieg, 2011). To date, 114 genome sequences are publically available, with 18 completely sequenced genomes. The genome of this species is one of the largest in the *Lactobacillus* group, with a size of approximately 3.4 Mbp, and a GC content of 44.4 %. In an extensive study, 185 isolates from different environments were phenotypically characterised, and based on the observed phenotypic diversity, a set of 42 candidates were selected for genomic analysis (Siezen et al., 2010). The core genome of *L. plantarum* was found to comprise 2050-2200 genes. Approximately 120 fully conserved genes were unique to *L. plantarum*. Many of the unique genes encode hypothetical proteins, while some genes encode functions that could be used for phenotyping. The two candidates are a conserved cluster for tartrate and sulfur uptake and metabolism, which are associated with plant habitats (Siezen et al., 2010). The reference genome

WCFS1 itself has over 50 genes not found in any of the other selected strains isolated from different environments. Most notable are three gene clusters encoding exopolysaccharide, a putative macrolide and a non-ribosomal synthesised hybrid peptide-polyketide, all of which take part in the interaction with environment. They were most likely acquired in a recent evolutionary event due to their GC content, suggesting adaptations necessary for survival in a specific niche (Siezen et al., 2010). Apart from these 50 genes, all other strains were estimated to lack between 9 % and 20 % of genes present in the reference genome, WCFS1. These genes are mainly organised in functional gene clusters, or cassettes as parts of operons and they encode prophages, restriction/modification systems, exopolysaccharide, bacteriocin and non-ribosomal peptide biosynthesis and carbohydrate utilisation components and are located on genomic islands (described in details in the next section) (Siezen et al., 2010, Siezen and van Hylckama Vlieg, 2011).

1.4 Niche adaptability of lactobacilli

The widespread dissemination of members of the lactobacilli in different environments testifies to their extraordinary niche adaptability. Lactobacilli are present in grass and on plant material, in dairy products, on human skin, in the mouth, intestine and in the female reproductive system (Claesson et al., 2007), habitats with many contrasting environmental conditions (temperature, pH value, available nutrients, and competing microorganisms). Comparative genomic analysis has revealed that adaptation to such highly variable environments is a result of genome evolution and the genetic basis for niche specialisation appears to be the result of eliminating anabolic systems that are not needed through adaptation to nutritionally rich habitats, such as milk. On the other hand, in all LAB, including lactobacilli, duplications of genes coding for transporters and metabolism of carbohydrates, amino acid transporters and peptidases occurred, further enhancing the ability of these species to live in nutrient-rich environments (Fig. 1a) (Makarova and Koonin, 2007, Mayo et al., 2008, Douglas and Klaenhammer, 2010).

1.4.1 Horizontal gene transfer (HGT) is the main pathway of niche adaptability in lactobacilli

Although gene loss and acquisition, which are the principal events resulting in niche adaptation, occur in different ways, HGT *via* bacteriophages, transposons and other mobile elements appears to be an especially dominant force of adaptation to novel environments in *Lactobacillus* species (Broadbent et al., 2012), and it is responsible for various genome rearrangements (Rossi et al., 2014). Such events have made the LAB amenable to adaption to different habitats, including milk and other food matrices, plant material, and GIT. Transposons and plasmids present the main

mechanism of gene exchange that occurs amongst different taxonomic groups that do not possess strictly controlled restriction/modification systems (Rossi et al., 2014). Both niche specialists and generalists have undergone multiple genetic changes which have led to restriction or broadening of the possible habitats in which these strains could survive.

Apart from the traditional classes of mobile genetic elements (plasmids and prophages), structures acquired by the host bacteria through HGT comprising mobile elements and genes contributing to the ability of the host to adapt to specific conditions of habitat, are known as genomic islands (GI) (Bellanger et al., 2014). The first record of "lifestyle adaptation" islands in *Lactobacillus* was in the genome of *L. plantarum* WCFS1, where numerous genes involved in sugar transportation and metabolism are grouped together in a region characterised by lower GC content (41.5 %) than the rest of the genome (44.45 %), suggesting recent acquisition by HGT (Kleerebezem et al., 2003). Apart from strong overrepresentation of genes involved in energy metabolism, regulatory proteins coordinating sugar metabolism are also present on GI (Molenaar et al., 2005). In strain *L. helveticus* DPC4571, a number of amino acid metabolism genes along with lipid biosynthesis genes were also identified in a region characterised with higher GC content (42 % compared to 37 % in the rest of the genome) and insertion sequences flanking this region suggest a recent transfer of this GI (Callanan et al., 2008). One of the GI of *L. casei* BL23 carries genes for catabolism of *myo*-inositol, a cyclic polyol not commonly metabolised by LAB and potentially present in degrading plant material (Yebra et al., 2007, Cai et al., 2009). Genomic islands of *L. casei* ATCC 334 encode hypothetical proteins and transcriptional regulators, sugar transporters and metabolic enzymes and are characterised by high prevalence of insertion sequences,

recombinases, and integrases with higher GC content supporting their recent acquisition and a heterologous origin (Cai et al., 2009). The 26 genomic islands of *L. rhamnosus* ATCC 53103, isolated from the human gut, include six carbohydrate utilisation gene clusters, which seem to have secured the survival of the strain in a less nutritionally rich environment, such as the human intestine (Toh et al., 2013). These examples of different genes present in GI confirm their importance for adaptation and survival in specific environmental conditions.

1.4.2 Niche adaptation studies reveal lactobacilli as niche specialists

Niche specialists can be described as strains that are able to live in a limited number of habitats, while niche generalists have the capacity to populate various environments. Genome analysis of dairy specialists show that these strains have an abundance of sugar transportation, proteolysis and amino acid transportation encoding genes, some of which have undergone duplication as they enable the organism to uptake nutrients from the rich milk environment (Makarova et al., 2006). On the other hand, substantial gene decay has been confirmed in some lactobacilli, such as in the dairy *L. casei* strains, which have more than 120 CDS absent. As a result, these strains have improved their ability to survive in the dairy niche but have a reduced capacity for survival in other niches (Cai et al., 2009). In the genomes of dairy LAB, more than 10 % of coding genes are present only as pseudogenes (Zhu et al., 2009), which are non-functional due to frameshifts, nonsense mutations, deletions or truncations (O'Sullivan et al., 2009). For example, the dairy isolate *L. helveticus* DPC4571 is reported to have 217 pseudogenes, while *L. bulgaricus* ATCC 11842 carries a staggering 533 pseudogenes coding for proteins involved in regulating amino acid and nucleotide metabolism and bile salt hydrolysis (Callanan et al., 2008, O'Sullivan et al., 2009). In contrast, species mainly present in

the gut, such as *L. acidophilus*, *L. gasseri*, *L. reuteri* and *L. johnsonii* have either no pseudogenes or a low abundance of pseudogenes, which is likely the genetic basis supporting survival of these species in the gut environment (O'Sullivan et al., 2009).

Efforts have been made to find at least a partial correlation between genome characteristics and niche for such a versatile group as *Lactobacillus*. The study of O'Sullivan et al. (2009) compared the genomes of 11 LAB (ten *Lactobacillus* and one *Streptococcus thermophilus*) arising from different sources. In total, nine genes were identified as niche determinative as they insured survival in the gut or dairy environments. These genes were grouped into four classes that could be used as niche-specific genes for gut and dairy LAB: sugar metabolism, the proteolytic system, restriction/modification systems and bile salt hydrolysis. In contrast to this study, Kant et al. (2011) did not reveal any niche-specific genes in a study that analysed 20 genomes of 14 different *Lactobacillus* species. The possible cause of this observation is that the isolation source does not always correspond to the actual habitat, but rather a transient habitat (Fig. 1b), as some species, like *L. plantarum* can be found in various environments (Kant et al., 2011).

Correlation between gene loss and niche adaptation was examined by growing nine *L. casei* strains from various isolation sources in chemically defined amino acid media supplemented with one of the substrates representing plant, gut or dairy habitats (Broadbent et al., 2012). The two cheese specialists had the most restricted substrate profiles, with no genes for inulin, sucrose or cellobiose utilisation present in their genomes, while the other strains used a higher number of different substrates, with corn silage isolates growing on 26 different substrates (Broadbent et al., 2012). In the study of Smokvina et al. (2013), niche affinity of *L. paracasei* was examined through utilisation of carbon sources as growth factors for a set of strains with

diverse origins: plant, mammalian and dairy. The analysis revealed the clustering of seven out of the 16 dairy isolates that could be considered as niche specialists, which had smaller genomes compared to the others (2.8 Mbp average), limited numbers of sugar cassettes and an absence of genes involved in utilisation of plant-derived sugars. This was expected, as the spectrum of sugars in the dairy environment is narrow with lactose dominating. On the other hand, no clear clustering pattern was revealed for plant and mammalian isolates. Plant isolates originate from a broad range of ecosystems that differ in environmental and nutritional conditions, while mammalian isolates come from the gut where they are exposed to constantly changing surroundings due to the presence of food and other microorganisms, and this complicates their precise grouping (Smokvina et al., 2013).

Lactobacilli occupy habitats that differ considerably in environmental conditions. The dairy niche bacteria have to be robust enough to survive manufacture and storage conditions encountered during industrial production. In the gut, strains need to be able to survive in the presence of other intestinal microbiota and resist bile salts and other harsh conditions found in the gut (Senan et al., 2014). A genome-scale study based on genes involved in stress responses of the *L. helveticus* strains MTCC 5463 (probiotic strain isolated from a vaginal swab of a healthy volunteer, Senan et al. (2015)) and DPC4571 (a dairy isolate, Callanan et al. (2008)) gave an insight into genes responsible for adaptation to various environments (Senan et al., 2014). When comparing these two genomes for the ability of the strains to survive in a bile-rich environment, it was shown that the MTCC 5463 genome exhibited multiple coding sequences for bile salt hydrolase (bsh). However, the cheese starter DPC4571, adapted to a dairy niche, displayed a total lack of active *bsh* genes. The probiotic strain is exposed to other gut microbiota and in constant competition for successful

colonisation and available nutrients. In order to survive in these conditions, it carries a higher number of starvation-induced genes. By contrast, while the dairy strain possessed some genes for starvation proteins, such as phosphate starvation inducible stress-related protein, it was deficient in the gene for the carbon starvation protein CstA. Both strains carried a substantial number of genes that allow response to heat and cold shock, but the molecular chaperones were far more prevalent in the probiotic genome (Senan et al., 2014). Another study performed on *L. helveticus* strains confirmed loss of genes encoding mucus-binding proteins from strains adapted to the milk environment, but confirmed their maintenance in probiotic strain R0052, where they are essential for survival and residence of the strain in the gut (Cremonesi et al., 2012).

Another noteworthy conclusion regarding niche adaptability was made when genome sequences of two strains, *L. helveticus* DPC4571 and *L. acidophilus* NCFM, were compared. The remarkable level of identity of 98 % for 16S rRNA sequences was observed. Additionally, 75 % of ORFs in DPC4571 were found in NCFM, which confirmed a close relationship between the two strains that inhabit significantly different environments (milk and gut). The genetic differences between these two strains were examined and they explained the genetic basis for niche specialisation. It was shown that the dairy strain lacked many genes that were retained in the probiotic strain, such as PTS systems, cell wall-anchoring proteins and the already mentioned mucus-binding proteins (Callanan et al., 2008).

In the previously mentioned study that analysed 100 *L. rhamnosus* strains, interesting observations regarding niche adaptability and clustering were made. Most dairy isolates clustered together, while intestinal and probiotic strains shared similarities with other human isolates. When both the phenotypic and genomic data

of each strain were joined, two geno-phenotypes were identified. Firstly, the strains in group A were characterised by the absence of SpaCBA pili, lactose, maltose and rhamnose metabolism all of which point to dairy adaptation. Secondly, group B strains were bile resistant, pili possessing and L-fucose utilising, all characteristics important for intestinal tract survival. Although isolates of the same origin could be found in both groups, cheese isolates mainly belonged to group A, while intestinal isolates belonged mainly to group B. Intestinal isolates in group A may have originated from the consumption of food and represent rather a transient flora, while isolates from group B represent typical GIT residents (Fig. 1b). Interestingly, vaginal and oral isolates shared geno-phenotype A, which suggests a connection with dairy isolates (Douillard et al., 2013). Another study attempted to link genotypes and carbohydrate utilisation profiles of 65 *L. rhamnosus* strains isolated from diverse habitats, such as human, baby and goat feces, cheese and fermented milk (Ceapa et al., 2015). Genomic fingerprinting was performed by amplified fragment length polymorphism (AFLP) genotyping, and 11 genotypic groups were determined. Although not seen as a strict rule, strains of the same origin clustered together. Some clusters contained strains from various origins, indicating that these strains frequently change habitats (Fig. 1b). Conversely, some clusters had members of a single isolation niche, such as dairy. Following on from this, 25 isolates that represent all 11 clusters obtained by AFLP were tested for the carbon sources they could potentially use. Based on 72 carbon sources, three metabolic groups were determined, with group A including strains that could use plant derived carbohydrates, group B including strains with no ability to use lactose and group C containing strains that could use various carbohydrates. Although group B had no ability to use lactose, some strains isolated from cheese did belong to this group,

where they were present as non-starter flora and had a role in proteolysis in the later stages of ripening. Interestingly, there was no direct correlation between metabolic groups and niche isolation, but strains coming from the same AFLP cluster appeared in the same metabolic group. This work again confirms that origin of isolation gives only an indication of potential metabolic capacity of the strain, but other approaches also have to be employed to fully understand strain fitness. For example, *L. rhamnosus* strain HN001 is present as a cheese isolate, but it has the ability to use 53 different carbon sources, which contradicts the general tendency of niche specialists to use a more narrow range of carbohydrates indicating that this strain was most probably very recently introduced into cheese environment. On the other hand, strain ATCC 53103 (GG) which originated from the intestine, belongs to a metabolically specialist group, possibly because it was transferred from dynamic environment such as GIT to more stable industrial habitat, which may have led to the metabolic simplification (Ceapa et al., 2015).

Finally, the effect of niche adaptation could be seen even within different dairy products. In the multi locus sequence typing (MLST) study of 11 housekeeping genes in 245 *L. helveticus* isolates from natural fermented products, particular branches of isolates could be associated with the dairy product from which they originated (koumiss group, tarag group and coumiss-tarag group). These results suggest that even ecological niches representing different dairy environments may impact evolution of *L. helveticus* strains because genetic relationships are generally correlated with the ecological niches (Sun et al., 2015b).

1.5 A genomic perspective on key dairy traits: flavour formation and phage resistance

The successful application of lactobacilli in the industrial environment depends on the robustness of selected strains and their ability to contribute to the desirable properties of the final product. Apart from their metabolic potential which affects the technological and organoleptic characteristics of dairy products, the ability of dairy lactobacilli to combat phage attacks which are frequent in dairy plants also contributes to the overall quality of product. Thus, a genomic perspective of these two features of dairy related lactobacilli will be discussed in more details.

1.5.1 Diverse proteolytic and flavour formation abilities of dairy lactobacilli

Flavour formation in dairy products is the result of a complex network of processes which ends in specific combinations of flavour compounds and aroma development. Three major processes contribute to flavour development: glycolysis, lipolysis and proteolysis (Van Kranenburg et al., 2002, Smit et al., 2005, Settanni and Moschetti, 2010). Glycolysis refers mainly to the metabolism of lactose and citrate. While lactose, the primary milk sugar, is mostly metabolised to lactic acid, a proportion of it can be converted to flavour compounds such as diacetyl, acetoin, acetaldehyde, or acetic acid, depending on the organism (Van Kranenburg et al., 2002). Certain organisms also have the ability to metabolise citrate. Citrate is generally metabolised to pyruvate, which can be further metabolised to acetoin in the final product (Medina de Figueroa et al., 2001, Mortera et al., 2013). Lipolysis in fermented milk products arises mainly from the activity of microbial lipolytic enzymes (Collins et al., 2003). Esterases hydrolyse hydrosoluble ester chains between 2 and 8 C atoms, and lipases are more active on longer ester chains (10 C atoms). Free fatty acids contribute to

cheese flavour, particularly short and intermediate-chain fatty acids, which represent the starting molecules for catabolic reactions resulting in the production of numerous flavour and aromatic compounds (Collins et al., 2003). Of all the metabolic processes responsible for flavour development in dairy products, proteolysis is considered the most important and complex one, affecting texture, hardness, elasticity and the overall flavour of the fermented product (Savijoki et al., 2006). The proteolysis cascade starts with casein degradation by cell envelope proteinases (CEP, Prt). The peptides released in this processes are then transported in the cell, where peptidases with varying specificities cleave them, releasing amino acids. These amino acids are the substrates for various metabolic reactions, with aminotransferases being the first enzymes in the subsequent catabolic cascade. Diverse and numerous aromas are released in these reactions (aldehydes, ketones, carboxylic acids and volatile sulfur compounds) (Marilley and Casey, 2004). In this section, the genomics of the components of proteolytic system of *Lactobacillus* will be discussed, as proteolysis represents a critical process in flavour development in dairy products.

Cell envelope proteinases (CEPs) are multi-subunit, cell wall-associated proteinases and their main role during growth in milk is degradation of casein into smaller peptides (Sun et al., 2015a). The importance of surface proteinases is made clear in studies that showed that knock-out strains lack the ability to grow in milk (Mayo et al., 2010).

In an extensive study performed on the genomes of 213 *Lactobacillus* and associated genera, intriguing diversity in CEP characteristics was revealed (Sun et al., 2015a). In total, genes for 60 CEPs were identified and presence of genes for CEPs was highly correlated with phylogenetic clades. Three different anchoring mechanisms

were observed: a SLAP domain (S-layer anchoring domain) responsible for non-covalent interactions was present, particularly in the *L. delbrueckii* sub-clade; a LPXTG motif for covalent linkage to peptidoglycan and a derivative of the LPXTG motif. In thirteen cases, no anchoring domain for CEP was identified as sequences were terminated exactly before the typical start of the anchoring domain sequence. Multiple alignments indicated the sequences of these 13 CEPs differ from other CEPs along the entire length of the protein. Besides this, the possibilities of various domain combinations in the CEPs enable a diversity of potential substrates to be utilised, resulting in a range of final products, which could contribute to improvement of dairy products flavour (Sun et al., 2015a).

The vast majority of LAB have only one CEP, but for certain strains of *L. helveticus,* it has been confirmed through multiplex PCR analysis that at least four different proteinases exist (Broadbent et al., 2011) and four *prt* genes were described in the genome of *L. helveticus* CNRZ32 (Broadbent et al., 2013). The presence of a higher number of proteinases with different substrate and cleavage specificities could explain the efficiency of the *L. helveticus* proteolytic system. CEPs have different and complimentary properties and some strains could have acquired additional genes because they provide an adaptive advantage regarding milk protein hydrolysis (Genay et al., 2009). In the study by Broadbent et al. (2011), 51 *L. helveticus* strains were tested for presence of *prt* paralogs. The distribution of *prt* genes varied among *L. helveticus* strains and the most abundant gene was *prtH3,* which contradicts the study by (Genay et al., 2009) who found that *prtH2* was in fact a ubiquitous gene in *L. helveticus* strains. The reasons for this contradiction are that sequences for *prtH4* were not available, and *prtH3* gene from DPC4571 strain was described as an allele of *prtH2* (Broadbent et al., 2011). From the dairy industry perspective, the diverse

proteinase gene content in *L. helveticus* may be a crucial factor in determining the function and behaviour of these strains with regard to desired flavour formation (Broadbent et al., 2011).

The correct maturation of CEP depends on the presence of the maturation proteins, PrtM. For instance, while *L. helveticus* CNRZ32 has 2 *prtM* paralogs designated as *prtM* and *prtM2*, in other analysed *L. helveticus* strains *prtM* was found only in strains that possessed *prtH*, and *prtM2* was encoded in genomes of all tested strains. It has been proposed that *prtM* is needed for activation of *prtH*, and *prtM2* is responsible for folding and activation of other *prt* paralogs (Broadbent et al., 2011). On the other hand, no *prtM* gene for this protein was found in any of the four completely sequenced *L. delbrueckii* strains (Liu et al., 2012). However, the foldase protein (PrsA) involved in maturation of extracellular proteinase and folding and stability of subtilisins in *Bacillus subtilis* was detected. PrsA might be involved in maturation of PrtB, as PrsA from four *L. delbrueckii* strains were homologous with known PrtM proteins (Liu et al., 2012).

Peptides released by the activity of CEP are transported by various transport systems inside the cell, where they are cleaved by peptidases of different activities, releasing amino acids. Several studies that took into consideration various LAB genomes concluded that the general peptidases (PepN, PepC) and dipeptidyl-peptidase PepX were widely distributed among *Lactobacillus*, including species of interest in dairy fermentation (Cai et al., 2009, Liu et al., 2010). A closer look suggests that PepN and PepX are encoded by single genes, but genes for other peptidases, such as PepC/E and PepO were detected as multiple copies in strains belonging to species generally seen as important for dairy industry, enabling higher adaption in habitat abundant in proteins and peptides (Cai et al., 2009).

The diversity in peptidase content is observed on the same species level, where strains differ in numbers of peptidases and transport system components. Upon analysis of four fully sequenced genomes of *Lactobacillus delbrueckii* (ATCC 11842, BAA-365, 2038 and ND02), strain ND02 possessed the highest number of proteinase and peptidase genes, as well as the highest number of peptidase and amino acid transport systems. Intracellular peptidases showed some differences between the four strains, such as three unique peptidases in strain ND02. In the case of strain 2038, two cell surface peptidases EnlA and Pep-D4 were present as complete genes, indicating that this strain has a more powerful proteolytic capability and potentially produces more free amino acids than the other strains (Liu et al., 2012). All four sequenced strains possessed two complete Opp systems, but they differed in numbers and organisation of substrate binding protein OppA. The highest number of OppA genes was found in the industrial strain 2038 and their products enable transport of different oligopeptides (Liu et al., 2012).

The next step in the protein degradation cascade is the metabolism of free amino acids, following which a large number of flavour compounds arise. Aminotransferases are the first enzymes in the cascade, transferring amino groups from amino acids to α-keto acids, most often α-ketoglutarate. In a comparative study of enzymes involved in amino acid metabolism contributing to generation of flavour compounds in 21 genomes of different LAB species, (12 of which were lactobacilli), a homolog of the *bcaT* gene, coding for branched-chain aminotransferase activity, was present in all *Lactobacillus* strains considered as important in dairy production, while a larger number of homologs for the *araT* gene, coding for aromatic aminotransferase activity, were usually present (Liu et al., 2008). The distribution of amino acid metabolising enzymes amongst starter and NSLAB including the species

discussed in this review, were compared by Gobbetti et al. (2015), and it confirmed the diversity of the metabolic capability of lactobacilli and underlined the importance of genomic analysis as part of a knowledge-based approach to strain selection.

Cysteine and methionine are precursors for the production of volatile sulfur compounds (VSCs) which are important flavour compounds that are found in many cheese varieties. The metabolism of sulfur containing amino acids is complex as multiple alternative metabolic pathways exist (Mayo et al., 2010). One of the enzymes involved in metabolism of methionine is cystathionine gamma lyase (CGL), which was found in several *L. casei* strains isolated from cheese and milk (Irmler et al., 2008). Two variants of the gene encoding CGL shared 81 % of similarity and were named *ctl1* and *ctl2*. Homologs of *ctl1* and *ctl2* were found in other LAB: *L. helveticus, L. bulgaricus L. rhamnosus* and *S. thermophilus*, but they were not present in three publicly available genomes of *L. casei* (ATCC 334, Zhang and BL23) and it is likely that these strains uptake sulfur-containing peptides and amino acids from the environment (Irmler et al., 2009). Analysis of nucleotides upstream from a *ctl* gene cluster found an ORF encoding for a putative transposase, supporting the possibility of horizontal transfer of the cluster to *L. casei* strains. The gene cluster forms an operon important in cysteine biosynthesis, as its expression was downregulated when L-cysteine is added to the medium (Bogicevic et al., 2012). Furthermore, when these strains were used in cheese production, significantly higher levels of VSC were detected at the end of ripening (Bogicevic et al., 2013).

Glutamate dehydrogenase (GDH) is an enzyme that acts as a cofactor for aminotransferase function, as it enables recycling of α-ketoglutarate, the receptor of the amino group during transamination. When genomes of 12 species of *Lactobacillus* were analysed, the presence of a *gdh* gene was confirmed only in *L.*

plantarum WCFS1 and *L. salivarius* UCC118 (Liu et al., 2008), which agrees with the strain dependency of *gdh* presence and higher prevalence in natural strains commonly found in cheese manufacture (Tanous et al., 2002). However, the majority of *L. casei*, *L. rhamnosus* and *L. plantarum* genomes possess the *gdh* gene (Gobbetti et al., 2015), but no *gdh* gene was found in any of the sequenced *L. delbrueckii* strains (Liu et al., 2012, Gobbetti et al., 2015). Nevertheless, two genes encoding proteins homologous to aspartate aminotransferase were found in *L. delbrueckii* and which could potentially catalyse the formation of glutamate from 2-oxoglutarate (α-ketoglutarate) and L-aspartate (Liu et al., 2012).

Collective data from genomic analysis of dairy-related strains present a first step in knowledge based strain selection. The insight into the number and characteristics of genes of interest enables strategic choice of cultures for dairy manufacture. Besides that, selection of strains with variable key enzyme presence and activities opens the possibilities for development of products with diverse flavour and broadens the overall portfolio offered to the final customer.

1.5.2 CRISPR regions of dairy-related lactobacilli

Bacteriophages present a serious problem in dairy industry affecting continuity of quality for the final product as they affect survival of starter and adjunct cultures in the fermentation process. Although huge efforts are made to prevent and control phage levels, phage infections regularly cause disruptions in production and product downgrading (Marco et al., 2012).

Several mechanisms of phage resistance were previously described for lactic acid bacteria and they include prevention of phage adsorption, blocking the entry of phage DNA, cutting phage nucleic (restriction/modification systems) acid and

abortive infection (Garneau and Moineau, 2011). However, recently, a new system that enables effective resistance to phage attacks was discovered, and it was shown that this system was almost universally present in bacteria, including LAB. CRISPR (clustered regularly interspaced short palindromic repeats), together with CRISPR-associated genes (*cas*) form a bacterial immune system against foreign DNA, such as phage or plasmids (Barrangou and Horvath, 2012). The typical CRISPR locus, located behind the leader sequence, contains a string of DNA repeats and spacers, which represent short sequences corresponding to foreign DNA inserted between two repeats (Deveau et al., 2010). The efficient defence from foreign DNA attack involves the incorporation of short sequences of foreign DNA in CRISPR loci (acquisition) (Fig. 2a). In the event of foreign DNA being present in the cell, these short sequences are transcribed into small interfering RNAs, called CRISPR RNA (crRNA), which guide multifunctional protein complexes to recognise and cleave matching foreign DNA (Fig. 2b) (Barrangou and Horvath, 2012).

Two genes, *cas1* and *cas2*, are regularly present in CRISPR-Cas systems, and they are involved in the acquisition process (Barrangou, 2013). Based on the signature genes which confer interference, three types of CRISPR-Cas systems are well described. Type I systems have *cas3* as the signature gene, which encodes an endonuclease involved in the cleavage of DNA. Another feature of this type is the Cascade complex, participating in processing of crRNA and recognition of target DNA. The signature gene of Type II systems is *cas9*, which encodes a protein important for the crRNA synthesis and target DNA cleavage. Specificity of Type II systems is trans-activating CRISPR RNA (tracrRNA) that hybridizes to crRNA and enables its maturation by endoribonuclease RNAse III. Type III systems are defined by the signature gene *cas10* and they are mechanistically diverse, with IIIA systems

cleaving DNA and IIIB systems cleaving RNA molecules (Barrangou, 2013, Selle and Barrangou, 2015). Besides these three systems, novel types (IV, V and VI) were discovered more recently (Wright et al., 2016).

In LAB, eight different families of CRISPR loci were found and these families did not correlate with phylogeny of LAB indicating their independent evolution from other elements on the chromosome. The analysis of CRISPR loci at the level of the LAB showed that highly similar loci were found in distant genera and species. This could be explained by HGT and indeed, these loci have different GC content compared to the rest of the host genome. Interestingly, the comparison of CRISPRs of two closely related species, *L. helveticus* and *L. casei*, showed that they belong to different families, once again confirming the high level of variability of these regions (Horvath et al., 2009).

In the analysis of 213 genomes of *Lactobacillus* and associated genera, 137 CRISPR loci were found in 63 % of all analysed genomes. All three types of systems were found in *Lactobacillus* and the size of loci varied between 2 and 135 spacers. Type II systems were found to be the most prevalent (36 % of analysed genomes). In addition, novel Type II systems with heterogeneous *cas9* sequences were detected, and their potential use could be as tool for specific DNA cleavage in genome editing in both prokaryotes and eukaryotes (Sun et al., 2015a).

CRISPR profiles of 100 *L. rhamnosus* strains were generated by spacer oligotyping, a method firstly described by Kamerbeek et al. (1997), and a considerable level of strain variety was revealed (Douillard et al., 2013). Additionally, in certain cases, correlation between CRISPR loci and specific niche was observed. In total, 24 spacers were identified from both plasmids and phage DNA. Spacers that corresponded to phages belonged *to L. rhamnosus* phages or *L. casei* phages. The

study defined two general geno-phenotypes (discussed above) and the CRISPR locus profiles were substantially different in these two groups (Douillard et al., 2013). A comparative study of CRISPR in *Lactobacillus delbrueckii* ssp. *bulgaricus* that took into consideration 33 strains showed that these strains possessed either Type II or Type III CRISPR systems (Urshev and Ishlimova, 2015). However, in the genome of recently sequenced strain CFL1 both CRISPR types (II and III) were present simultaneously (Meneghel et al., 2016).

As described previously, *L. casei* represents a highly genomically diverse species of lactobacilli, while *L. acidophilus* is characterised by remarkable genome stability. These differences are also apparent in the comparison of CRISPR systems in the two species. The CRISPR spacers of *L. casei* show a high level of variability and homology to *Lactobacillus* phages and plasmids. It was noted that strains isolated from commercial cheeses possess higher numbers of spacer sequences highlighting potential interactions with phage in the dairy manufacturing environment (Broadbent et al., 2012). Conversely, CRISPR loci of *L. acidophilus* show striking stability. When CRISPRs of La-14 and NCFM were compared, a high level of identity was observed, and similar sequences were found in strain ATCC 4796 (Stahl and Barrangou, 2013). In addition, CRISPR loci of 20 *L. acidophilus* strains also showed stability and uniformity (Bull et al., 2014). This may suggest that *L. acidophilus* has not recently encountered phage attack, as this species does not encode for an active phage and there is no recent report of validated phages of this species. The fact that *L. acidophilus* is resistant to phage attack supports its wide and successful commercial application (Bull et al., 2014).

1.5.2.1 Applications of CRISPR systems

Analysis of the CRISPR loci present in strains provides the evidence of previous phage interaction and opens possibilities for enhancing phage resistance of industrial strains. A potential strategy would be to improve the CRISPR systems both in resistance level and spectrum, which would contribute to the robustness of the industrial strains. This could be achieved by selecting CRISPR mutants after repeated exposure to different phages selected from a diverse collection. Mutants with novel spacers with high homology to conserved phage sequences could be used in culture rotation schemes of dairy strains. Another benefit of mutant selection, as described by Barrangou and Horvath (2012), is the development of tagging system for proprietary strains.

Due to their hypervariability in spacer regions, CRISPR loci could be used in strain typing studies, as nearly identical strains could be distinguished, and this typing has already been performed for pathogens such as *Mycobacterium tuberculosis* or *Yersinia pestis*, as well as for industrially important LAB (Barrangou and Horvath, 2012). High level of diversity in CRISPR loci represents a basis for comparative analysis of strains originating from different habitats, and it may be used in phylogenetic relationship studies (Horvath et al., 2009).

Genome editing represents a novel and elegant approach that has revolutionised the idea of genetic engineering. This approach was inspired by the mechanism of action of Type II CRISPR systems, where crRNA introduces double-stranded DNA breaks (DBS) of invading DNA (Jiang and Marraffini, 2015). DBS and targeted genome editing was successfully performed by adapting the Type II CRISPR system from *Streptococcus pyogenes* (Jinek et al., 2012). For the genome engineering process, two components have to be present in the cell: Cas9 nuclease that makes the DBS

and a guide RNA, a chimeric molecule combined of crRNA and tracrRNA that leads the Cas9 to a specific DNA site (Fig. 2c). The DNA break can be followed by non-homologous end joining which induces indels, or homology-directed repair that introduces site-specific insertion from DNA donor templates (Sander and Joung, 2014). This simple and highly specific approach has moved the boundaries of genetic and biochemical research, and it is almost ideal for genome editing applications due to its efficiency and affordability (Selle and Barrangou, 2015).

1.6 Genome scale metabolic models and metabolic engineering of *Lactobacillus* species

While comparative genomic studies represent the starting point for advancing our understanding of the evolution, diversity and metabolism of LAB, systems biology approaches, which combine mathematical modelling with 'omics' information, can predict how cells will behave and what modifications could be made to improve their performance (King et al., 2015). An example of this are genome-scale metabolic models (GSMM), which represent a catalogue of all the metabolic reactions and their associations in a single organism from gene to final metabolic process based on merging information about gene functions, the biochemical reactions in which the product is involved and theoretical background (Teusink et al., 2011). GSMMs connect the genotypic and phenotypic data and combine them with transcriptomic, proteomic and metabolomics data (Steele et al., 2013). Some of applications of GSMM constructed for LAB include design of metabolic engineering experiments, detection of differences between the strains and testing of characteristics of potential probiotic strains (Vinay-Lara et al., 2014). From the perspective of the dairy lactobacilli, the development of such models could be of immense importance for desired product design (Steele et al., 2013) and metabolic engineering projects (Gaspar et al., 2013) (Fig. 3).

The metabolic network of an organism is based on genomic information, and this network connects the information of genes and the metabolic reactions they are involved in (Lewis et al., 2012). After detailed revision and correction of the (genome-scale) metabolic model, it is then transformed to a stoichiometric matrix, which is a mathematical representation of metabolic reactions. The purpose of this

step is to convert GSMM to a computational one (O'Brien et al., 2015). Constraint-based reconstruction and analysis (COBRA) models are the most widely used in GSMM analysis (Lewis et al., 2012). Flux Balance Analysis (FBA) is the oldest, most basic and commonly used COBRA method (Orth et al., 2010, Lewis et al., 2012, O'Brien et al., 2015) for simulating GSMM. Detailed explanation of how FBA operates can be found in Orth et al. (2010). Flux variability analysis (FVA), introduced by Mahadevan and Schilling (2003), modifies the FBA approach as it considers the effect of metabolic uncoupling. FVA determines, for each reaction in the model, the range of possible fluxes that correspond to experimental values of constraints (Smid and Hugenholtz, 2010).

Lactococcus lactis was the first LAB to have a genome-scale model constructed (Oliveira et al., 2005), followed by *L. plantarum*, (Teusink et al., 2006) and *Streptococcus thermophilus* (Pastink et al., 2009) and most recently, *L. casei* (Vinay-Lara et al., 2014, Xu et al., 2015). Here, we will review the most important findings of models designed for some species of *Lactobacillus*.

The GSMM of *L. plantarum* WCFS1 was used to compare a traditional view of ATP production from lactate and acetate and ATP production based on the constraints approach when experimental constraints were applied. The traditional approach has certain disadvantages as it takes into account lactate and acetate production in other metabolic processes which do not contribute to ATP yield, like amino acid or citrate metabolism. After comparison of ATP production in both approaches, the same result was obtained in both cases, meaning that the effects of amino acid and citrate metabolism were not crucial. Additionally, the model identified catabolic reactions such as transamination of aromatic and branched-chain amino acids to generate ATP. These reactions are seen as a major factor in flavour development, but have not been

previously connected with ATP production. Further on, the model attempted to assess the effect of uncoupling on metabolic capacities. FVA was used to calculate the spectrum of flux values consistent with the experimental constraints and showed higher flexibility of the flux ranges for the uncoupled energy production and consumption. However, FBA was not able to correctly predict *L. plantarum* biomass production, as it did not take into account inefficient lactate production. FBA predicted higher growth, as it detected lactate production as incompatible with optimised growth. In reality though, *L. plantarum* produces lactate and tends to utilise a route that is less efficient even under limited energy conditions, and this event cannot be predicted by FBA, which proposed higher yield as a result of mixed acid fermentation (Teusink et al., 2006).

The study by Vinay-Lara et al. (2014) compared metabolic networks from two *L. casei* strains that are fully sequenced, ATCC 334 and 12A. FBA was used to analyse the properties and capabilities of both models. Both tested strains have similar amino acid requirements- branched-chain and aromatic amino acids and arginine are essential. It is most likely that the rich environment (cheese and corn silage) that these strains were isolated from reduced the need for synthesising all amino acids. Although models initially did not predict glutamate as an essential amino acid, excluding this amino acid from the culture medium significantly reduced the growth of ATCC 334 and resulted in no growth for 12A. However, in both metabolic models glutamine can be converted into glutamate, and the experimental studies suggested that this interconversion of glutamine to glutamate results in low yields of synthesised glutamate, thus explaining why glutamate is needed even in the presence of glutamine. A correction of the metabolic pathway was possible in the case of ATCC 334, but fixing the inconsistency in 12A was not successful, and the model

was not able to determine the strain's requirements for glutamate. Carbohydrate utilisation analysis of these strains once again confirmed the hypothesis of gene decay during adaptation to nutrient rich environments. Strain 12A, isolated from corn silage (Cai et al., 2007) possesses an ABC transporter for uptake of raffinose and enzymes needed for pullulan and panose degradation, sugars frequently present in plant material. On the contrary, ATCC 334, a cheese isolate, lacks these genes as they are most likely redundant in the dairy environment. Interestingly, the metabolic model for strain 12A shows that all the genes for converting *myo*-inositol to glyceraldehyde-3-phosphate are present. *Myo*-inositol can be used as phosphate storage molecule in plants. Although the majority of LAB cannot use this sugar as carbon source, strain 12A has all the genes needed for conversion of *myo*-inositol, but this metabolic pathway is not active in 12A probably due to regulatory effects (Vinay-Lara et al., 2014). In other *L. casei* models it was shown that, *in silico* growth of *L. casei* LC2W was improved by *myo*-inositol under aerobic conditions, suggesting that this strain could utilise energy sources that seemed inappropriate under anaerobic conditions (Xu et al., 2015).

A genome-scale metabolic model of *L. casei* LC2W was used for the analysis of the oxygen effect on flavour compound synthesis and three new *in silico* knockout targets were selected for acetoin production. In *L. casei* LC2W, the main precursor of flavour compounds is α-acetolactate. Acetoin and diacetyl are produced from α-acetolactate by acetolactate-decarboxylase or through non-enzymatic processes. Although acetoin could accumulate in LC2W in both aerobic and anaerobic conditions, production of diacetyl was dependent on oxygen and it was possible to maintain diacetyl production at a high level with the increase of oxygen uptake. Additionally, FBA suggested three new *in silico* knockout targets for acetoin

production: dihydrofolate reductase, methylen-tetrahydrofolate dehydrogenase and glycerol-phospho-transferase (Xu et al., 2015).

Regarding the flavour potential of LAB, a completely different approach was recently proposed. As seen, GSMM contain numerous gaps which cannot always be completed. Although there are many known pathways involved in flavour formation, the overall process of flavour development is highly complex. Compounds that are often seen as flavour contributors are products of amino acid metabolism: alcohols, aldehydes and acids, and especially sulfur compounds, products of methionine metabolism (Curioni and Bosset, 2002, Smit et al., 2005, Yvon, 2006). Reverse pathway engineering (RPE) (Liu et al., 2014) takes small molecules as a starting point and looks for enzymatic or chemical reactions that can track these compounds back to the known precursors. This method was used in LAB to predict so far unknown reactions in metabolic pathways by combining retrosynthesis and genomic information. To confirm that the proposed approach is correct, the relatively well-known pathway of leucine degradation in LAB was tested in the model. Not only were the main branches confirmed, but it also suggested a novel route of generating 3-methyl-butanoic acid, one of the most important flavour compounds of leucine metabolism. This novel route starts with the transamination product of leucine, α-keto-isocaproate, which is further reduced to α-hydroxy-isocaproate. The second step suggests formation of 3-methyl-butanoic acid from α-hydroxy-isocaproate, and the related reaction found in the database was a lactate oxidation reaction catalysed by lactate-2-monoxygenase (LOX), so it was assumed that LOX could possibly catalyse oxidation of α-hydroxy-isocaproate. Broader activity of LOX seems to be dependent on the amino acid at position 95 and it could be obtained if alanine in position 95 was mutated to glycine (detailed explanation in Liu et al. (2014)). The RPE method

also revealed a non-enzymatic reaction of converting α-keto-isocaproate to 2-methyl-propanal, and this reaction connects valine and leucine catabolism. Regarding the methionine degradation, RPE discovered an enzymatic reaction responsible for the conversion of methanethiol to dimethyl-sulfide (DMS), using DMS as an input. Enzymes homocystein-S-methyl-transferase, methionine synthase and thiol-S-methyl-transferase were proposed using the bioinformatics approach. The prediction of novel reactions using RPE opens up new possibilities for metabolic engineering. For example, hydroxy-isocaproate is often seen as an off-flavour in cheese products, but the proposed conversion to the flavour compound 3-methyl-butanoic acid could be implemented in novel strategies for production of flavour by utilising off-flavours as precursors (Liu et al., 2014).

1.6.1 Metabolic engineering as a future application of lactobacilli

A vast amount of knowledge on genetics and metabolism of LAB opened the door for implementation of LAB in novel biotechnological applications (Gaspar et al., 2013). Application of LAB is not limited only to classical food fermentation and the use of LAB as cell factories is expected to increase (Gaspar et al., 2013). LAB are characterised by limited biosynthetic capacity and metabolic versatility and their physiology is relatively simple. They are characterised by relatively small genomes (2-3 Mbp), fast growth, high sugar uptake rates and less high-level control systems, all of which make them suitable candidates for metabolic engineering (Papagianni, 2012, Gaspar et al., 2013). Genetic engineering made possible the production of molecules not natively present in the host, but also enabled engineering of native genes (Keasling, 2012). Genetic engineering proved successful in the development of strains producing recombinant proteins and small chemicals, but development of tools that exceed genetic engineering is needed, as some molecules are synthesised

in multiple reactions (Bution et al., 2015). Metabolic engineering summarizes previous knowledge regarding cell metabolic features and it uses molecular tools to deliberately change cellular metabolism for the purpose of the efficient production of target molecules (Bution et al., 2015) (Fig. 3). However, the host cell needs to meet several requirements to ensure efficient metabolic engineering occurs. Host cells should be genetically stable, not interfering with heterologous genes on the introduced vector, and have optimal traits for industrial applications. Apart from these, genomic information can help in the choice of host, as new pathways can induce stress response and impede gene expression (Keasling, 2012).

Metabolic engineering of lactic acid bacteria presents a novel approach for re-routing metabolic reactions in LAB so specific and desired compounds are produced in higher amounts. Several different types of molecules can be produced by LAB as cell factories: lactic acid, flavour compounds (diacetyl, acetaldehyde), sweeteners (L-alanine, mannitol, sorbitol, xylitol), exopolysaccharide, vitamins etc. (Papagianni, 2012). Historically, the first attempt of engineering of LAB was oriented towards improving production of the butter aroma compound diacetyl in *Lactococcus lactis*. Subsequently, many other studies expanded the species of LAB that were subject to engineering as well as the types of molecules produced. Several recent review articles (Papagianni, 2012, Gaspar et al., 2013, Mazzoli et al., 2014) give detailed information about achievements in production of industrially important compounds in LAB. Production of food ingredients, commodity compounds, vitamins and ethanol are thoroughly reviewed with methods of engineering and future perspectives anticipated. Besides this, metabolic engineering is used as a tool for improvement of adherence and immunomodulatory properties of probiotic strains (described and reviewed in Yebra et al. (2012)). While most of results come from

Lactococcus lactis as most widely used LAB, novel information comes from *Lactobacillus* species as well. Here we review studies performed on strains of *Lactobacillus* spp. mainly associated with dairy food.

Lactic acid is used as a preservative and flavour enhancing agent by the food industry, and also in cosmetic and pharmaceutical industries (Papagianni, 2012). In addition, L-lactic acid is used as the starting material in the production of biopolymers (Gaspar et al., 2013). Unlike chemical synthesis, which often leads to racemic mixture of L- and D-lactic acid, microbial fermentation can be optimised for production of a single enantiomer (Gaspar et al., 2013). The L-isomer is a preferred for two reasons: D-isomer is not metabolised in humans and has a toxic effect and L-isomer polymerises which is important in polymers production (Kyla-Nikkila et al., 2000, Papagianni, 2012). The initial attempts to influence lactic acid production in lactobacilli date in 1990's, when the enhancement of L-lactic acid was achieved by the inactivation of *ldhD* in *L. helveticus* (Bhowmik and Steele, 1994), but the overexpression of *ldhL* in *L. plantarum* did not cause an increase of L-lactic acid synthesis, although increased activity of L-LDH was observed (Ferain et al., 1994). More recently, selective L-lactate production was tested in *L. helveticus* CNRZ32 and two approaches were used (Kyla-Nikkila et al., 2000). The promoter of the *ldhD* gene was deleted in the construct GRL86 while in the other construct, GRL89, the structural gene of *ldhD* was replaced with an additional copy of the structural gene of *ldhL*. Both constructs produced only L-lactic acid in amounts that were on the level of total lactate produced by the wild type strain and no difference in growth profiles for either construct was observed compared to the wild strain. Additionally, the L-lactic acid production phase of mutant strains was prolonged compared to the wild strain (Kyla-Nikkila et al., 2000).

Ethanol represents an important biofuel and the high demand for renewable energy sources puts efficient ways of ethanol production in focus (Mazzoli et al., 2014). Although many bacteria have low ethanol tolerance, some species of LAB, especially lactobacilli are relatively tolerant to high concentration of alcohols (Mazzoli et al., 2014). Initial efforts to enhance ethanol production were focused on the overexpression of heterologous genes encoding pyruvate decarboxylase (*pdc*) and alcohol dehydrogenase (*adh*), the enzymes responsible for conversion of pyruvate to ethanol. When *pet* operon, which carries *pdc* and *adh* genes from *Zymomonas mobilis* (Gram-negative bacteria) was used for the transformation of *L. casei* 686, the recombinant strains showed more than a two-fold increase in ethanol production (Gold et al., 1996). In a later study (Nichols et al., 2003), the *pet* operon was modified for expression in Gram-positive bacteria and several strains of *L. plantarum* and *L. casei* were transformed. After glucose fermentations were carried out, some engineered strains showed higher ethanol production compared to the parental strains, but lactic acid was detected as a major metabolic product (Nichols et al., 2003). In the study of Liu et al. (2006), *pdc* gene from Gram-positive bacteria *Sarcina ventriculi* (*Spdc*) was expressed in *ldh* deficient *L. plantarum* TF103, which accumulated pyruvate. Three different promoters and native *Spdc* 5' flanking sequences were fused with *Spdc* gene and introduced in T103. All constructs produced higher amounts of ethanol than the control carrying an empty vector, but they also produced significant amounts of lactate and the level was higher than in the control (wt) strain (Liu et al., 2006).

Sorbitol is a sugar alcohol largely used in the food industry as a sweetener (Gaspar et al., 2013). It is poorly absorbed in small intestine and as it has low calorie value, is used in diabetic appropriate foods (Ladero et al., 2007), but also as a softener and

texturing agent (Yebra et al., 2012). An attempt to construct sorbitol-producing LAB was performed by introducing the *gutF* gene coding for sorbitol-6-phosphate dehydrogenase, into the *lac* operon of *L. casei.* The strain with the integrated *gutF* was named BL232 and the expression was controlled as in other *lac* genes. Additionally, a L-lactate dehydrogenase (*ldhL)* knockout of BL232 was constructed, and designated as BL233. Resting cells of both of these strains produced sorbitol from glucose, and the *ldhL* knockout showed higher production of sorbitol compared to BL232. It was proposed that *ldhL* inactivation leads to a higher NADH/NAD+ ratio and the cell uses this for the sorbitol production (Nissen et al., 2005). In further studies, metabolic engineering of *L. casei* led to a strain that could produce sorbitol without consequent uptake after glucose exhaustion, by introducing a mutation in the sorbitol-specific phospho-transferase system. Sorbitol producing *L. casei* were constructed through a series of transformations of strain BL232: deletion of *ldh1* gene encoding the main lactate dehydrogenase (BL251) followed by deletion of *gutB* gene (BL283) involved in transport of sorbitol and subsequent mutation of the mannitol-1-phosphate dehydrogenase (*mtlD)* gene (BL300). While mutant BL251 used sorbitol after glucose consumption, BL283 was not able to transport sorbitol and levels of sorbitol did not drop after glucose exhaustion. To avoid synthesis of mixed polyols (sorbitol and mannitol, as occurred in the study of Nissen et al. (2005)), a gene encoding mannitol-1-phosphate dehydrogenase was inactivated (BL300) and this knockout strain did not produce mannitol, and sorbitol production was doubled compared to BL283. In addition, the resting cells of BL300 were able to produce sorbitol from lactose in 1 % supplemented MRS, especially at pH 5.5 and 4.75, but this conversion was less efficient than the conversion of glucose.

Additionally, BL300 cells were able to produce sorbitol as a sole polyol from whey permeate, a by-product of the dairy industry (De Boeck et al., 2010).

In order to obtain *L. plantarum* producing sorbitol, a different approach was used. In the genome of *L. plantarum* NCIMB8826, two genes for the enzyme sorbitol-6-phosphate dehydrogenase (*srlD1* and *srlD2*) were present. The two *srlD* coding regions were overexpressed in transformed *L. plantarum* strain VL103 which is lactate dehydrogenase deficient. High sorbitol-6-phosphate dehydrogenase activities as well as sorbitol levels were detected in the overexpressing strains VL103, while no activity could be detected in the wild-type and VL103 strains harbouring the empty vector, used as a control strain. The deficiency in LDH was essential and LDH-positive control did not produce sorbitol under any of conditions examined (Ladero et al., 2007).

Succinic acid is a starting block in synthesis of biodegradable plastic (Babu et al., 2013) and can be used as a food additive (Beauprez et al., 2010). In a study by Tsuji et al. (2013), production of succinic acid was examined in the previously described lactate-dehydrogenase deficient strain *L. plantarum* VL103. Three enzymes involved in succinic acid production: pyruvate carboxylase (PC), phospho-enol-pyruvate carboxykinase (PEPCK) and malic enzyme (ME) were overexpressed in this strain, and all transformants showed increased activity of the corresponding enzyme, up to 2.4 fold in the case of PC. However, although PC overexpression was the most effective for succinic acid production in *L. plantarum*, a mutant with PEPCK enzyme overexpressed, exhibited a higher specific growth rate, compared to the two others, and seemed a better candidate for LAB succinic acid production, as PC overexpression was effective but slowed down the growth rate. Additionally, combined levels of succinic acid production were observed in mutants displaying

overproduction of the two enzymes and the co-expression of PC and PEPCK increased succinic acid yield and biomass (Tsuji et al., 2013).

Engineered *L. casei* were used to increase the production of diacetyl and acetoin from whey permeate (Nadal et al., 2009). These two compounds have a buttery flavour and are used as additives in the food industry (Yebra et al., 2012). The presence of the lactococcal aceto-hydroxy-acid synthase (*ilvBN*) gene and deletion of lactate dehydrogenase gene (*ldh*) resulted in an increase in diacetyl/acetoin synthesis from glucose, but strain with only *ldh* deletion showed a similar result. By contrast, when the bacterial cells were exposed to lactose, strains carrying the *ilvBN* gene showed four times higher production of the desired compounds. The strain containing *ilvBN* and *ldh* mutations and a strain with additional *pdhC* (gene coding the E2-dihydrolipoamide-acetyl-transferase, component of pyruvate dehydrogenase complex Pdh) mutation were used for whey permeate fermentations. Having found the most suitable conditions for pH, the total amount of diacetyl/acetoin production was higher for the strain with the *pdhC* mutation. Fed batch experiments with this strain were done with the addition of whey permeate and yeast extract, but no further increase in diacetyl/acetoin concentrations was observed, and it was proposed that higher concentrations of product might have inhibitory effect. However, the amount of product obtained was still lower compared to engineered *Lactococcus lactis* (Nadal et al., 2009).

Exopolysaccharides (EPS) have been widely used in food industry, as they impact on the texture of food products, but they have also been shown to possess prebiotic characteristics (Papagianni, 2012). The EPS production levels in LAB are relatively low, and there have been several attempts to increase its production, mainly in *Lactococcus lactis* (for review see Gaspar et al. (2013)). In an attempt to increase

EPS production in *L. casei*, the effects of cofactors involved in EPS biosynthesis were investigated. The gene encoding NADH oxidase (*nox*), from *Streptococcus mutans*, was cloned and overexpressed in *L. casei* LC2W. The strain obtained grew slower than the wild type, but showed 46 % increase in EPS production (Li et al., 2015b). Furthermore, several other genes believed to be involved in EPS production were chosen from different *Lactobacillus* strains (*L. plantarum, L. casei* and *L. rhamnosus*) and their effect on EPS biosynthesis was tested. The genes *tga* (trans-glutaminase), *pfk* (phospho-fructokinase), *pgm* (phospho-glucomutase), *galtf* (galacto-transferase), *rhatf* (rhamnosyl-transferase), *rfbB* (dTDP-glucose-4,6-dehydratase) and *galT* (galactose-1-phosphate-urydil-transferase), and previously described *nox* (NADH oxidase), all involved in various steps of EPS production were successfully cloned and overexpressed in *L. casei* LC2W. Although recombinant strains had slower growth rates, some of them showed the positive effect of overexpressed genes (*pfk, rfbB* and *galT*) on EPS production (Li et al., 2015a), but lower than for the previously described *nox*-mutant. Besides that, the *nox*-mutant was shown to produce EPS in higher amounts in aerobic conditions, although growth was less than in anaerobic conditions. In aerobic conditions, the strain with overexpression of NADH oxidase reduced used more NADH and produced lower amounts of lactate, all of which led to the increased EPS production (Li et al., 2015a).

The question remains, however, would engineered bacteria be acceptable for direct use in food production. According to the current EU legislation (Directive 2009/41/EC of the European Parliament and of the Council), a genetically modified microorganism (GMM) is any microorganism that has foreign DNA introduced in a way that does not occur naturally. Many of these modified bacteria

could potentially be used in dairy food production where they could contribute to flavour and texture or fermented products containing these LAB could be used as a vehicle for probiotic delivery. However, these foods would have GMO status and fall under specific legislation, and guidelines for their applications have been proposed (European Food Safety Authority, 2011). It also raises issues in applicability and market potential as well as consumer acceptance of the modified LAB and careful analysis of variations in legislatives as well as possibilities and limits in applying genetically modified LAB in food, mainly in regard to consumers risk and benefits, should be taken into consideration (Pedersen et al., 2005, Sybesma et al., 2006). In addition, new approaches of genome editing with employment of CRISPR-Cas system would not be seen as GMM-generating tools according to the current definition, as it was recently discussed in case of genetically edited crops (Kanchiswamy et al., 2015), as only oligonucleotides that correspond to native molecules are needed for this reaction and the complex that derives edition is further degraded in the cell. This opens questions about redefining GMM and their use in the food industry. One issue that has to be considered is the fact that although the CRISPR systems have a high specificity level, the problem of unexpected negative effects remains a possibility, which could have massive effect on global food market (Au, 2015).

On the other hand, less restriction embraces the usage of modified LAB as potential cell factories. The era of application of recombinant bacteria for molecules started with human insulin production by recombinant *E. coli* developed in late 1970's (Goeddel et al., 1979). In general, LAB are recognised as safe and non-pathogenic, which makes them suitable for engineering projects. Even though these cells are engineered, the final product is purified and separated from the bacterial producer

and is used as a sole chemical in food or other industries. However, the disposal of GMM in these cases presents a challenge, and optimal destruction and prevention of environmental dissemination of engineered strains have to be implemented in industrial strategies (Gautier, 2008).

1.7 Concluding remarks

The *Lactobacillus* genus represents a versatile group of LAB that continues to intrigue scientists from different fields of microbiology. Their genetic characteristics are constantly being supplemented with new data. The rising number of available genomes provides greater opportunities for implementation of the data to give a better understanding of and improved application of these microorganisms. Construction of pangenomes reveals genetic and phenotypic diversity, and explains adaptability of lactobacilli to various habitats. Genetic data can be also used to anticipate the potential of strains for application in various industrial fields.

The construction of genome scale computational models gives an indication of a strains metabolic potential and facilitates identification of genes most suitable for engineering studies (Bution et al., 2015). The introduction of next generation sequencing (NGS) methods and metabolite profiling reveals new and unexpected features of LAB. The construction of metabolic models of industrial microorganisms is becoming an essential step in the development of fermented foods and food ingredients (Smid and Hugenholtz, 2010). The overall knowledge obtained after deployment of all approaches described in this review contributes to a better understanding of the physiology of *Lactobacillus* cultures during dairy production, which encourages the development of novel production technologies that will provide continuous product quality improvement (Steele et al., 2013).

1.8 Acknowledgments

The authors wish to thank to Dr Mary Rea and Dr Brian Healy for their critical reading of this manuscript. Ewelina Stefanovic is in a receipt of Teagasc Walsh Fellowship.

1.9 References

Altermann, E., Russell, W. M., Azcarate-Peril, M. A., Barrangou, R., Buck, B. L., McAuliffe, O., Souther, N., Dobson, A., Duong, T., Callanan, M., Lick, S., Hamrick, A., Cano, R. & Klaenhammer, T. R. 2005. Complete genome sequence of the probiotic lactic acid bacterium *Lactobacillus acidophilus* NCFM. *Proc Natl Acad Sci U S A,* 102**,** 3906-12.

Au, R. 2015. From genetic engineering to genome engineering: what impact has it made on science and society. *Adv Biol Biotechnol Genet,* 2**,** 1-8.

Babu, R. P., O'Connor, K. & Seeram, R. 2013. Current progress on bio-based polymers and their future trends. *Progr Biomater,* 2**,** 1-16.

Barrangou, R. 2013. CRISPR-Cas systems and RNA-guided interference. *Wiley Interdiscip Rev RNA,* 4**,** 267-78.

Barrangou, R. & Horvath, P. 2012. CRISPR: new horizons in phage resistance and strain identification. *Annu Rev Food Sci Technol,* 3**,** 143-162.

Beauprez, J. J., De Mey, M. & Soetaert, W. K. 2010. Microbial succinic acid production: Natural versus metabolic engineered producers. *Process Biochem,* 45**,** 1103-1114.

Bellanger, X., Payot, S., Leblond-Bourget, N. & Guedon, G. 2014. Conjugative and mobilizable genomic islands in bacteria: evolution and diversity. *FEMS Microbiol Rev,* 38**,** 720-760.

Berger, B., Pridmore, R. D., Barretto, C., Delmas-Julien, F., Schreiber, K., Arigoni, F. & Brussow, H. 2007. Similarity and differences in the *Lactobacillus acidophilus* group identified by polyphasic analysis and comparative genomics. *J Bacteriol,* 189**,** 1311-1321.

Bermudez-Humaran, L. G., Kharrat, P., Chatel, J. M. & Langella, P. 2011. Lactococci and lactobacilli as mucosal delivery vectors for therapeutic proteins and DNA vaccines. *Microb Cell Fact,* 10 Suppl 1**,** S4.

Bhowmik, T. & Steele, J. L. 1994. Cloning, characterization and insertional inactivation of the *Lactobacillus helveticus* D(-) lactate dehydrogenase gene. *Appl Microbiol Biotechnol,* 41**,** 432-439.

Bogicevic, B., Fuchsmann, P., Breme, K., Portmann, R., Guggenbuhl, B. & Irmler, S. 2013. A preliminary study on the effect of *Lactobacillus casei* expressing cystathionine lyase1/cystathionine lyase2 on Cheddar cheese and the formation of sulphur-containing compounds. *Int Dairy J,* 33**,** 97-103.

Bogicevic, B., Irmler, S., Portmann, R., Meile, L. & Berthoud, H. 2012. Characterization of the *cysK2-ctl1-cysE2* gene cluster involved in sulfur metabolism in *Lactobacillus casei*. *Int J Food Microbiol,* 152**,** 211-219.

Broadbent, J. R., Cai, H., Larsen, R. L., Hughes, J. E., Welker, D. L., De Carvalho, V. G., Tompkins, T. A., Ardo, Y., Vogensen, F., De Lorentiis, A., Gatti, M., Neviani, E. & Steele, J. L. 2011. Genetic diversity in proteolytic enzymes and amino acid metabolism among *Lactobacillus helveticus* strains. *J Dairy Sci,* 94**,** 4313-4328.

Broadbent, J. R., Hughes, J. E., Welker, D. L., Tompkins, T. A. & Steele, J. L. 2013. Complete Genome Sequence for *Lactobacillus helveticus* CNRZ 32, an Industrial Cheese Starter and Cheese Flavor Adjunct. *Genome Announc,* 1**,** e00590-13.

Broadbent, J. R., Neeno-Eckwall, E. C., Stahl, B., Tandee, K., Cai, H., Morovic, W., Horvath, P., Heidenreich, J., Perna, N. T., Barrangou, R. & Steele, J. L. 2012. Analysis of the *Lactobacillus casei* supragenome and its influence in species evolution and lifestyle adaptation. *BMC Genomics,* 13**,** 533.

Bull, M. J., Jolley, K. A., Bray, J. E., Aerts, M., Vandamme, P., Maiden, M. C., Marchesi, J. R. & Mahenthiralingam, E. 2014. The domestication of the probiotic bacterium *Lactobacillus acidophilus*. *Sci Rep,* 4**,** 7202.

Buriti, F. C. A., da Rocha, J. S. & Saad, S. M. I. 2005. Incorporation of *Lactobacillus acidophilus* in Minas fresh cheese and its implications for textural and sensorial properties during storage. *Int Dairy J,* 15**,** 1279-1288.

Bution, M. L., Molina, G., Abrahao, M. R. & Pastore, G. M. 2015. Genetic and metabolic engineering of microorganisms for the development of new flavor compounds from terpenic substrates. *Crit Rev Biotechnol,* 35**,** 313-325.

Cai, H., Rodriguez, B. T., Zhang, W., Broadbent, J. R. & Steele, J. L. 2007. Genotypic and phenotypic characterization of *Lactobacillus casei* strains isolated from different ecological niches suggests frequent recombination and niche specificity. *Microbiology,* 153**,** 2655-2665.

Cai, H., Thompson, R., Budinich, M. F., Broadbent, J. R. & 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.

Callanan, M., Kaleta, P., O'Callaghan, J., O'Sullivan, O., Jordan, K., McAuliffe, O., Sangrador-Vegas, A., Slattery, L., Fitzgerald, G. F., Beresford, T. & Ross, R. P. 2008. Genome sequence of *Lactobacillus helveticus,* an organism distinguished by selective gene loss and insertion sequence element expansion. *J Bacteriol,* 190**,** 727-735.

Ceapa, C., Lambert, J., van Limpt, K., Wels, M., Smokvina, T., Knol, J. & Kleerebezem, M. 2015. Correlation of *Lactobacillus rhamnosus* Genotypes and Carbohydrate Utilization Signatures Determined by Phenotype Profiling. *Appl Environ Microbiol,* 81**,** 5458-5470.

Claesson, M. J., van Sinderen, D. & O'Toole, P. W. 2007. The genus *Lactobacillus*-a genomic basis for understanding its diversity. *FEMS Microbiol Lett,* 269**,** 22-28.

Claesson, M. J., van Sinderen, D. & O'Toole, P. W. 2008. *Lactobacillus* phylogenomics-towards a reclassification of the genus. *Int J Syst Evol Microbiol,* 58**,** 2945-2954.

Cleveland, J., Montville, T. J., Nes, I. F. & Chikindas, M. L. 2001. Bacteriocins: safe, natural antimicrobials for food preservation. *Int J Food Microbiol,* 71**,** 1-20.

Collins, R. E. & Higgs, P. G. 2012. Testing the infinitely many genes model for the evolution of the bacterial core genome and pangenome. *Mol Biol Evol,* 29**,** 3413-3425.

Collins, Y. F., McSweeney, P. L. H. & Wilkinson, M. G. 2003. Lipolysis and free fatty acid catabolism in cheese: a review of current knowledge. *Int Dairy J,* 13**,** 841-866.

Corsetti, A. & Settanni, L. 2007. Lactobacilli in sourdough fermentation. *Food Res Int,* 40**,** 539-558.

Cotter, P. D., Hill, C. & Ross, R. P. 2005. Bacteriocins: developing innate immunity for food. *Nat Rev Microbiol,* 3**,** 777-788.

Cremonesi, P., Chessa, S. & Castiglioni, B. 2012. Genome sequence and analysis of *Lactobacillus helveticus*. *Front Microbiol,* 3**,** 435.

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.

De Boeck, R., Sarmiento-Rubiano, L. A., Nadal, I., Monedero, V., Perez-Martinez, G. & Yebra, M. J. 2010. Sorbitol production from lactose by engineered *Lactobacillus casei* deficient in sorbitol transport system and mannitol-1-phosphate dehydrogenase. *Appl Microbiol Biotechnol,* 85**,** 1915-1922.

De Vuyst, L. & Leroy, F. 2007. Bacteriocins from lactic acid bacteria: production, purification, and food applications. *J Mol Microbiol Biotechnol,* 13**,** 194-199.

Dellaglio, F., Felis, G. E. & Torriani, S. 2002. The status of the species *Lactobacillus casei* (Orla-Jensen 1916) Hansen and Lessel 1971 and *Lactobacillus paracasei* Collins et al. 1989. Request for an opinion. *Int J Syst Evol Microbiol,* 52**,** 285-287.

Deveau, H., Garneau, J. E. & Moineau, S. 2010. CRISPR/Cas system and its role in phage-bacteria interactions. *Annu Rev Microbiol,* 64**,** 475-493.

Douglas, G. L. & Klaenhammer, T. R. 2010. Genomic evolution of domesticated microorganisms. *Annu Rev Food Sci Technol,* 1**,** 397-414.

Douillard, F. P., Ribbera, A., Kant, R., Pietila, T. E., Jarvinen, H. M., Messing, M., Randazzo, C. L., Paulin, L., Laine, P., Ritari, J., Caggia, C., Lahteinen, T., Brouns, S. J., Satokari, R., von Ossowski, I., Reunanen, J., Palva, A. & de Vos, W. M. 2013. Comparative genomic and functional analysis of 100 *Lactobacillus rhamnosus* strains and their comparison with strain GG. *PLoS Genet,* 9**,** e1003683.

European Food Safety Authority (EFSA). 2011. Guidance on the risk assessment of genetically modified microorganisms and their products intended for food and feed use. *EFSA Journal,* 9**,** 2193.

Ejtahed, H. S., Mohtadi-Nia, J., Homayouni-Rad, A., Niafar, M., Asghari-Jafarabadi, M., Mofid, V. & Akbarian-Moghari, A. 2011. Effect of probiotic yogurt containing *Lactobacillus acidophilus* and *Bifidobacterium lactis* on lipid profile in individuals with type 2 diabetes mellitus. *J Dairy Sci,* 94**,** 3288-3294.

El Kafsi, H., Binesse, J., Loux, V., Buratti, J., Boudebbouze, S., Dervyn, R., Kennedy, S., Galleron, N., Quinquis, B., Batto, J. M., Moumen, B., Maguin, E. & van de Guchte, M. 2014. *Lactobacillus delbrueckii* ssp. *lactis* and ssp. *bulgaricus:* a chronicle of evolution in action. *BMC Genomics,* 15**,** 407.

Ferain, T., Garmyn, D., Bernard, N., Hols, P. & Delcour, J. 1994. *Lactobacillus plantarum ldhL* gene: overexpression and deletion. *J Bacteriol,* 176**,** 596-601.

Garneau, J. E. & Moineau, S. 2011. Bacteriophages of lactic acid bacteria and their impact on milk fermentations. *Microb Cell Fact,* 10 Suppl 1**,** S20.

Gaspar, P., Carvalho, A. L., Vinga, S., Santos, H. & Neves, A. R. 2013. From physiology to systems metabolic engineering for the production of biochemicals by lactic acid bacteria. *Biotechnol Adv,* 31**,** 764-788.

Gautier, M., 2008. Ethical Issues Raised by Genetically Modified Microorganisms. http://bioethics.agrocampus-ouest.eu/infoglueDeliverLive/digitalAssets/57484_41EN-ethical-issues-ogm.pdf. Last accessed on 20.07.16.

Genay, M., Sadat, L., Gagnaire, V. & Lortal, S. 2009. prtH2, not prtH, is the ubiquitous cell wall proteinase gene in *Lactobacillus helveticus*. *Appl Environ Microbiol,* 75**,** 3238-3249.

Giraffa, G. 2014. *Lactobacillus helveticus*: importance in food and health. *Front Microbiol,* 5**,** 338.

Giraffa, G., De Vecchi, P. & Rossetti, L. 1998. Note: identification of *Lactobacillus delbrueckii* subspecies *bulgaricus* and subspecies *lactis* dairy isolates by amplified rDNA restriction analysis. *J Appl Microbiol,* 85**,** 918-924.

Gobbetti, M., De Angelis, M., Di Cagno, R., Mancini, L. & Fox, P. F. 2015. Pros and cons for using non-starter lactic acid bacteria (NSLAB) as secondary/adjunct starters for cheese ripening. *Trends Food Sci Tech,* 45**,** 167-178.

Goeddel, D. V., Kleid, D. G., Bolivar, F., Heyneker, H. L., Yansura, D. G., Crea, R., Hirose, T., Kraszewski, A., Itakura, K. & Riggs, A. D. 1979. Expression in *Escherichia coli* of chemically synthesized genes for human insulin. *Proc Natl Acad Sci U S A,* 76**,** 106-110.

Gold, R. S., Meagher, M. M., Tong, S. X., Hutkins, R. W. & Conway, T. 1996. Cloning and expression of the *Zymomonas mobilis* "production of ethanol" genes in *Lactobacillus casei*. *Curr Microbiol,* 33**,** 256-260.

Hao, P., Zheng, H., Yu, Y., Ding, G., Gu, W., Chen, S., Yu, Z., Ren, S., Oda, M., Konno, T., Wang, S., Li, X., Ji, Z. S. & Zhao, G. 2011. Complete sequencing and pan-genomic analysis of *Lactobacillus delbrueckii* subsp. *bulgaricus* reveal its genetic basis for industrial yogurt production. *PLoS One,* 6**,** e15964.

Herias, M. V., Koninkx, J. F., Vos, J. G., Huis in't Veld, J. H. & van Dijk, J. E. 2005. Probiotic effects of *Lactobacillus casei* on DSS-induced ulcerative colitis in mice. *Int J Food Microbiol,* 103**,** 143-155.

Horvath, P., Coute-Monvoisin, A. C., Romero, D. A., Boyaval, P., Fremaux, C. & Barrangou, R. 2009. Comparative analysis of CRISPR loci in lactic acid bacteria genomes. *Int J Food Microbiol,* 131**,** 62-70.

Iartchouk, O., Kozyavkin, S., Karamychev, V. & Slesarev, A. 2015. Complete Genome Sequence of *Lactobacillus acidophilus* FSI4, Isolated from Yogurt. *Genome Announc,* 3**,** e00166-15

Innocente, N., Biasutti, M., Rita, F., Brichese, R., Comi, G. & Iacumin, L. 2016. Effect of indigenous *Lactobacillus rhamnosus* isolated from bovine milk on microbiological characteristics and aromatic profile of traditional yogurt. *LWT-Food Sci Technol,* 66**,** 158-164.

Irmler, S., Raboud, S., Beisert, B., Rauhut, D. & Berthoud, H. 2008. Cloning and characterization of two *Lactobacillus casei* genes encoding a cystathionine lyase. *Appl Environ Microbiol,* 74**,** 99-106.

Irmler, S., Schafer, H., Beisert, B., Rauhut, D. & Berthoud, H. 2009. Identification and characterization of a strain-dependent cystathionine beta/gamma-lyase in *Lactobacillus casei* potentially involved in cysteine biosynthesis. *FEMS Microbiol Lett,* 295**,** 67-76.

Jiang, W. & Marraffini, L. A. 2015. CRISPR-Cas: New Tools for Genetic Manipulations from Bacterial Immunity Systems. *Annu Rev Microbiol,* 69**,** 209-228.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A. & Charpentier, E. 2012. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science,* 337**,** 816-821.

Kamerbeek, J., Schouls, L., Kolk, A., van Agterveld, M., van Soolingen, D., Kuijper, S., Bunschoten, A., Molhuizen, H., Shaw, R., Goyal, M. & van Embden, J. 1997. Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology. *J Clin Microbiol,* 35**,** 907-914.

Kanchiswamy, C. N., Malnoy, M., Velasco, R., Kim, J. S. & Viola, R. 2015. Non-GMO genetically edited crop plants. *Trends Biotechnol,* 33**,** 489-491.

Kant, R., Blom, J., Palva, A., Siezen, R. J. & de Vos, W. M. 2010. Comparative genomics of *Lactobacillus*. *Microb Biotechnol,* 4**,** 323-332.

Kant, R., Rintahaka, J., Yu, X., Sigvart-Mattila, P., Paulin, L., Mecklin, J. P., Saarela, M., Palva, A. & von Ossowski, I. 2014. A comparative pan-genome perspective of niche-adaptable cell-surface protein phenotypes in *Lactobacillus rhamnosus*. *PLoS One,* 9**,** e102762.

Keasling, J. D. 2012. Synthetic biology and the development of tools for metabolic engineering. *Metab Eng,* 14**,** 189-195.

King, Z. A., Lloyd, C. J., Feist, A. M. & Palsson, B. O. 2015. Next-generation genome-scale models for metabolic engineering. *Curr Opin Biotechnol,* 35**,** 23-29.

Kleerebezem, M., Boekhorst, J., van Kranenburg, R., Molenaar, D., Kuipers, O. P., Leer, R., Tarchini, R., Peters, S. A., Sandbrink, H. M., Fiers, M. W., Stiekema, W., Lankhorst, R. M., Bron, P. A., Hoffer, S. M., Groot, M. N., Kerkhoven, R., de Vries,

M., Ursing, B., de Vos, W. M. & Siezen, R. J. 2003. Complete genome sequence of *Lactobacillus plantarum* WCFS1. *Proc Natl Acad Sci U S A,* 100**,** 1990-1995.

Kyla-Nikkila, K., Hujanen, M., Leisola, M. & Palva, A. 2000. Metabolic engineering of *Lactobacillus helveticus* CNRZ32 for production of pure L-(+)-lactic acid. *Appl Environ Microbiol,* 66**,** 3835-3841.

Ladero, V., Ramos, A., Wiersma, A., Goffin, P., Schanck, A., Kleerebezem, M., Hugenholtz, J., Smid, E. J. & Hols, P. 2007. High-level production of the low-calorie sugar sorbitol by *Lactobacillus plantarum* through metabolic engineering. *Appl Environ Microbiol,* 73**,** 1864-1872.

Lebeer, S., Vanderleyden, J. & De Keersmaecker, S. C. 2008. Genes and molecules of lactobacilli supporting probiotic action. *Microbiol Mol Biol Rev,* 72**,** 728-764.

Lewis, N. E., Nagarajan, H. & Palsson, B. O. 2012. Constraining the metabolic genotype-phenotype relationship using a phylogeny of *in silico* methods. *Nat Rev Microbiol,* 10**,** 291-305.

Li, N., Huang, Y., Liu, Z., You, C. & Guo, B. 2015a. Regulation of EPS production in *Lactobacillus casei* LC2W through metabolic engineering. *Lett Appl Microbiol,* 61**,** 555-561.

Li, N., Wang, Y., Zhu, P., Liu, Z., Guo, B. & Ren, J. 2015b. Improvement of exopolysaccharide production in *Lactobacillus casei* LC2W by overexpression of NADH oxidase gene. *Microbiol Res,* 171**,** 73-77.

Liu, E., Hao, P., Konno, T., Yu, Y., Oda, M., Zheng, H. & Ji, Z.-S. 2012. Amino Acid Biosynthesis and Proteolysis in *Lactobacillus Bulgaricus* Revisited: A Genomic Comparison. *Comput Mol Biosci,* 02**,** 61-77.

Liu, M., Bayjanov, J. R., Renckens, B., Nauta, A. & Siezen, R. J. 2010. The proteolytic system of lactic acid bacteria revisited: a genomic comparison. *BMC Genomics,* 11**,** 36.

Liu, M., Bienfait, B., Sacher, O., Gasteiger, J., Siezen, R. J., Nauta, A. & Geurts, J. M. 2014. Combining chemoinformatics with bioinformatics: *in silico* prediction of

bacterial flavor-forming pathways by a chemical systems biology approach "reverse pathway engineering". *PLoS One,* 9**,** e84769.

Liu, M., Nauta, A., Francke, C. & Siezen, R. J. 2008. Comparative genomics of enzymes in flavor-forming pathways from amino acids in lactic acid bacteria. *Appl Environ Microbiol,* 74**,** 4590-4600.

Liu, S., Nichols, N. N., Dien, B. S. & Cotta, M. A. 2006. Metabolic engineering of a *Lactobacillus plantarum* double ldh knockout strain for enhanced ethanol production. *J Ind Microbiol Biotechnol,* 33**,** 1-7.

Ljungh, A. & Wadstrom, T. 2006. Lactic acid bacteria as probiotics. *Curr Issues Intest Microbiol,* 7**,** 73-90.

Mahadevan, R. & Schilling, C. H. 2003. The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. *Metab Eng,* 5**,** 264-276.

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

Makarova, K. S. & Koonin, E. V. 2007. Evolutionary genomics of lactic acid bacteria. *J Bacteriol,* 189**,** 1199-1208.

Marco, M. B., Moineau, S. & Quiberoni, A. 2012. Bacteriophages and dairy fermentations. *Bacteriophage,* 2**,** 149-158.

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

Mayo, B., Aleksandrzak-Piekarczyk, T., Fernández, M., Kowalczyk, M., Álvarez-Martín, P. & Bardowski, J. 2010. Updates in the Metabolism of Lactic Acid Bacteria. *In*: Mozzi, F., Raya, R. & Vignolo, G. (eds.) *Biotechnology of Lactic Acid Bacteria.* Wiley-Blackwell.

Mayo, B., van Sinderen, D. & Ventura, M. 2008. Genome analysis of food grade lactic Acid-producing bacteria: from basics to applications. *Curr Genomics,* 9**,** 169-183.

Mazzoli, R., Bosco, F., Mizrahi, I., Bayer, E. A. & Pessione, E. 2014. Towards lactic acid bacteria-based biorefineries. *Biotechnol Adv,* 32**,** 1216-1236.

McAuliffe, O. 2017. Genetics of lactic acid bacteria. *In*: Fox, P.F., McSweeney, P.L.H., Cotter, P., Everett, D.W. (eds.) *Cheese: Chemistry, Physics and Microbiology,* fourth ed. Elsevier.

Medina de Figueroa, R., Oliver, G. & Benito de Cádenas, I. L. 2001. Influence of temperature on flavour compound production from citrate by *Lactobacillus rhamnosus* ATCC 7469. *Microbiological Research,* 155**,** 257-262.

Medini, D., Donati, C., Tettelin, H., Masignani, V. & Rappuoli, R. 2005. The microbial pan-genome. *Curr Opin Genet Dev,* 15**,** 589-594.

Meneghel, J., Dugat-Bony, E., Irlinger, F., Loux, V., Vidal, M., Passot, S., Beal, C., Layec, S. & Fonseca, F. 2016. Draft Genome Sequence of *Lactobacillus delbrueckii* subsp. *bulgaricus* CFL1, a Lactic Acid Bacterium Isolated from French Handcrafted Fermented Milk. *Genome Announc,* 4**,** e00052-16.

Milesi, M. M., Wolf, I. V., Bergamini, C. V. & Hynes, E. R. 2010. Two strains of nonstarter lactobacilli increased the production of flavor compounds in soft cheeses. *J Dairy Sci,* 93**,** 5020-5031.

Molenaar, D., Bringel, F., Schuren, F. H., de Vos, W. M., Siezen, R. J. & Kleerebezem, M. 2005. Exploring *Lactobacillus plantarum* genome diversity by using microarrays. *J Bacteriol,* 187**,** 6119-6127.

Mortera, P., Pudlik, A., Magni, C., Alarcon, S. & Lolkema, J. S. 2013. Ca2+-citrate uptake and metabolism in *Lactobacillus casei* ATCC 334. *Appl Environ Microbiol,* 79**,** 4603-4612.

Nadal, I., Rico, J., Perez-Martinez, G., Yebra, M. J. & Monedero, V. 2009. Diacetyl and acetoin production from whey permeate using engineered *Lactobacillus casei*. *J Ind Microbiol Biotechnol,* 36**,** 1233-1237.

Nichols, N. N., Dien, B. S. & Bothast, R. J. 2003. Engineering lactic acid bacteria with pyruvate decarboxylase and alcohol dehydrogenase genes for ethanol production from *Zymomonas mobilis*. *J Ind Microbiol Biotechnol,* 30**,** 315-321.

Nissen, L., Perez-Martinez, G. & Yebra, M. J. 2005. Sorbitol synthesis by an engineered *Lactobacillus casei* strain expressing a sorbitol-6-phosphate dehydrogenase gene within the lactose operon. *FEMS Microbiol Lett,* 249**,** 177-183.

O'Brien, E. J., Monk, J. M. & Palsson, B. O. 2015. Using Genome-scale Models to Predict Biological Capabilities. *Cell,* 161**,** 971-987.

O'Sullivan, O., O'Callaghan, J., Sangrador-Vegas, A., McAuliffe, O., Slattery, L., Kaleta, P., Callanan, M., Fitzgerald, G. F., Ross, R. P. & Beresford, T. 2009. Comparative genomics of lactic acid bacteria reveals a niche-specific gene set. *BMC Microbiol,* 9**,** 50.

Oliveira, A. P., Nielsen, J. & Forster, J. 2005. Modeling *Lactococcus lactis* using a genome-scale flux model. *BMC Microbiol,* 5**,** 39.

Ong, L., Henriksson, A. & Shah, N. P. 2007. Chemical analysis and sensory evaluation of Cheddar cheese produced with *Lactobacillus acidophilus, L. casei, L. paracasei* or *Bifidobacterium* sp. *Int Dairy J,* 17**,** 937-945.

Orth, J. D., Thiele, I. & Palsson, B. O. 2010. What is flux balance analysis? *Nat Biotechnol,* 28**,** 245-248.

Papagianni, M. 2012. Metabolic engineering of lactic acid bacteria for the production of industrially important compounds. *Comput Struct Biotechnol J,* 3**,** e201210003.

Pastink, M. I., Teusink, B., Hols, P., Visser, S., de Vos, W. M. & Hugenholtz, J. 2009. Genome-scale model of *Streptococcus thermophilus* LMG18311 for metabolic comparison of lactic acid bacteria. *Appl Environ Microbiol,* 75**,** 3627-3633.

Pedersen, M. B., Iversen, S. L., Sorensen, K. I. & 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.

Pfeiler, E. A. & Klaenhammer, T. R. 2007. The genomics of lactic acid bacteria. *Trends Microbiol,* 15**,** 546-553.

Pridmore, R. D., Berger, B., Desiere, F., Vilanova, D., Barretto, C., Pittet, A. C., Zwahlen, M. C., Rouvet, M., Altermann, E., Barrangou, R., Mollet, B., Mercenier, A., Klaenhammer, T., Arigoni, F. & Schell, M. A. 2004. The genome sequence of the probiotic intestinal bacterium *Lactobacillus johnsonii* NCC 533. *Proc Natl Acad Sci U S A,* 101**,** 2512-2517.

Ramachandran, P., Lacher, D. W., Pfeiler, E. A. & Elkins, C. A. 2013. Development of a tiered multilocus sequence typing scheme for members of the *Lactobacillus acidophilus* complex. *Appl Environ Microbiol,* 79**,** 7220-7228.

Rossi, F., Rizzotti, L., Felis, G. E. & Torriani, S. 2014. Horizontal gene transfer among microorganisms in food: current knowledge and future perspectives. *Food Microbiol,* 42**,** 232-243.

Sander, J. D. & Joung, J. K. 2014. CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat Biotechnol,* 32**,** 347-355.

Savijoki, K., Ingmer, H. & Varmanen, P. 2006. Proteolytic systems of lactic acid bacteria. *Appl Microbiol Biotechnol,* 71**,** 394-406.

Selle, K. & Barrangou, R. 2015. Harnessing CRISPR-Cas systems for bacterial genome editing. *Trends Microbiol,* 23**,** 225-232.

Senan, S., Prajapati, J. B. & Joshi, C. G. 2014. Comparative genome-scale analysis of niche-based stress-responsive genes in *Lactobacillus helveticus* strains. *Genome,* 57**,** 185-192.

Senan, S., Prajapati, J. B. & Joshi, C. G. 2015. Whole-genome based validation of the adaptive properties of Indian origin probiotic *Lactobacillus helveticus* MTCC 5463. *J Sci Food Agric,* 95**,** 321-328.

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

Sgarbi, E., Lazzi, C., Tabanelli, G., Gatti, M., Neviani, E. & Gardini, F. 2013. Nonstarter lactic acid bacteria volatilomes produced using cheese components. *J Dairy Sci,* 96**,** 4223-34.

Siezen, R. J., Tzeneva, V. A., Castioni, A., Wels, M., Phan, H. T., Rademaker, J. L., Starrenburg, M. J., Kleerebezem, M., Molenaar, D. & van Hylckama Vlieg, J. E. 2010. Phenotypic and genomic diversity of *Lactobacillus plantarum* strains isolated from various environmental niches. *Environ Microbiol,* 12**,** 758-773.

Siezen, R. J. & van Hylckama Vlieg, J. E. 2011. Genomic diversity and versatility of *Lactobacillus plantarum,* a natural metabolic engineer. *Microb Cell Fact,* 10 Suppl 1**,** S3.

Smid, E. J. & Hugenholtz, J. 2010. Functional genomics for food fermentation processes. *Annu Rev Food Sci Technol,* 1**,** 497-519.

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

Smokvina, T., Wels, M., Polka, J., Chervaux, C., Brisse, S., Boekhorst, J., van Hylckama Vlieg, J. E. & Siezen, R. J. 2013. *Lactobacillus paracasei* comparative genomics: towards species pan-genome definition and exploitation of diversity. *PLoS One,* 8**,** e68731.

Stahl, B. & Barrangou, R. 2013. Complete Genome Sequence of Probiotic Strain *Lactobacillus acidophilus* La-14. *Genome Announc,* 1**,** e00376-13.

Steele, J., Broadbent, J. & Kok, J. 2013. Perspectives on the contribution of lactic acid bacteria to cheese flavor development. *Curr Opin Biotechnol,* 24**,** 135-141.

Strahinic, I., Lozo, J., Terzic-Vidojevic, A., Fira, D., Kojic, M., Golic, N., Begovic, J. & Topisirovic, L. 2013. Technological and probiotic potential of BGRA43 a natural isolate of *Lactobacillus helveticus*. *Front Microbiol,* 4**,** 2.

Sun, Z., Harris, H. M., McCann, A., Guo, C., Argimon, S., Zhang, W., Yang, X., Jeffery, I. B., Cooney, J. C., Kagawa, T. F., Liu, W., Song, Y., Salvetti, E., Wrobel, A., Rasinkangas, P., Parkhill, J., Rea, M. C., O'Sullivan, O., Ritari, J., Douillard, F. P., Paul Ross, R., Yang, R., Briner, A. E., Felis, G. E., de Vos, W. M., Barrangou, R., Klaenhammer, T. R., Caufield, P. W., Cui, Y., Zhang, H. & O'Toole, P. W. 2015a. Expanding the biotechnology potential of lactobacilli through comparative genomics of 213 strains and associated genera. *Nat Commun,* 6**,** 8322.

Sun, Z., Liu, W., Song, Y., Xu, H., Yu, J., Bilige, M., Zhang, H. & Chen, Y. 2015b. Population structure of *Lactobacillus helveticus* isolates from naturally fermented dairy products based on multilocus sequence typing. *J Dairy Sci,* 98**,** 2962-2972.

Sybesma, W., Hugenholtz, J., de Vos, W. M. & Smid, E. J. 2006. Safe use of genetically modified lactic acid bacteria in food. Bridging the gap between consumers, green groups, and industry. *Electron J Biotechn,* 9**,** 424-448.

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. *Ant Leeuwenhoek,* 82**,** 271-278.

Taverniti, V. & Guglielmetti, S. 2012. Health-Promoting Properties of *Lactobacillus helveticus*. *Front Microbiol,* 3**,** 392.

Teusink, B., Bachmann, H. & Molenaar, D. 2011. Systems biology of lactic acid bacteria: a critical review. *Microb Cell Fact,* 10 Suppl 1**,** S11.

Teusink, B., Wiersma, A., Molenaar, D., Francke, C., de Vos, W. M., Siezen, R. J. & Smid, E. J. 2006. Analysis of growth of *Lactobacillus plantarum* WCFS1 on a complex medium using a genome-scale metabolic model. *J Biol Chem,* 281**,** 40041-40048.

Toh, H., Oshima, K., Nakano, A., Takahata, M., Murakami, M., Takaki, T., Nishiyama, H., Igimi, S., Hattori, M. & Morita, H. 2013. Genomic adaptation of the *Lactobacillus casei* group. *PLoS One,* 8**,** e75073.

Tsuji, A., Okada, S., Hols, P. & Satoh, E. 2013. Metabolic engineering of *Lactobacillus plantarum* for succinic acid production through activation of the reductive branch of the tricarboxylic acid cycle. *Enzyme Microb Technol,* 53**,** 97-103.

Tuo, Y., Zhang, W., Zhang, L., Ai, L., Zhang, Y., Han, X. & Yi, H. 2013. Study of probiotic potential of four wild *Lactobacillus rhamnosus* strains. *Anaerobe,* 21**,** 22-27.

Urshev, Z. & Ishlimova, D. 2015. Distribution of clustered regularly interspaced palindrome repeats CRISPR2 and CRISPR3 in *Lactobacillus delbrueckii* ssp. *bulgaricus* strains. *Biotechnol Biotec Eq,* 29**,** 541-546.

Van de Guchte, M., Penaud, S., Grimaldi, C., Barbe, V., Bryson, K., Nicolas, P., Robert, C., Oztas, S., Mangenot, S., Couloux, A., Loux, V., Dervyn, R., Bossy, R., Bolotin, A., Batto, J. M., Walunas, T., Gibrat, J. F., Bessieres, P., Weissenbach, J., Ehrlich, S. D. & Maguin, E. 2006. The complete genome sequence of *Lactobacillus bulgaricus* reveals extensive and ongoing reductive evolution. *Proc Natl Acad Sci U S A,* 103**,** 9274-9279.

Van Hoorde, K., Van Leuven, I., Dirinck, P., Heyndrickx, M., Coudijzer, K., Vandamme, P. & Huys, G. 2010. Selection, application and monitoring of *Lactobacillus paracasei* strains as adjunct cultures in the production of Gouda-type cheeses. *Int J Food Microbiol,* 144**,** 226-235.

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.

Villena, J., Oliveira, M. L., Ferreira, P. C., Salva, S. & Alvarez, S. 2011. Lactic acid bacteria in the prevention of pneumococcal respiratory infection: future opportunities and challenges. *Int Immunopharmacol,* 11**,** 1633-1645.

Vinay-Lara, E., Hamilton, J. J., Stahl, B., Broadbent, J. R., Reed, J. L. & Steele, J. L. 2014. Genome-scale reconstruction of metabolic networks of *Lactobacillus casei* ATCC 334 and 12A. *PLoS One,* 9**,** e110785.

Wright, A. V., Nunez, J. K. & Doudna, J. A. 2016. Biology and Applications of CRISPR Systems: Harnessing Nature's Toolbox for Genome Engineering. *Cell,* 164**,** 29-44.

Wyszynska, A., Kobierecka, P., Bardowski, J. & Jagusztyn-Krynicka, E. K. 2015. Lactic acid bacteria-20 years exploring their potential as live vectors for mucosal vaccination. *Appl Microbiol Biotechnol,* 99**,** 2967-2977.

Xu, N., Liu, J., Ai, L. & Liu, L. 2015. Reconstruction and analysis of the genome-scale metabolic model of *Lactobacillus casei* LC2W. *Gene,* 554**,** 140-147.

Ya, T., Zhang, Q., Chu, F., Merritt, J., Bilige, M., Sun, T., Du, R. & Zhang, H. 2008. Immunological evaluation of *Lactobacillus casei* Zhang: a newly isolated strain from koumiss in Inner Mongolia, China. *BMC Immunol,* 9**,** 68.

Yebra, M. J., Monedero, V., Pérez-Martínez, G. & Rodríguez-Díaz, J. 2012. Genetically Engineered Lactobacilli for Technological and Functional Food Applications. *In*: Valdez, B. (ed.) *Food Industrial processes - Methods and Equipment,* InTech.

Yebra, M. J., Zuniga, M., Beaufils, S., Perez-Martinez, G., Deutscher, J. & Monedero, V. 2007. Identification of a gene cluster enabling *Lactobacillus casei* BL23 to utilize myo-inositol. *Appl Environ Microbiol,* 73**,** 3850-3858.

Yu, S., Peng, Y., Zheng, Y. & Chen, W. 2015. Comparative Genome Analysis of *Lactobacillus casei*: Insights into Genomic Diversification for Niche Expansion. *Indian J Microbiol,* 55**,** 102-107.

Yvon, M. 2006. Key enzymes for flavour formation by lactic acid bacteria. *Aust J Dairy Technol,* 61**,** 88-96.

Zhu, Y., Zhang, Y. & Li, Y. 2009. Understanding the industrial application potential of lactic acid bacteria through genomics. *Appl Microbiol Biotechnol,* 83**,** 597-610.

Table 1: General genomic features of the most important dairy related *Lactobacillus* species. All data were obtained at www.ncbi.nlm.nih.gov, last assessed in July 2016.

Species of *Lactobacillus*	Number of sequences available	Median total length (Mbp)	Median number of proteins	Median GC content (%)
L. delbrueckii	32	1.865	1637	49.8
L. helveticus	22	2.077	1784	36.8
L. casei	35	3.036	2736	46.4
L. paracasei	53	2.961	2749	46.3
L. acidophilus	16	1.979	1815	34.6
L. rhamnosus	102	2.937	2641	46.6
L. plantarum	114	3.275	2912	44.4

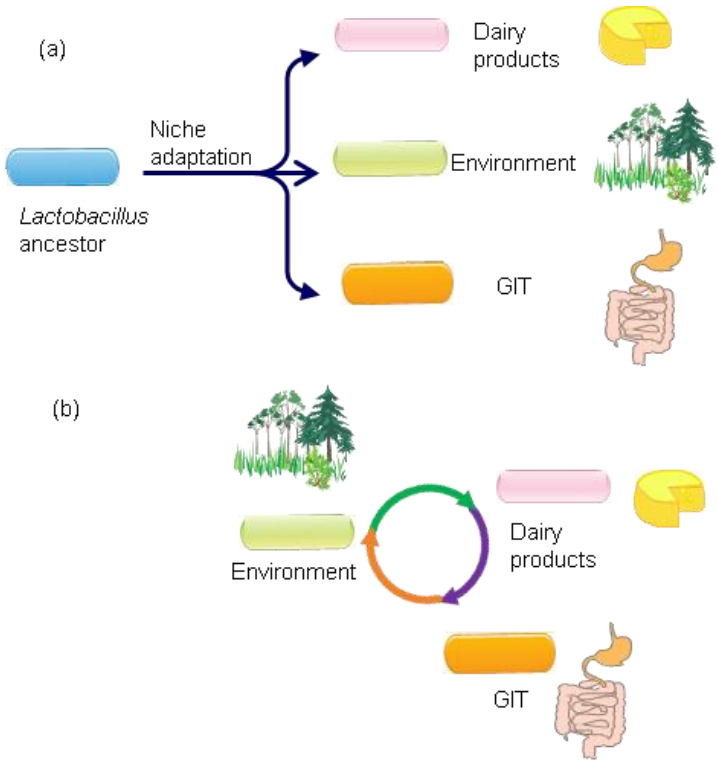


Figure 1: Process of niche adaptation. (a) Ancestor of *Lactobacillus* spp. had undergone multiple genome changes, such as decay of superfluous genes and acquisition of genes that support survival in specific environmental conditions, which all led to niche specialisation for various habitats, three of which have been depicted here (dairy, environment, human and animal GIT). However, strains of *Lactobacillus* could change their habitat (b), for instance during human consumption of dairy or plant food, and this is why isolation source does not always correspond to the strains' natural environment. This has to be kept in mind while analysing characteristics of strains isolated from different ecological niches, as origin of isolation gives only an indication of potential metabolic capacity of an organism.

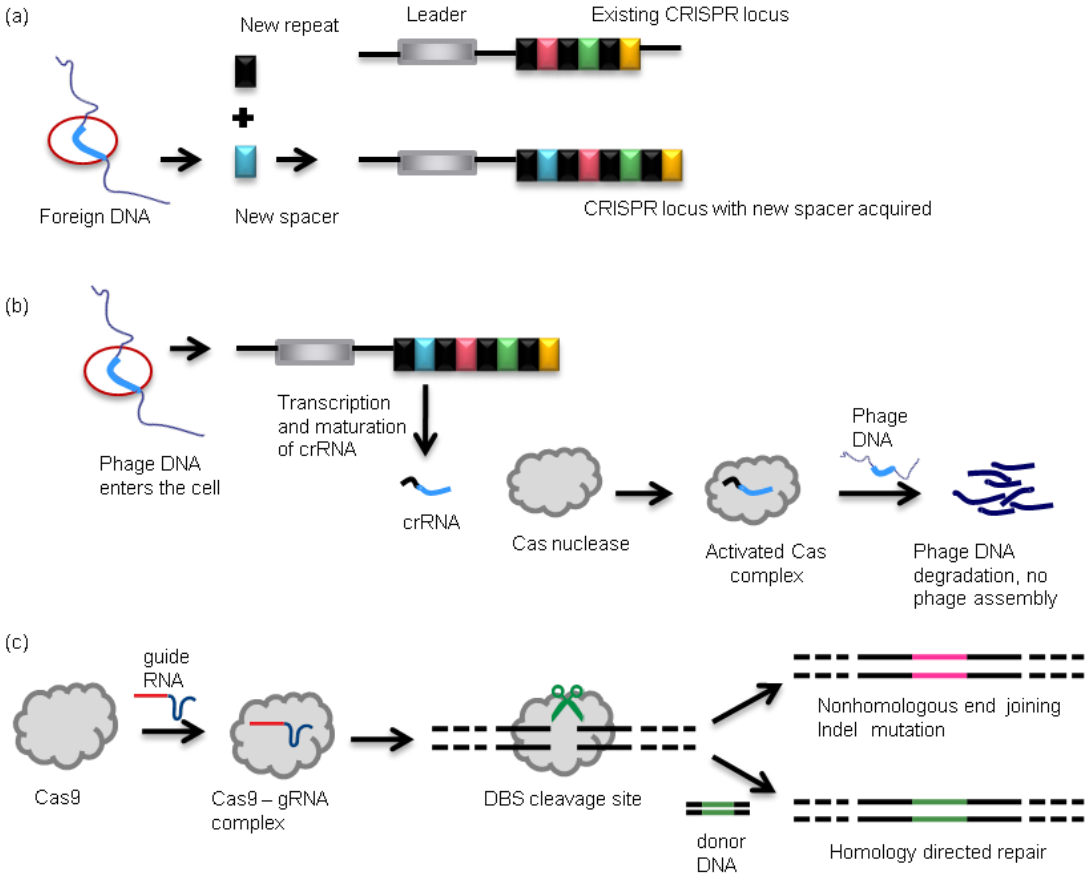


Figure 2: (a) CRISPR-Cas system of bacteria enables efficient resistance to phage attack. For example, in case of dairy lactobacilli, when the cells encounter the dairy phage for the first time, its DNA is cleaved and a sequence that includes repeater (black box) and spacer (blue box) is integrated in CRISPR-Cas locus, directly behind the leader sequence. (b) In the event of repeated attack by the same phage, its DNA sequence corresponding to an existing spacer induces transcription and maturation of CRISPR RNA (crRNA), which activates Cas complex and efficiently cleaves the foreign DNA. Further stages of phage reproduction are terminated, and there are no newly assembled phage particles. As the dairy strain combats the phage, normal fermentation process occurs. (c) CRISPR systems mechanism initiated development of genome editing tool. Here, Cas9 nuclease interacts with chimeric guide RNA, that provides the enzyme to the specific site in DNA, after which precise double stranded break (DBS) occurs. After DBS, breaks can be either nonhomologously joined leading to an indel mutation, or, in presence of a donor DNA, this sequence is precisely inserted in a homology directed repair event.

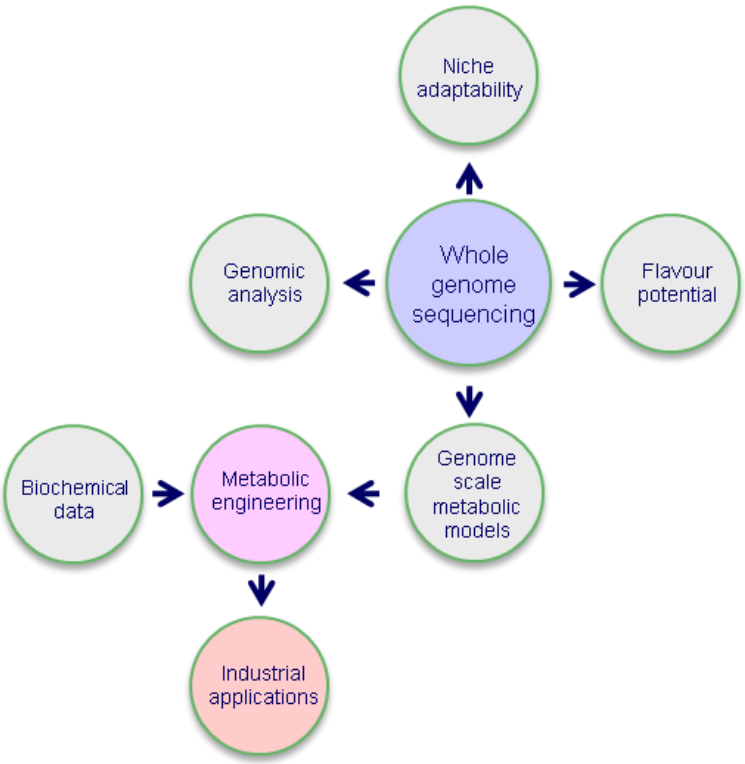


Figure 3: Schematic view of range of applications of available genome sequences. The whole genome sequencing (WGS) data provides the basis for genomic characterisation of species or genera, as well as evolutionary studies, such as niche adaptability. Insight in genetic content of a strain can predict the presence of metabolic machinery that could generate flavour compounds. Additionally, they enable the construction of genome scale metabolic models, which coupled to genetic information and biochemical data lead to the development of metabolic engineering studies. Results of these studies reveal strains capacity for plausible industrial applications.

Chapter 2

Genetic, enzymatic and metabolite profiling of the *Lactobacillus casei* group reveals strain biodiversity and potential applications for flavour diversification

All of this chapter has been published in Journal of Applied Microbiology (2017). 122, 1245-1261.
DOI: 10.1111/jam.13420

Ewelina Stefanovic chapter contributions:

Experimental work:

- Designed and performed all experimental work described in this chapter

Results interpretation:

- Analysis of all data of work described in this manuscript

Manuscript preparation:

- Major contribution to manuscript preparation

2.1 Abstract

Aims: The *Lactobacillus casei* group represents a widely explored group of lactic acid bacteria, characterised by a high level of biodiversity. In this study, the genetic and phenotypic diversity of a collection of more than 300 isolates of the *L. casei* group and their potential to produce volatile metabolites important for flavour development in dairy products was examined.

Methods and Results: Following confirmation of species by 16S rRNA PCR, the diversity of the isolates was determined by pulsed-field gel electrophoresis. The activities of enzymes involved in the proteolytic cascade were assessed and significant differences among the strains were observed. Ten strains were chosen based on the results of their enzymes activities and they were analysed for their ability to produce volatiles in media with increased concentrations of a representative aromatic, branched-chain and sulfur amino acid. Volatiles were assessed using gas chromatography coupled to mass spectrometry. Strain-dependent differences in the range and type of volatiles produced were evident.

Conclusions: Strains of the *L. casei* group are characterised by genetic and metabolic diversity which supports variability in volatile production.

Significance and Impact of the Study: This study provides a screening approach for the knowledge-based selection of strains potentially enabling flavour diversification in fermented dairy products.

Keywords: dairy, diversity, *Lactobacillus,* PFGE, proteinase

2.2 Introduction

Lactobacillus is the largest and most diverse genus of the lactic acid bacteria (LAB), and to date (Nov 2016), comprises more than 170 species (www.ncbi.nlm.nih.gov/Taxonomy/Browser). The species of this genus show remarkable niche adaptation, and have been isolated from dairy products and other fermented foods, the human and animal gastrointestinal tract and from plant material (Claesson et al., 2007). The species *Lactobacillus casei* and *Lactobacillus paracasei*, along with *Lactobacillus rhamnosus*, are referred to as the *Lactobacillus casei* group, and are regarded as closely related, both phylogenetically and phenotypically. *Lactobacillus casei* and *Lactobacillus paracasei* represent some of the best explored species within the *Lactobacillus* genus with 89 genome sequences available (Nov 2016) for these species (www.ncbi.nlm.nih.gov/genome/genomes). However, the taxonomic classification of strains is far from straightforward, as often new isolates are named as *L. casei* when they should be named *L. paracasei* since they are more closely related to *L. paracasei* type strain ATCC 334 than to ATCC 393, the type strain of *L. casei,* according to the Judicial Commission of the International Committee of Systematics of Bacteria (Tindall, 2008). This affects overall nomenclature. Strains of these two species have been isolated from all the usual niches for lactobacilli (fermented products, gastrointestinal tract, environment) (Cai et al., 2009). Their broad ecological distribution reflects their metabolic flexibility and widespread application. The niche adaptability of these two species has been explained through genomic studies where the presence and absence of certain genes important for survival in different niches (specific carbohydrate metabolism, bile salt resistance genes etc.) and comparative analysis has demonstrated the plasticity of

their genomes and their liability to evolutionary changes (Cai et al., 2009, Broadbent et al., 2012, Smokvina et al., 2013, Yu et al., 2015). Acquisition of foreign genes, mainly through horizontal gene transfer, has enabled changes in the metabolic and nutritional capacities of these species and has led to adaptation to more dynamic habitats, such as the gastrointestinal tract and plant materials. Conversely, gene decay, evident in dairy isolates, has narrowed the potential habitats and accommodated dairy niche specialisation (Makarova et al., 2006, Cai et al., 2009, Broadbent et al., 2012). The intra-species heterogeneity and the associated metabolic diversity have provided an opportunity to harness the metabolic potential of strains of the *Lactobacillus casei* group for application in a broad spectrum of fields, from health improvement to food production.

While certain strains of these species are perhaps best known for their characteristic health benefits (Sgouras et al., 2004, Herias et al., 2005, Ivory et al., 2008, Chen et al., 2014), other strains of *L. casei* and *L. paracasei* are commonly found as the dominant species of nonstarter lactic acid bacteria (NSLAB) in ripening cheese (Gobbetti et al., 2015) and are likely to play a role in the development of flavour in these products (Swearingen et al., 2001, Thage et al., 2005, Van Hoorde et al., 2010). The development of flavour results from a complex network of metabolic reactions, which include three main processes: sugar metabolism (glycolysis), lipid degradation (lipolysis), and protein catabolism (proteolysis). Although sugars, mainly lactose, and lipids can be metabolised to flavour compounds, the proteolytic process is seen as particularly important for flavour development (Smit et al., 2005). In LAB, this cascade begins with the activity of a surface proteinase, often called a cell wall, or cell envelope proteinase (CEP). The peptides produced by the activity of CEP are transported into the cell and degraded by the coordinated action of peptidases with

different, but often partially overlapping, specificities. This joint activity of peptidases is crucial for achieving the desired level of proteolysis in cheese (Stressler et al., 2013). As a result of peptidase activity, free amino acids are released. Free amino acids can directly contribute to flavour (McSweeney and Sousa, 2000), but it is their further metabolism that is seen as a key process in flavour formation (McSweeney and Sousa, 2000, Yvon and Rijnen, 2001, Rijnen et al., 2003). There are several pathways of amino acid metabolism in cheese, initiated by the activity of aminotransferases, lyases or decarboxylases (Ardo, 2006). However, the majority of the most important flavour compounds originate in transamination pathway. Aminotransferases (AT) transfer the amino group to α-keto acid (most often α-ketoglutarate) (Jensen and Ardo, 2010). Nevertheless, transamination depends on the presence of an amino group receptor, usually α-ketoglutarate, which is produced by glutamate dehydrogenase (GDH) (Kieronczyk et al., 2004), although low level of α-ketoglutarate can be produced in cheese through glutamate catabolism (Christensen et al., 1999). GDH activity has been shown to be a limiting factor for transamination (Tanous et al., 2002) and as such, indirectly represents one of the key enzymes responsible for the high flavour potential of LAB (Kieronczyk et al., 2003, Kieronczyk et al., 2004, Thage et al., 2005). The assessment of activities of enzymes of the proteolytic cascade could provide information regarding the flavour development capacity of strains.

The aim of this study was to investigate the genotypic and phenotypic diversity of strains of the *L. casei* group and to examine their potential to contribute to flavour development and diversification in dairy products. We focused on two species, *L. casei* and *L. paracasei*, as these species are most commonly associated with the non-starter flora in dairy products, and we designated the strains as belonging to *L. casei*

group. The strategy employed included defining the genomic diversity of a selected bank of strains, subsequent assessment of activities of enzymes involved in proteolysis, and finally the determination of volatile flavour production in a single amino acid-enhanced media. The diversity observed at the genetic level was borne out at the phenotypic level. These differences facilitated variations in the metabolic activity resulting in the development of diverse volatile profiles among strains.

2.3 Materials and methods

2.3.1 Bacterial strains and culture conditions

A total of 252 strains of the *L. casei* group from the Teagasc Food Research Centre DPC Culture Collection were used in this study. The strains were isolated as a part non-starter microbiota of dairy products including Cheddar, Provola, Comte, and Gouda cheeses, and from other fermented products such as yoghurt and sourdough. The strains were cultivated in MRS media (Oxoid, Basingstoke, UK) at 30°C in aerobic conditions. Strain *Lactococcus lactis* subsp. *cremoris* Wg2 was cultivated in LM17 (Merck, Darmstadt, Germany) at 30°C in aerobic conditions.

2.3.2 DNA isolation, PCR and sequence analysis of the 16S rRNA gene

For DNA isolation, the GeneElute® Bacterial Genomic DNA Kit (SigmaAldrich, St. Louis, MO, USA) was used according to the manufacturer's instruction. Subsequently, PCR amplification of the 16S rRNA gene was performed with 16S universal primers: UNI16_F: 5'-AGAGTTTGATCCTGGCTCAGG-3', UNI16_R: 5'-ACGGCAACCTTGTTACGAGTT-3' (Alander et al., 1999), which amplify nearly the entire length of 16S rRNA gene (Frank et al., 2008). PCR was performed using the following amplification conditions: initialisation at 94°C for 5 min, 40 cycles of denaturation at 94°C for 45 s, annealing at 55°C for 35 s and elongation at 72°C for 1 min, and final elongation at 72°C for 10 min (Eppendorf Mastercycler Pro, Hamburg, Germany). Amplicons of size of about 1500 bp were purified (Isolate II PCR and Gel Kit, Bioline, London, UK) and sequenced using the Sanger method (GATC Biotech AG, Koln, Germany). The sequence data generated was compared to the NCBI nucleotide database (www.ncbi.nlm.nih.gov) using the BLAST

algorithm (Altschul et al., 1990). The top BLAST hit was taken as confirmation of species.

2.3.3 Pulsed-Field Gel Electrophoresis (PFGE)

PFGE was performed as described by Simpson et al. (2002) with slight modifications. Bacterial strains were grown overnight at 30°C in MRS broth containing 20 mmol/L threonine. For each strain, 1 mL of cell suspension was centrifuged (15000 g, 5 min), washed once in 500 µL of 10 mmol/L Tris-HCl, 1 mol/L NaCl, pH 7.6 and resuspended in 200 µL of the same solution, mixed with 200 µL of 2 % low melting point agarose in 0.125 mol/L EDTA pH 7.6 and left to solidify in moulds at room temperature. Plugs were subjected to cell lysis with a mix of 10 mg/mL lysozyme and 20 units/mL mutanolysin in EC buffer (1 mol/L NaCl, 6 mmol/L Tris-HCl, 100 mmol/L EDTA, 1 % (w/v) sarkosyl, pH 7.6) for 24 h at 37°C. Subsequently, plugs were subjected to proteolysis with proteinase K (0.5 mg/mL in 0.5 mol/mL EDTA, 1 % (w/v) sarkosyl, pH 8.0) and incubated for 24 h at 55°C. Proteolysis was performed twice, and plugs were washed in 1 mmol/L phenylmethanesulfonyl fluoride (PMSF) prepared in TE 10/1 buffer (10 mmol/L Tris-HCl, 1mmol/L EDTA, pH 8.0) for 1 h at 37°C. Slices (1-2 mm) were cut from the agarose plugs and washed 3 times for 30 min at room temperature with gentle shaking in TE 10/0.1 buffer (10 mmol/L Tris-HCl, 0.1 mmol/L EDTA, pH 8.0). Slices were then incubated with 100 µL of the restriction buffer Cut Smart® (New England Biolabs, Ipswich, MA, USA) at 4°C for at least 30 min. The buffer was removed and plugs were incubated for 24 h at 37°C with *Asc*I restriction enzyme (New England Biolabs) in the same buffer. The reaction was stopped by the addition of 0.5 mL of 0.5 mol/L EDTA pH 8.0. Following digestion, slices were loaded into the wells of a 200 mL 1 % PFGE grade agarose gel (prepared in 0.5 × dilution of

TRIS borate-EDTA buffer concentrate, SigmaAldrich). The gels were run in the same 0.5 × TRIS-borate buffer using a CHEF-DR® II PFGE apparatus (Biorad, Hercules, CA, USA) at 1 V (6 V/cm) for 16 h at 14°C with the pulse ramped from 1 to 20 s. Gels were stained with ethidium bromide (0.5 µg/mL) for 1 h, and then destained in water for 1 h. Gels were photographed using Alpha Imager® 3400 (Alpha Innotech Corp, San Leandro, CA, USA).

2.3.4 Phylogenetic analysis

PFGE images were processed using BioNumerics® 7.5 software (Applied Maths, Austin, TX, USA). Dendrograms were made using the Unweighted Pair Group Method Using Average Linkage (UPGMA) distance matrix method (Sokal and Michener 1958) and curve based Pearson correlation.

2.3.5 Determination of cell envelope proteinase activity

Cell envelope proteinase (CEP) activity was determined using a modification of the method previously described by Weimer et al. (1997) and Gaudreau et al. (2005), which is based on the EnzCheck® kit Green Fluorescence E-6638 (Molecular Probes, Eugene, OR, USA). Strains were grown in 35 mL 10 % (w/v) reconstituted skim milk (RSM) for 18 h at 30°C. Cells were centrifuged (4000 g, 10 min, 4°C), and washed 3 times with 50 mmol/L Tris-HCl buffer pH 7.8 with 2 mmol/L $CaCl_2$ added. After washing, the optical density (OD_{600nm}) of cells was adjusted to approximately OD_{600}=10 in the same buffer. Components of the kit were prepared according to manufacturer's instructions. In a 96-well microplate, 100 µL of cell suspension and 100 µL of prepared BODIPY®FL casein solution were mixed and incubated for 24 h at 30°C. Fluorescence (Ex/Em 505/513 nm) was measured on a Synergy 2 reader (BioTek Multi Detection Plate Reader, Winsooski, VT, USA), using optimal filters:

485/20 nm for extinction and 528/20 nm for emission. A proteinase K solution (2 µg/mL) was used as a positive control. Enzyme activities for each strain were expressed as direct fluorescence readings. All strains were evaluated in triplicate. *Lactococcus lactis* subsp. *cremoris* Wg2 was used as a control strain. A set of trypsin standards from 0.2 ng/mL to 70 µg/mL was made and their activity was measured as for the samples.

2.3.6 Determination of aminopeptidase activities

After incubation of strains in 10 mL MRS broth at 30°C for 18 h, cells were centrifuged (4000 g, 10 min, 4°C) and washed twice with 50 mmol/L sodium-phosphate buffer pH 7.5, and resuspended in the same buffer to a final volume of 2 mL. To obtain cell-free extracts (CFE), cells were disrupted by sonication (Soniprep 150, MSE LTD, London, UK) in 5 cycles of 15 s sonication on maximum amplitude (20 amplitude microns) and 45 s of cooling on ice. Sonicated samples were centrifuged (12000 g, 10 min, 4°C) to remove cell debris.

Aminopeptidase assays were performed using a modified method for aminopeptidase activities defined by Jensen and Ardo (2010). Chromogenic substrates (L-Lys-*para*-nitroanilide (*p*NA) (Sigma-Aldrich), H-Gly-Pro-*p*NA and H-Arg-*p*NA (Bachem, Bubendorf, Switzerland) for PepN, PepX and PepC, respectively) were prepared as 1 mmol/L solutions in 50 mmol/L sodium-phosphate buffer pH 7.5. The assay mixture contained 50 µL of substrate solution and 50 µL of CFE. Absorbance was measured at 405 nm (Synergy HT, BioTek Multi Detection Plate Reader) after 30 min of incubation at 30°C. The amount of *p*-nitroaniline released was determined by including a standard curve previously obtained for standard samples of *p*-nitroaniline ranging between 0 to 50 nmol. Aminopeptidase activities were expressed as nmol of

p-nitroaniline released per min and mg of protein. No positive control was included in this assay, as no commercial enzyme was available. Blanks contained water instead of CFE. Development of yellow colour in the samples, originating from *p*-nitroaniline, and no colour development in the blank after incubation were considered as a sign of enzyme activity of CFE. Protein content was determined by using bicinchoninic acid assay (Pierce®BCA Protein Assay Kit, Thermoscientific, Walthanm, MA, USA). All strains were analysed for all selected aminopeptidase activities in triplicate.

2.3.7 Determination of aromatic aminotransferase activity

The assay to determine aromatic aminotransferase (ArAT) activity was performed by following the conversion of phenylalanine to phenylpyruvate. The assay was based on a method described in Brandsma et al. (2008) with modifications. The assay mixture contained 20 mmol/L L-phenylalanine, 10 mmol/L α-ketoglutarate, 0.5 mmol/L sodium EDTA, 0.05 mmol/L pyridoxal-5'-phsophate all dissolved in 25 mmol/L borate buffer pH 8.5. In each well, 150 µL of mixture and 100 µL of CFE (prepared as described above) were mixed, and absorbance was measured after 12 hours incubation at 30°C at 290 nm (Synergy HT, BioTek Multi Detection Plate Reader). The amount of phenylpyruvate released was determined from a standard curve obtained for a set of standards ranging from 5 to 450 nmol of sodium-phenylpyruvate. The ArAT activity was expressed as µmol of phenylpyruvate released per mg of protein. No positive control was included in this assay, as no commercial enzyme was available. Negative controls included CFE without added phenylalanine as a substrate and blanks contained water instead of CFE. Change of absorbance in samples containing CFE during incubation time and no change of absorbance in negative controls and blanks were considered as an evidence of

enzyme activity of CFE. Protein content was determined as previously described. All strains were analysed in triplicate.

2.3.8 Determination of glutamate dehydrogenase activity

The glutamate dehydrogenase assay was performed based on the principle described by Kieronczyk et al. (2003) using a modification of the Megazyme L-Glutamic Acid Kit assay (K-GLUT®, Megazyme International Ireland Ltd, Bray, Ireland). The modification involved supplementing the GDH of the kit with bacterial CFE, which allowed for quantitative determination of GDH activity of the bacterial strains. The original assay conditions and volumes were modified in order to quantify GDH activity in CFEs as follows. The final reaction mixture contained 10 µL of diaphorase, 40 µL of TEA buffer, 20 µL of glutamic acid solution (0.1 mg/mL), 20 µL of INT-NAD^+ solution (all of these supplied in K-GLUT® Kit) and 100 µL of CFE. Absorbance was measured at 492 nm after 1 h of incubation at 37°C (Synergy HT, BioTek Multi Detection Plate Reader). One unit of GDH activity corresponded to the amount of enzyme that resulted in an increase of absorbance of 0.01 per 1 min. No positive control was included in this assay. Blank contained water instead of CFE. The development of red colour product (INT-formazan) in samples and no colour development in the blank after incubation were considered as an evidence of enzyme activity of CFE. Protein content was determined as previously described. Specific enzyme activity was expressed as the number of units (U) per mg of protein. All strains were analysed in duplicate.

2.3.9 Production of volatile compounds from single amino acid metabolism

In order to evaluate the metabolic activity of selected strains in the presence of a predominance of a single amino acid, a set of specific media was prepared. The

media contained 50 g/L of Bacto®Tryptone (BD, Oxford, UK), 12 g/L NaCl and the specific selected amino acid added to final concentration of 50 mmol/L. The amino acids chosen for this analysis were phenylalanine, leucine and methionine to demonstrate the metabolic activity of strains towards aromatic, branched-chain and sulfur amino acids, respectively. The corresponding media were designated as PEM (phenylalanine-enhanced media), LEM (leucine-enhanced media) and MEM (methionine-enhanced media), and for this analysis, ten strains were chosen based on the results of all the enzymatic assays described above. Strain DPC1116 had the highest activity of PepN, one of the highest PepX activities, high PepN activity and medium CEP, AT, and GDH activities. Strain DPC2068 had medium CEP activity, high PepN, PepC and PepX activities, medium ArAT activity and low GDH activity. Strain DPC2071, had the highest CEP activity, high PepN, PepC activities, medium PepX and ArAT activities and low GDH activity. Strain DPC3990 had high CEP activity, high PepN, PepC activities, medium PepX activity and low ArAT and GDH activities. Strain DPC4026 had low CEP activity, medium PepN and PepC activities, low PepX, ArAT and GDH activities. Strain DPC4206 had high activities in all enzyme assays. Strain DPC4536 had low CEP activity, medium activities for PepN, PepC and PepX and ArAT assays, but the highest activity of GDH. Strain DPC5408 had high CEP, medium PepN, PepC and PepX activities, and low activities of ArAT and GDH. Strain DPC6753 had high activity of CEP, low activities of PepN, PepC and PepX, and high activities of ArAT and GDH. Strain DPC6800 had low CEP activity, high PepN and PepC activities, and the highest PepX activity, high ArAT and medium GDH activity. Cells were prepared according to the protocol described by Van de Bunt et al. (2014) with some modifications. Briefly, strains were incubated overnight (30°C) in 10 mL of MRS, and were re-inoculated (1 % v/v) into

500 mL of MRS and incubated for 24 h at 30°C until they reached stationary phase. Cells were centrifuged (4000 g, 10 min, 4°C), washed twice with 0.1 mol/L sodium phosphate buffer pH 6, and resuspended in 5 mL of the same buffer containing 15 % glycerol and kept at -80°C until required. Thawed cell suspensions (1 mL) were inoculated into 9 mL of prepared each of three media described above (PEM, LEM, MEM). Three replicate vials were made for each strain. For one replicate of each strain, cell counts (CFU/mL) were performed. A 100 µL aliquot was taken for plate counting at t=0 h and the strains were incubated for 48 h, after which another 100 µL aliquot was taken for plate counting (t=48 h). Plate counting was performed on MRS agar plates, which were incubated aerobically for 72 h at 30°C. Samples were kept in -20°C until required for volatile analysis. The control consisted of un-inoculated media. The samples and the control were tested in triplicate.

2.3.10 Head-Space Solid Phase Microextraction Gas Chromatography-Mass Spectrometry analysis (HS-SPME GC-MS)

For each sample of each of three media, 2 g of the sample was placed in an amber 20 mL screw capped HS-SPME vial with a silicone/PTFE septum (Apex Scientific, Maynooth, Ireland). The vials were equilibrated to 40°C for 10 min with pulsed agitation of 5 seconds at 500 rpm using a heated stirrer module on a Shimadzu AOC 5000 plus autosampler. A single 50/30 µm Carboxen®/ divinylbenzene/ polydimethylsiloxane (CAR/DVB/PDMS, Agilent Technologies, Cork, Ireland) fibre was used to perform solid phase microextraction (SPME). The SPME fibre was exposed to the headspace above the samples for 20 min at 40°C. After extraction, the fibre was injected into the GC inlet *via* a merlin microseal and desorbed for 2 min at 250 °C into a SPL injector with a SPME liner. Injections were made on Shimadzu 2010 Plus GC with a DB-5 (60 m × 0.25 mm × 0.25 µm, Agilent Technologies)

column in splitless mode using a split/splitless injector. Helium was used as a carrier gas, which was maintained at 23 psi. The temperature of the column oven was set at 35°C, held for 5 min, increased at 6.5°C/min to 230°C then increased at 15°C/min to 320°C, yielding at total GC run time of 41.5 min. The mass spectrometer detector Shimadzu TQ8030 was run in single quad mode. The ion source temperature was 230°C, the interface temperature were set at 280°C and the MS mode was electronic ionization (-70 eV) with the mass range m/z scanned between 35 and 250 amu.

All samples were analysed in the same GC run. A set of external standards (dimethyl-sulfide, benzaldehyde, cyclohexanone, butyl acetate, acetone, and ethanol at concentrations of 10 ppm) was also run at the start of the sample set to ensure that both the HS-SPME extraction and MS detection were within specification. Blanks (empty vials) were injected regularly to monitor possible carry over. The SPME fibre was cleaned between samples using a bake-out station on the AOC 5000 at 270°C for 3 min to ensure no carry over between samples.

2.3.11 Data processing and compound identification

Chromatograms obtained by GC analysis were converted to .cdf format and processed by TargetView® (Markes International Ltd, Llantrisant, UK). Compounds of interest were chosen according to previously published reviews of flavour contributing compounds (Curioni and Bosset, 2002, Singh et al., 2003, Smit et al., 2005). Identification of compounds was based on the results of a comparison with the NIST 2011 Mass Spectral Library (Scientific Instrument Services, Ringoes, NJ, USA) and an in-house library for flavour compounds and confirmed by calculating linear retention indices as described in (Van den Dool and Kratz, 1963).

2.3.12 Statistical analysis

All enzymatic assay results were statistically analysed using one-way Analysis of Variance (ANOVA) followed by the least significant difference (LSD) *post hoc* test. ANOVA and LSD were also used for testing the significance of differences in cell counts in three media (PEM, LEM and MEM) at t=0 h and t=48 h. Principal component analysis (PCA) was used for analysis of GC-MS data. All listed statistical tests were performed in R statistical software (R Core Team 2015, R Foundation for Statistical Computing, Austria, www.r-project.org).

2.4 Results

2.4.1 Origins of *L. casei* group strain bank and confirmation of species

Initially, 310 isolates were selected for analysis in this study. The isolates had previously been assigned as *Lactobacillus casei, L. paracasei* or simply as *Lactobacillus* isolates in the Teagasc DPC Culture Collection and they originated from dairy products or sourdough. Each of the 310 isolates was subject to 16S rRNA sequencing, and based on the BLAST analysis of the amplified sequences, 252 out of 310 isolates were confirmed as either *L. casei* or *L. paracasei*.

2.4.2. Comparative phylogenetic analysis reveals extensive genomic diversity in *L. casei* group strain bank

To assess the diversity of the 252 confirmed isolates belonging to the *L. casei* group, PFGE was used to generate genomic fingerprints. Grouping of the distinct strains was performed by comparing PFGE fingerprints with BioNumerics®7.5 software, but also by a simple visual comparison, as some of fingerprints were distant in the dendrogram but very similar when checked manually. Analysis of the generated PFGE patterns revealed 98 distinct profiles among the 252 isolates, representing 98 distinct strains. Figure 1a represents a dendrogram of the PFGE profiles of the 98 diverse strains. Additionally, strain DPC4536, that has indistinguishable fingerprint to strain DPC4206 was included in further enzyme activity evaluation, to observe the potential phenotypic differences between these two strains (Fig. 1b). This means that in total, 99 strains were analysed in enzymatic assays.

2.4.3 Key enzymatic assays show diverse activities among selected strains

To assess the cell envelope proteinase activity (CEP) of the strains, a kit based on the proteolysis of BODIPY®FL-labelled casein derivatives which release highly fluorescent peptides, was used. This assay is based on the principle that the measured increase in fluorescence is proportional to the proteinase activity. All 99 strains showed CEP activity, but the levels varied significantly from strain to strain. The CEP activity was expressed as measured fluorescence and it ranged from 80.3 arbitrary fluorescence units for strain DPC4764 to 229.7 arbitrary fluorescence units for strain DPC2071 (Fig. 2), which corresponded to fluorescence measured when standard solutions of trypsin in the range of 0.4 to 4.0 µg/mL were used (data not shown). *Lactococcus lactis* subsp. *cremoris* Wg2 was used as a CEP-positive control strain, having been confirmed as such in a previous study (Kok et al., 1988, Laan and Konings 1989, Nikolic et al., 2009), and its CEP activity was 205.33 arbitrary fluorescence units. The strains DPC2071, DPC4206, DPC3990, (Fig. 2) had similar or even higher activity than the control strain *Lactococcus lactis* subsp. *cremoris* Wg2 and in total, 25 strains displayed activity that was not statistically different to the activity of *Lactococcus lactis* subsp. *cremoris* Wg2 (Supporting Information Table 1).

The strains showed significantly different activities towards the various aminopeptidase substrates tested (L-Lys-*para*-nitroanilide (*p*NA), H-Gly-Pro-*p*NA and H-Arg-*p*NA for PepN, PepX and PepC, respectively). PepN activities ranged from 0 to 54.2 nmol *para*-nitroaniline/(min*mg protein) for DPC5336 and DPC1116 respectively; PepC activities ranged from 0 for strains DPC4139, DPC4140 and DPC5410 to 50.3 nmol/(min*mg protein) for DPC4680, and PepX activities from 0 (31 strains) to 39.2 nmol/(min*mg protein) for DPC6800 (Fig. 3, Supporting

Information Table 1). Statistical analysis confirmed that aminopeptidase activities differed significantly between the strains. For example, for PepX, 39 strains showed activity below 5 nmol/(min*mg protein), and therefore only eight of these strains were carried forward for further analysis of enzyme activities.

When aromatic AT activity was measured, strain DPC5411 showed the highest activity with 3.28 µmoles of phenylpyruvate released per mg protein while DPC4805 had the lowest activity of 0.25 µmol/mg protein. Statistical analysis revealed significant differences among the strains for aromatic AT activity (Fig. 4). The GDH activity ranged from 0 for strains DPC6084, DPC4802 and DPC4026 and 17.5 U/mg of protein for strain DPC4536 and differences in GDH activities were shown to be significant (Fig. 5).

2.4.4 HS-SPME GC-MS volatile analysis confirms metabolic diversity of selected strains

Strains DPC1116, DPC2068, DPC2071, DPC3990, DPC4026, DPC5408, DPC6753, DPC6800, and two strains which have indistinguishable genetic fingerprints (DPC4206 and DPC4536) were selected as candidates for volatile analysis on the basis of their spectrum of key proteolytic enzyme activities. The strains were assessed for their capacity to metabolise amino acids in three distinct media containing elevated levels of a single amino acid (phenylalanine (PEM), leucine (LEM) or methionine (MEM)). The enumeration of cells in these media, pre- and post-incubation, is presented in Supporting Information Table 2. In all three amino acid-enhanced media, six of the ten strains (DPC1116, DPC2071, DPC4206, DPC4536, DPC5408, DPC6800) showed significantly lower cell numbers after 48

hours incubation. In the case of strain DPC4026, significantly lower cell number after 48 hours of incubation was observed in only MEM.

Compounds selected as flavour-contributing in samples in all three media are listed in Table 3 of Supporting Information. The compounds of interest were selected according to the previously published reviews of compounds considered as main flavour contributors in cheese (Curioni and Bosset, 2002, Singh et al., 2003, Smit et al., 2005). The highest number of the relevant volatile compounds (47) were present in PEM media, with metabolites containing aromatic ring structures, such as benzeneacetaldehyde, 3-ethyl-benzaldehyde, 1,3-xylene, tetramethyl-benzene, hexyl-benzene and methyl-naphthalene exclusively present in this medium. Acetic and butanoic acids and long-chain ketones (C7-C13) were also only detected in PEM. In addition, PEM samples had the highest number of alcohols, aldehydes, and ketones. Of the 47 compounds selected, significant differences in the relative abundance of 27 compounds were observed. In LEM samples, 25 volatiles were detected, and 24 were present in significantly different amounts for various strains. In MEM samples, 22 volatiles were detected and significant differences in abundances of all 22 compounds were observed between the strains tested. No specific metabolites, present in LEM or MEM exclusively, were identified.

The relative abundances of compounds for which significant differences among the strains, including the control, were observed, are presented in Figure 6. In all three media, 1-butanol was present at the highest abundances compared to all other volatiles (no significant difference between the strains including the control, in PEM), and in PEM and MEM, the abundance of this compound was the highest in the control. Several strains showed unique abilities to produce certain volatiles. Strain DPC2068 produced significantly higher amounts of butyl-3-methyl butanoate,

a metabolite of leucine, in PEM compared to all other strains, while in MEM and LEM, this strain was the only producer of this compound. The same observation was made in the case of butyl-2-methyl propanoate (metabolite of leucine), butyl butanoate, butyl propanoate (secondary metabolites of amino acids) in all three media. In MEM and LEM, this strain produced a significantly higher abundance of dimethyl-trisulfide (DMTS), originating from methionine, compared to all other strains. An interesting observation was made for diacetyl and acetoin, important flavour compounds originating from sugar metabolism. Strain DPC4026 was the only producer of diacetyl in both LEM and MEM, and in PEM the production by DPC4026 was significantly higher compared to all other strains. DPC4026 also produced the highest abundance of acetoin in PEM and LEM, and significantly higher abundance of acetoin compared to all other strains in MEM. Strain DPC4206 produced the highest abundance of dimtehyl-disulfide (DMDS) and methanethiol, methionine metabolites, in both MEM and LEM, and the abundance of methanethiol in LEM was significantly higher compared to all other strains. In LEM, this strain produced the highest amount of 3-methyl-butanal (leucine metabolite), while in PEM it produced the highest levels of butanoic acid. Strains DPC4206 and DPC1116 produced the highest abundances of 3-methyl-butanol (leucine metabolite) in all three media, and DPC1116 produced significantly higher abundance of this compound in MEM compared to all other strains. Strain DPC6800 produced the highest abundance of the two main compounds arising from phenylalanine metabolism, benzaldehyde and benzyl-alcohol, and the abundance of benzyl-alcohol was significantly higher compared to all other strains in PEM.

PCA plots obtained after analysis of total ion chromatograms for ten chosen strains in three media (PEM, LEM and MEM) are presented in Figure 7. PCA plots were

generated by using only those compounds (variables) for which significant differences in abundance among the strains, including the control, were observed. In the case of PEM samples, dimension (PC1) described 26.1 % variation and dimension 2 (PC2) described 22.0 % total variation between the strains. In the PCA plot for LEM samples, PC1 described 32.2 % of variation, while PC2 described 26.6 % variation between the strains. In the PCA plot for strains inoculated in MEM, PC1 described 33.8 % variation, and PC2 described 28.1 % variation. The control clearly separated from all strains tested on plots for PEM and MEM, while on LEM plot, the control was positioned in the central part of the plot, along with the majority of other strains. In all three plots, two strains, DPC2068 and DPC4206, were positioned more separately from other strains and the control. The position of DPC2068 was determined by the relative abundance of butyl-2-methyl propanoate, butyl-3-methyl butanoate, butyl propanoate and butyl butanoate, in all plots, and additionally by 3-methyl-butyl acetate and butyl butanoate in PEM and MEM, and DMTS in MEM. The position of strain DPC4206 was associated with DMDS, butanoic acid and ethyl benzene in PEM, 3-methyl-butanal, 3-methyl-butanoic acid, 3-methyl-butanol, DMDS and methanethiol in LEM and DMDS, 3-methyl-butanol, acetone, butanone and methanethiol in MEM.

2.4.5 Two strains with the same genomic fingerprint show different phenotypic characteristics

In addition to the 98 strains with diverse genomic fingerprints, strain DPC4536, that has an indistinguishable PFGE pattern from strain DPC4206 (Fig. 1b) was included in the enzymatic assays. Different activities were determined for all enzymes analysed and significant differences in activities were observed for CEP and GDH (Fig. 2-5, Supporting Information Table 1).

These two strains were also compared for their abilities to produce volatile compounds in the three amino acid-enriched specific media. The volatile profiles differed in types and abundances of compounds detected (Fig. 6), with specificities of DPC4206 metabolic characteristics described in the above section. The differences in the metabolite production is also visible in PCA plots (Fig. 7), where DPC4536 is located closer to other strains, while DPC4206 is one of two strains that were the most separated from the other strains.

2.5 Discussion

Lactobacillus is by far the largest and most diverse genus of LAB, with *L. casei* and *L. paracasei* being some of the best explored species of this genus (Broadbent et al., 2012). The genome evolution was elucidated in comparative genome analysis and it explained the adaptation of this species to numerous dynamic, nutritionally variable environments, such as gut, plant and milk. The wide ranging habitats of these species make them relevant subjects for research on genetic diversity and niche expansion (Broadbent et al., 2012).

In the literature, interchangeable use of the names *L. casei* and *L. paracasei* occurs. Any newly isolated strains are often designated as *L. casei*, but they should be named *L. paracasei*, according to current nomenclature and defined type strains (Tindall, 2008). Several studies aimed to distinguish *L. casei* and *L. paracasei*, and confirm whether they constitute one or two species (Dellaglio et al., 1991, Dicks et al., 1996, Dellaglio et al., 2002). Attempts have been made to introduce novel approaches for differentiation of *L. casei, L. paracasei, L. zeae* and *L. rhamnosus* such as the use of species specific 16S primers (Ward and Timmins, 1999), and Temporal temperature gradient gel electrophoresis (TTGE) of obtained 16S rRNA PCR amplicons (Vasquez et al., 2001). Additionally, combining results of restriction endonuclease analysis of total DNA, TTGE of 16S rRNA PCR and ribotyping (Vasquez et al., 2005) showed that numerous heterogeneities found in 16S rRNA genes in *L. casei/paracasei* and related species *L. zeae* and *L. rhamnosus* make definitive separation of these species complex and difficult. Recently, a new method based on high resolution melting analysis of PCR amplicons obtained with sets of species discriminating primers (Iacumin et al., 2015) has been proposed. Although

for the majority of the 194 strains tested, this method gave satisfactory results in species identification. However, six strains showed inconsistencies in identification. In the case of strains from the DPC culture collection, analysis with strain-specific primers did not enable satisfactory differentiation of the two species (data not shown) and to that end, a more general approach that involved conventional 16S rRNA sequencing was used to confirm the isolates as part of the *L. casei* group. In addition, the strain genomic profiling was assessed and 98 distinct PFGE profiles were detected, confirming the variability of the isolates. Surprisingly, although the strains originated mainly from cheese, the observed level of diversity was high. The reason behind this could be that the isolation source does not necessarily match the original habitat, as they could be part of the NSLAB flora or starter mixture used during cheese manufacture, or come from other sources during cheese handling and ripening, e.g. contamination from personnel, surfaces etc. Besides that, isolates were collected over a period of more than 20 years, and this time-span contributed most likely to the variety of isolates available in the DPC culture collection.

It is envisaged that differences in genomic structure present a natural basis for genetic variation among the strains, which would further facilitate strain-to-strain variation in their phenotypic characteristics, including their flavour-forming ability. *L. casei* and *L. paracasei* are often part of NSLAB flora of cheese during ripening, and are seen as important contributors to flavour development, due to the metabolic capacity of these strains that balances the degradation of substrates present and leads to formation of volatiles affecting organoleptic characteristics of cheese (Banks and Williams, 2004, Sgarbi et al., 2013).

The metabolic activity of microorganisms present in cheese during ripening results in development of flavour compounds (Marilley and Casey, 2004), and the metabolic

products of three biochemical pathways: glycolysis, lipolysis and proteolysis are seen as cheese flavour contributors (McSweeney and Sousa, 2000). Lactate is the main metabolite of the primary milk sugar lactose, and it can be metabolised to flavour compounds such as diacetyl, acetoin, acetaldehyde, or acetic acid, *via* a pyruvate intermediate (Smit et al., 2005). Lipolytic reactions during cheese ripening result in free fatty acid production, and short and intermediate-chain fatty acids either contribute to flavour themselves, or represent the starting molecules for the production of other flavour compounds (Collins et al., 2003). Although products of both glycolysis and lipolysis can contribute to flavour, in bacterial ripened cheeses catabolic products of proteolytic reactions, mainly metabolites of free amino acids, represents the major flavour contributing metabolic pathway (Smit et al., 2005), and because of this, activities of the enzymes of the proteolytic cascade were assessed.

The activity of the CEP, which cleaves casein molecules to shorter peptides, was shown to vary from strain to strain. This enzyme has an important role in flavour development, as casein hydrolysis is a cascade process and greater proteolytic activity in the earlier steps will result in the exponential generation of more flavour compounds in later metabolic steps. Various approaches have been developed to determine CEP activity, such as ones based on absorbance measurement after cleavage of chromogenic substrate (Fernández de Palencia et al., 1997, Hebert et al., 2008), or fluorescent measurement after degradation of fluorescently labelled caseins such as used in this study (Weimer et al., 1997, Wakai et al., 2013). Our results show that CEP activity of strains of the *L. casei* group is quite variable. These results correspond to the results obtained by Weimer et al. (1997), which demonstrated inter- and intraspecies differences in CEP activities for *Lactococcus lactis, L. casei* and *L. helveticus.*

The next step of proteolysis, the degradation of peptides to free amino acids is catabolised by aminopeptidases, some of which cleave only one type of amino acid and have a very narrow activity (glutamyl-aminopeptidase, PepA), while others, such as general aminopeptidases break the bond between various amino acids at the N-terminal end of the peptide (general aminopeptidases PepN, PepC) (Magboul and McSweeney, 1999). Besides these, important enzymes in efficient peptidolysis are dipeptidyl aminopeptidases, such as PepX, as they remove proline residues thus enabling further degradation of proteins (Stressler et al., 2013). Gonzalez et al. (2010) demonstrated that cell-free extracts of *L. paracasei* isolated from cheese exhibited the highest activity towards Ala-, Lys-, Pro-, and Leu-*p*NA substrates compared to other LAB isolates which included leuconostocs, lactococci and enterococci. Similar findings were reported by Ayad et al. (2004) and Macedo et al. (2000), where strains of *L. paracasei* were shown to have the highest aminopeptidase activities compared to leuconostoc, lactococcal or enterococcal strains. In this study, strain dependent activities of *L. casei* group were confirmed.

A study conducted on a series of lactococcal strains revealed that the activity of aromatic AT resulted in a more diverse volatile profile than the activity of branched-chain amino acid AT (Rijnen et al., 2003). On the other hand, the specificity of AT towards a certain type of amino acid is not absolute as shown in lactococcal strains where aromatic AT was able to degrade aromatic amino acids, but also methionine, a sulfur-containing amino acid (Rijnen et al., 2003) and leucine, a branched-chain amino acid (Christensen et al., 1999). With this in mind, determination of aromatic AT activity was seen as a suitable test for general aminotransferase activity determination. The results obtained in this study for strains of the *L. casei* group illustrate a diverse range of aromatic aminotransferase activities, which is one of the

crucial steps for the diversity of flavour compounds produced and confirms strain-specificity of the enzyme activity previously reported (Thage et al., 2004). Similar results to the ones from this study were reported for *L. helveticus* and *'L. danicus'* (related to *L. wasatchensis*, (Oberg et al., 2016)) grown in MRS (Jensen and Ardo, 2010, Pedersen et al., 2013).

Transamination, the transfer of an amino group from an amino acid to a keto acid, is enhanced when α-ketoglutaric acid is present as the amino group acceptor. Strains possessing glutamate dehydrogenase (GDH), that converts glutamate to α-ketoglutarate, are more likely to have an impact on flavour (Williams et al., 2006). Activity of GDH can depend on NAD or NADP as cofactors. Previously, Kieronczyk et al., (2003) reported that *L. paracasei* strains (INF15D, 2756) and *L. casei* 1244 did not possess NAD-dependent GDH activity, while *Lactococcus lactis* subsp. *cremoris* NCDO763 expressed low activity. Similarly, no NAD-GDH activity was observed for any of the NSLAB lactobacilli (*L. plantarum, L rhamnosus, L. parabucknerii, L. curvatus*) (De Angelis et al., 2010) or for *L. plantarum* or *L. paracasei* strains in the study of Tanous et al. (2002), but it was detected in the case of *Lactococcus lactis* subsp. *lactis* NCDO 1867. Conversely, Williams et al. (2006) showed that several *Lactobacillus* species possessed both NAD and NADP dependent GDH activity. In the present study, NAD-GDH activity was detected in all strains and the activities were higher than those obtained for NADP-GDH activity by De Angelis et al. (2010), yet they correspond to those obtained by Kieronczyk et al. (2003) for NADP-activity of GDH in *L. paracasei* INF15D.

All of the enzyme assays performed showed that strains in our culture collection bank expressed a range of activities of proteolytic enzymes supporting the diversity observed on the genetic level. Statistical analysis revealed groups that differed in

activities in each enzymatic assay, confirming the significance of the observed activity variations.

Metabolism of amino acids is considered as the most important process contributing to the development of flavour compounds (Yvon and Rijnen, 2001). These compounds include short-chain acids, alcohols and aldehydes, often with an aliphatic branch (volatiles originating from leucine, valine, isoleucine metabolism), as well compounds with aromatic ring structures originating from phenylalanine or tyrosine, and sulfur-containing compounds, which are the product of methionine metabolism (Yvon and Rijnen, 2001, Singh et al., 2003). Some of the most important flavour-forming compounds that arise from phenylalanine metabolism are benzaldehyde (almond flavour), phenylethanol (rose flower) (Curioni and Bosset, 2002) and benzeneacetaldehyde (bitter almond aromatic flavour) (Jung et al., 2013). The leucine metabolites, such as 3-methyl-butanol, 3-methyl-butanal and 3-methyl-butanoic acid, have malty, fresh cheese and rancid-sweet odours, respectively (Smit et al., 2005). Volatile sulfur compounds are important in overall cheese flavour. They include DMDS, DMTS and methanethiol which are described as having an onion, garlic and cabbage odour, respectively (Singh et al., 2003, Yvon, 2006). Secondary products, such as esters, also contribute to flavour, mainly with sweet fruity notes. For this reason, diversity in activity of amino acid converting enzymes was assessed. The base of the media used was pancreatic digest of casein, which contained all of the amino acids ranging between 0.4 % for aspartic acid up to 5.5 % for lysine, expressed as % of free amino acids. This medium was modified by addition of a single predominant amino acid, to explore metabolic preferences of the strains. The strains for these assays were selected on the basis of their varying enzyme activities, but most importantly, their different aminotransferase activities

(Supporting Information Table 1). Since amino acids dominated as the metabolic precursors in these media, as expected, the volatiles identified included the breakdown products typical of amino acid catabolism, confirming the general amino acid metabolic activity of the strains.

Figure 6 presents the relative abundances of compounds identified in samples of all three media in significantly different abundances among the strains, including the control. The overall trend in these experiments was that while more metabolites were present in samples of PEM, many of these compounds were not detected in LEM and MEM. In addition, the relative abundances of compounds were higher, often significantly higher, in PEM, compared to MEM and LEM (data not shown); however, the reason for this observation is not clear. In samples of all three media, 1-butanol was present in the highest abundances. This compound most probably originated from components of media generated during sterilisation. Butanol present in such a high abundances was the substrate for butyl esters formation in samples of all three media. Long-chain ketones were detected particularly in PEM. This could mean that in PEM a greater ratio of cell division/lysis occurred, leading to more long-chain ketones, normally of fat origin, and in this fat free medium which contained no added fats probably came from the metabolism of lipids released from the cell membrane after cell lysis. Interestingly, 2,3-butanedione (diacetyl) was detected in LEM and MEM, and 3-hydroxy-butanone (acetoin), was detected in all three media (no significant difference in PEM). These two compounds are important flavour contributors (buttery flavour) and they most probably originated from low level of sugar present in Bacto®Tryptone itself, as no sugar source was added during media preparation. In LEM and MEM, only DPC4026 produced 2,3-butanedione, suggesting a potential energy source for this strain. Among the ten strains that were

analysed in PEM, LEM and MEM, strain DPC6800 showed the highest activity of aromatic AT as determined by the *in vitro* assay, and this strain was confirmed as the highest producer of the most important molecules arising from phenylalanine, such as benzaldehyde and benzyl-alcohol. However, ethyl-benzene, 1,3-xylene, tetramethyl-benzene, etc. were present in higher concentrations in samples of other strains, and they probably emerged in further degradation of phenylalanine and its metabolites.

Principal component analysis (PCA) was used to emphasize variation and determine strong patterns in the datasets. PCA identified differences in total volatile production between the strains, and revealed, two strains, DPC2068 and DPC4206, that differed in metabolic potential compared to other strains in all three media (Fig. 7). The position of strain DPC2068 in the PCA plots from all three media was predominately due to its association with metabolites originating from the metabolism of branched-chain amino acids. Apparently, this strain has high activity towards branched-chain amino acids even in the media with less availability of these and abundances of other amino acid, such as PEM and MEM. On the other hand, strain DPC4206 has the most diverse metabolic activity compared to other strains in all three media, and it was able to produce the broadest range of volatiles, often of the highest relative abundance (Fig. 6). These two strains showed outstanding volatiles patterns and their metabolic activity could lead to diverse flavour development in fermented dairy products.

Interestingly, strains DPC4206 and DPC4536, which have the same PFGE fingerprints, were shown to have considerably different phenotypic characteristics, based both on the enzymatic assays and volatile analysis (Fig. 6 and 7). The PFGE analysis for these two strains was performed with additional enzymes (*Apa*I, *Cla*I,

and combination of both *Apa*I and *Cla*I), as well as RAPD PCR, and no differences in band patterns for these two strains was observed (data not shown). Similar results were obtained during analysis of *Listeria monocytogenes* strains in our laboratory (Fox et al., 2017). These findings highlight that strains sharing the same PFGE pattern do not necessarily have same genetic and, subsequently, phenotypic characteristics. This study showed that PFGE is not a definitive tool to determine strains genetic diversity, but rather a robust method used for assessing differences in genomic structure and observing larger evolutional events, such as large insertions, deletions and rearrangements of DNA (Cai et al., 2007). Although used as a golden standard for assessing strain diversity based on whole genome restriction analysis for long time, nowadays PFGE has slowly being replaced by the whole genome sequencing, which enables deeper insight in gene content differences among the strains and becomes a preferable method to record subtle genetic differences which would not be apparent in the PFGE profiles alone. For this reason, whole genome sequence analysis of the two strains, DPC4206 and DPC4536 is currently performed in our laboratory, in order to reveal the genetic basis of different phenotypic characteristics of these two strains observed in this study.

The study presented in this paper gives an insight into both the genetic and phenotypic diversity of strains of *L. casei* group. The observed level of genetic diversity can be considered as very broad, since the majority of isolates have the same origin of isolation. The analysed strains, including the two strains with the identical genetic fingerprint, showed variable phenotypic traits, as observed in assays determining the activities of proteolytic cascade enzymes. Additionally, the strains demonstrated different capacities for production of flavour compounds from amino acids, and two strains, DPC2068 and DPC4206, were particularly diverse in their

volatiles production. It can be inferred that strains of *L. casei* group have different abilities for volatile production, which makes them potentially useful for dairy product flavour diversification.

2.6 Acknowledgments

The authors thank to Douwe van Sinderen for providing the strain *Lactococcus lactis* subsp. *lactis* Wg2. Also, the authors are grateful to Anne Thierry for assistance with R software for statistical analysis. Ewelina Stefanovic is in receipt of Teagasc Walsh Fellowship.

2.7 References

Alander, M., Satokari, R., Korpela, R., Saxelin, M., Vilpponen-Salmela, T., Mattila-Sandholm, T. & 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.

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

Ardo, Y. 2006 Flavour formation by amino acid catabolism. *Biotechnol Adv* 24**,** 238-242.

Ayad, E. H. E., Nashat, S., El-Sadek, N., Metwaly, H. & El-Soda, M. 2004. Selection of wild lactic acid bacteria isolated from traditional Egyptian dairy products according to production and technological criteria. *Food Microbiol,* 21**,** 715-725.

Banks, J. M. & Williams, A. G. 2004. The role of the nonstarter lactic acid bacteria in Cheddar cheese ripening. *Int J Dairy Technol,* 57**,** 145-152.

Brandsma, J. B., Floris, E., Dijkstra, A. R. D., Rijnen, L., Wouters, J. A. & Meijer, W. C. 2008. Natural diversity of aminotransferases and dehydrogenase activity in a large collection of *Lactococcus lactis* strains. *Int Dairy J,* 18**,** 1103-1108.

Broadbent, J. R., Neeno-Eckwall, E. C., Stahl, B., Tandee, K., Cai, H., Morovic, W., Horvath, P., Heidenreich, J., Perna, N. T., Barrangou, R. & Steele, J. L. 2012. Analysis of the *Lactobacillus casei* supragenome and its influence in species evolution and lifestyle adaptation. *BMC Genomics,* 13**,** 533.

Cai, H., Rodriguez, B.T., Zhang, W., Broadbent, J.R. & Steele, J.L. 2007 Genotypic and phenotypic characterization of *Lactobacillus casei* strains isolated from different ecological niches suggests frequent recombination and niche specificity. *Microbiology* 153**,** 2655-2665.

Cai, H., Thompson, R., Budinich, M. F., Broadbent, J. R. & 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.

Chen, P., Zhang, Q., Dang, H., Liu, X., Tian, F., Zhao, J., Chen, Y., Zhang, H. & Chen, W. 2014. Antidiabetic effect of *Lactobacillus casei* CCFM0412 on mice with type 2 diabetes induced by a high-fat diet and streptozotocin. *Nutrition,* 30**,** 1061-1068.

Christensen, J. E., Dudley, E. G., Pederson, J. A. & Steele, J. L. 1999. Peptidases and amino acid catabolism in lactic acid bacteria. *Ant Leeuwenhoek,* 76**,** 217-246.

Claesson, M. J., van Sinderen, D. & O'Toole, P. W. 2007. The genus *Lactobacillus*-a genomic basis for understanding its diversity. *FEMS Microbiol Lett,* 269**,** 22-28.

Collins, Y. F., McSweeney, P. L. H. & Wilkinson, M. G. 2003. Lipolysis and free fatty acid catabolism in cheese: a review of current knowledge. *Int Dairy J,* 13**,** 841-866.

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.

De Angelis, M., Calasso, M., Di Cagno, R., Siragusa, S., Minervini, F. & Gobbetti, M. 2010. NADP-glutamate dehydrogenase activity in nonstarter lactic acid bacteria: effects of temperature, pH and NaCl on enzyme activity and expression. *J Appl Microbiol,* 109**,** 1763-1774.

Dellaglio, F., Dicks, L. M. T., Toit, M. D. & Torriani, S. 1991. Designation of ATCC 334 in Place of ATCC 393 (NCDO 161) as the Neotype Strain of *Lactobacillus casei* subsp. casei and Rejection of the Name *Lactobacillus paracasei* (Collins et al., 1989). *Int J Syst Evol Microbiol,* 41**,** 340-342.

Dellaglio, F., Felis, G. E. & Torriani, S. 2002. The status of the species *Lactobacillus casei* (Orla-Jensen 1916) Hansen and Lessel 1971 and *Lactobacillus paracasei* Collins et al. 1989. Request for an opinion. *Int J Syst Evol Microbiol,* 52**,** 285-287.

Dicks, L. M. T., Plessis, E. M. D., Dellaglio, F. & Lauer, E. 1996. Reclassification of *Lactobacillus casei* subsp. *casei* ATCC 393 and *Lactobacillus rhamnosus* ATCC 15820 as *Lactobacillus zeae* nom. rev., Designation of ATCC 334 as the Neotype of *L. casei* subsp. *casei,* and Rejection of the Name *Lactobacillus paracasei*. *Int J Syst Evol Microbiol,* 46**,** 337-340.

Fernández de Palencia, P., Pelaez, C., Romero, C. & Martin-Hernandez, M. 1997. Purification and characterization of the cell wall proteinase of *Lactobacillus casei* subsp. *casei* IFPL 731 isolated from raw goat's milk cheese. *J Agric Food Chem,* 45**,** 3401-3405.

Frank, J.A., Reich, C.I., Sharma, S., Weisbaum, J.S., Wilson, B.A. & Olsen, G.J. 2008. Critical evaluation of two primers commonly used for amplification of bacterial 16S rRNA genes. *Appl Environ Microbiol* 74**,** 2461-2470.

Fox, E. M., Casey, A., Jordan, K., Coffey, A., Gahan, C. G. & McAuliffe, O. 2017. Whole genome sequence analysis; an improved technology that identifies underlying genotypic differences between closely related *Listeria monocytogenes* strains. *Innov Food Sci EmergTechnol,* In Press

Gaudreau, H., Champagne, C. P. & Jelen, P. 2005. The use of crude cellular extracts of *Lactobacillus delbrueckii* ssp *bulgaricus* 11842 to stimulate growth of a probiotic *Lactobacillus rhamnosus* culture in milk. *Enzyme Microb Technol,* 36**,** 83-90.

Gobbetti, M., De Angelis, M., Di Cagno, R., Mancini, L. & Fox, P. F. 2015. Pros and cons for using non-starter lactic acid bacteria (NSLAB) as secondary/adjunct starters for cheese ripening. *Trends Food Sci Tech,* 45**,** 167-178.

Gonzalez, L., Sacristan, N., Arenas, R., Fresno, J. M. & Eugenia Tornadijo, M. 2010. Enzymatic activity of lactic acid bacteria (with antimicrobial properties) isolated from a traditional Spanish cheese. *Food Microbiol,* 27**,** 592-597.

Hebert, E. M., Mamone, G., Picariello, G., Raya, R. R., Savoy, G., Ferranti, P. & Addeo, F. 2008. Characterization of the pattern of αs1-and β-casein breakdown and release of a bioactive peptide by a cell envelope proteinase from *Lactobacillus delbrueckii* subsp. *lactis* CRL 581. *Appl Environ Microbiol,* 74**,** 3682-3689.

Herias, M. V., Koninkx, J. F., Vos, J. G., Huis in't Veld, J. H. & van Dijk, J. E. 2005. Probiotic effects of *Lactobacillus casei* on DSS-induced ulcerative colitis in mice. *Int J Food Microbiol,* 103**,** 143-155.

Iacumin, L., Ginaldi, F., Manzano, M., Anastasi, V., Reale, A., Zotta, T., Rossi, F., Coppola, R. & Comi, G. 2015. High resolution melting analysis (HRM) as a new tool for the identification of species belonging to the *Lactobacillus casei* group and comparison with species-specific PCRs and multiplex PCR. *Food Microbiol,* 46**,** 357-367.

Ivory, K., Chambers, S. J., Pin, C., Prieto, E., Arques, J. L. & Nicoletti, C. 2008. Oral delivery of *Lactobacillus casei* Shirota modifies allergen-induced immune responses in allergic rhinitis. *Clin Exp Allergy,* 38**,** 1282-1289.

Jensen, M. P. & Ardo, Y. 2010. Variation in aminopeptidase and aminotransferase activities of six cheese related *Lactobacillus helveticus* strains. *Int Dairy J,* 20**,** 149-155.

Jung, H. J., Ganesan, P., Lee, S. J. & Kwak, H. S. 2013. Comparative study of flavor in cholesterol-removed Gouda cheese and Gouda cheese during ripening. *J Dairy Sci,* 96**,** 1972-1983.

Kieronczyk, A., Skeie, S., Langsrud, T., Le Bars, D. & Yvon, M. 2004. The nature of aroma compounds produced in a cheese model by glutamate dehydrogenase positive *Lactobacillus* INF15D depends on its relative aminotransferase activities towards the different amino acids. *Int Dairy J,* 14**,** 227-235.

Kieronczyk, A., Skeie, S., Langsrud, T. & Yvon, M. 2003. Cooperation between *Lactococcus lactis* and nonstarter lactobacilli in the formation of cheese aroma from amino acids. *Appl Environ Microbiol,* 69**,** 734-739.

Kok, J., Leenhouts, K. J., Haandrikman, A. J., Ledeboer, A. M. & Venema, G. 1988. Nucleotide sequence of the cell wall proteinase gene of *Streptococcus cremoris* Wg2. *Appl Environ Microbiol,* 54**,** 231-238.

Laan, H. & Konings, W. N. 1989. Mechanism of Proteinase Release from *Lactococcus lactis* subsp. *cremoris* Wg2. *Appl Environ Microbiol,* 55**,** 3101-3106.

Macedo, A. C., Vieira, M., Pocas, R. & Malcata, F. X. 2000. Peptide hydrolase system of lactic acid bacteria isolated from Serra da Estrela cheese. *Int Dairy J,* 10**,** 769-774.

Magboul, A. A. A. & McSweeney, P. L. H. 1999. Purification and characterization of an aminopeptidase from *Lactobacillus curvatus* DPC2024. *Int Dairy J,* 9**,** 107-116.

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

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

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

Nikolic, M., Tolinacki, M., Fira, D., Golic, N. & Topisirovic, L. 2009. Variation in specificity of the PrtP extracellular proteinases in *Lactococcus lactis* and *Lactobacillus paracasei* subsp. *paracasei*. *Folia Microbiol (Praha),* 54**,** 188-194.

Oberg, C. J., Oberg, T. S., Culumber, M. D., Ortakci, F., Broadbent, J. R. & McMahon, D. J. 2016. *Lactobacillus wasatchensis* sp. nov., a non-starter lactic acid bacteria isolated from aged Cheddar cheese. *Int J Syst Evol Microbiol***,** 66, 158-164.

Pedersen, T. B., Ristagno, D., McSweeney, P. L. H., Vogensen, F. K. & Ardo, Y. 2013. Potential impact on cheese flavour of heterofermentative bacteria from starter cultures. *Int Dairy J,* 33**,** 112-119.

Rijnen, L., Yvon, M., van Kranenburg, R., Courtin, P., Verheul, A., Chambellon, E. & 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.

Sgarbi, E., Lazzi, C., Tabanelli, G., Gatti, M., Neviani, E. & Gardini, F. 2013. Nonstarter lactic acid bacteria volatilomes produced using cheese components. *J Dairy Sci,* 96**,** 4223-4234.

Sgouras, D., Maragkoudakis, P., Petraki, K., Martinez-Gonzalez, B., Eriotou, E., Michopoulos, S., Kalantzopoulos, G., Tsakalidou, E. & Mentis, A. 2004. In Vitro and In Vivo Inhibition of *Helicobacter pylori* by *Lactobacillus casei* Strain Shirota. *Appl Environ Microbiol,* 70**,** 518-526.

Simpson, P. J., Stanton, C., Fitzgerald, G. F. & Ross, R. P. 2002. Genomic Diversity within the Genus *Pediococcus* as Revealed by Randomly Amplified Polymorphic DNA PCR and Pulsed-Field Gel Electrophoresis. *Appl Environ Microbiol,* 68**,** 765-771.

Singh, T., Drake, M. & Cadwallader, K. 2003. Flavor of Cheddar cheese: a chemical and sensory perspective. *Compr Rev Food Sci Food Saf,* 2**,** 166-189.

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

Smokvina, T., Wels, M., Polka, J., Chervaux, C., Brisse, S., Boekhorst, J., van Hylckama Vlieg, J. E. & Siezen, R. J. 2013. *Lactobacillus paracasei* comparative genomics: towards species pan-genome definition and exploitation of diversity. *PLoS One,* 8**,** e68731.

Sokal, R. R. & Michener, C. D. 1958. A statistical method for evaluating systematic relationships. *Kans Univ Sci Bull,* 28**,** 1409-1438.

Stressler, T., Eisele, T., Schlayer, M., Lutz-Wahl, S. & Fischer, L. 2013. Characterization of the recombinant exopeptidases PepX and PepN from *Lactobacillus helveticus* ATCC 12046 important for food protein hydrolysis. *PLoS One,* 8**,** e70055.

Swearingen, P. A., O'Sullivan, D. J. & Warthesen, J. J. 2001. Isolation, characterization, and influence of native, nonstarter lactic acid bacteria on Cheddar cheese quality. *J Dairy Sci,* 84**,** 50-59.

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. *Ant Leeuwenhoek,* 82**,** 271-278.

Thage, B. V., Broe, M. L., Petersen, M. H., Petersen, M. A., Bennedsen, M. & Ardo, Y. 2005. Aroma development in semi-hard reduced-fat cheese inoculated with *Lactobacillus paracasei* strains with different aminotransferase profiles. *Int Dairy J,* 15**,** 795-805.

Thage, B. V., Houlberg, U. & Ardo, Y. 2004. Amino acid transamination in permeabilised cells of *Lactobacillus helveticus, L. paracasei* and *L. danicus*. *J Dairy Res,* 71**,** 461-470.

Tindall, B. 2008. The type strain of *Lactobacillus casei* is ATCC 393, ATCC 334 cannot serve as the type because it represents a different taxon, the name *Lactobacillus paracasei* and its subspecies names are not rejected and the revival of the name *'Lactobacillus zeae'*contravenes Rules 51b (1) and (2) of the International Code of Nomenclature of Bacteria. Opinion 82. *Int J Syst Evol Microbiol,* 58**,** 1764-1765.

Van de Bunt, B., Bron, P. A., Sijtsma, L., de Vos, W. M. & Hugenholtz, J. 2014. Use of non-growing *Lactococcus lactis* cell suspensions for production of volatile metabolites with direct relevance for flavour formation during dairy fermentations. *Microb Cell Fact,* 13**,** 1-9.

Van den Dool, H. & Kratz, P. D. 1963. A generalization of the retention index system including linear temperature programmed gas-liquid partition chromatography. *J Chromatogr A,* 11**,** 463-471.

Van Hoorde, K., Van Leuven, I., Dirinck, P., Heyndrickx, M., Coudijzer, K., Vandamme, P. & Huys, G. 2010. Selection, application and monitoring of *Lactobacillus paracasei* strains as adjunct cultures in the production of Gouda-type cheeses. *Int J Food Microbiol,* 144**,** 226-235.

Vasquez, A., Ahrne, S., Pettersson, B. & Molin, G. 2001. Temporal temperature gradient gel electrophoresis (TTGE) as a tool for identification of *Lactobacillus casei, Lactobacillus paracasei, Lactobacillus zeae* and *Lactobacillus rhamnosus. Lett Appl Microbiol,* 32**,** 215-219.

Vasquez, A., Molin, G., Pettersson, B., Antonsson, M. & Ahrne, S. 2005. DNA-based classification and sequence heterogeneities in the 16S rRNA genes of *Lactobacillus casei/paracasei* and related species. *Syst Appl Microbiol,* 28**,** 430-441.

Wakai, T., Shinoda, T., Uchida, N., Hattori, M., Nakamura, Y., Beresford, T., Ross, R. P. & Yamamoto, N. 2013. Comparative analysis of proteolytic enzymes need for processing of antihypertensive peptides between *Lactobacillus helveticus* CM4 and DPC4571. *J Biosci Bioeng,* 115**,** 246-252.

Ward, L. J. & Timmins, M. J. 1999. Differentiation of *Lactobacillus casei, Lactobacillus paracasei* and *Lactobacillus rhamnosus* by polymerase chain reaction. *Lett Appl Microbiol,* 29**,** 90-92.

Weimer, B., Dias, B., Ummadi, M., Broadbent, J., Brennand, C., Jaegi, J., Johnson, M., Milani, F., Steele, J. & Sisson, D. 1997. Influence of NaCl and pH on intracellular enzymes that influence Cheddar cheese ripening. *Lait,* 77**,** 383-398.

Williams, A. G., Withers, S. E., Brechany, E. Y. & Banks, J. M. 2006. Glutamate dehydrogenase activity in lactobacilli and the use of glutamate dehydrogenase-producing adjunct *Lactobacillus* spp. cultures in the manufacture of cheddar cheese. *J Appl Microbiol,* 101**,** 1062-1075.

Yu, S., Peng, Y., Zheng, Y. & Chen, W. 2015. Comparative Genome Analysis of *Lactobacillus casei*: Insights into Genomic Diversification for Niche Expansion. *Indian J Microbiol,* 55**,** 102-107.

Yvon, M. 2006. Key enzymes for flavour formation by lactic acid bacteria. *Aust J Dairy Technol,* 61**,** 88-96.

Yvon, M. & Rijnen, L. 2001. Cheese flavour formation by amino acid catabolism. *Int Dairy J,* 11**,** 185-201.

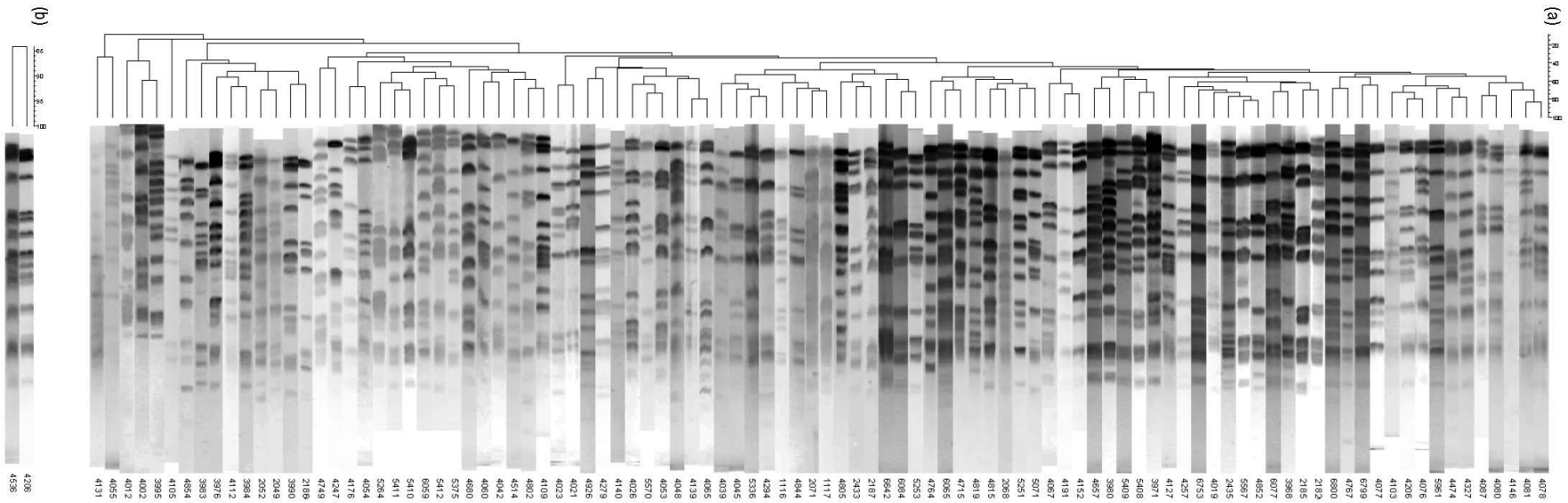


Figure 1: (a) Dendrogram of Pulsed-Field Gel Electrophoresis (PFGE) fingerprints of 98 diverse strains of *Lactobacillus casei* group generated by BioNumerics® 7.5 software, using UPGMA distance matrix method, and Pearson correlation, and (b) genetic fingerprints of two strains, DPC4206 and DPC4536, with indistinguishable PFGE profiles.

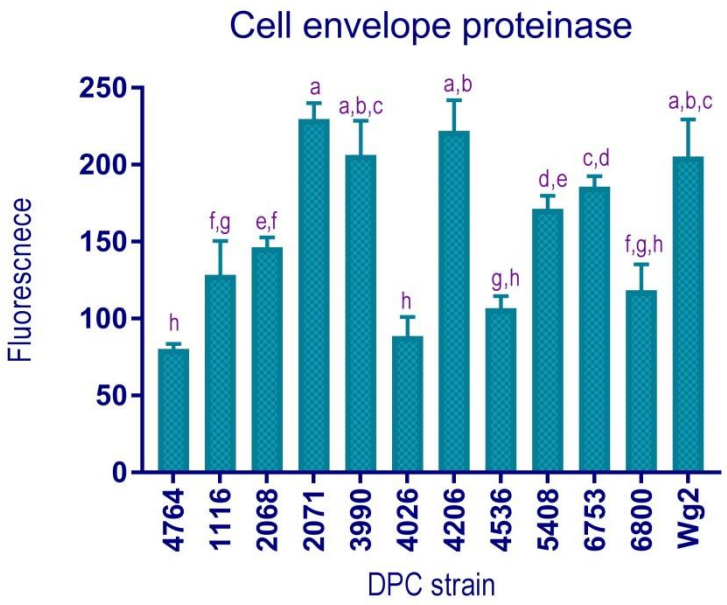


Figure 2: Cell envelope proteinase (CEP) activities of strains of *Lactobacillus casei* group as determined by EnzCheck® kit following incubation at 30°C for 24 h. Bars sharing the same letter show no significant difference according to least significant difference (LSD) test ($p<0.05$). Strains were analysed in triplicate. Error bars present standard deviation. The graph presents activities of ten representative strains, including the DPC4764 with the lowest activity observed, and *Lactococcus lactis* spp. *cremoris* Wg2, which was used as a positive control.

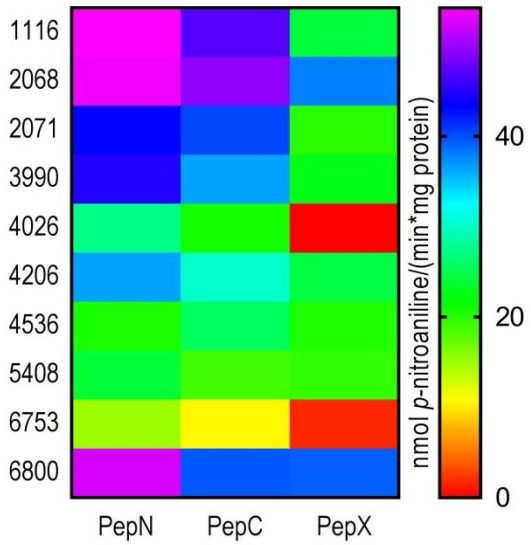


Figure 3: Heat map of aminopeptidase (PepN, PepC, PepX) activities of ten representative strains of ten representative strains of *Lactobacillus casei* group as determined by measuring cleavage of corresponding chromogenic substrates (L-Lys-*p*NA, Arg-*p*NA and Gly-Pro-*p*NA) for PepN, PepC and PepX, respectively. Results are expressed as nmol of released *p*-nitroaniline/ (min*mg of protein). Strains were analysed in triplicate.

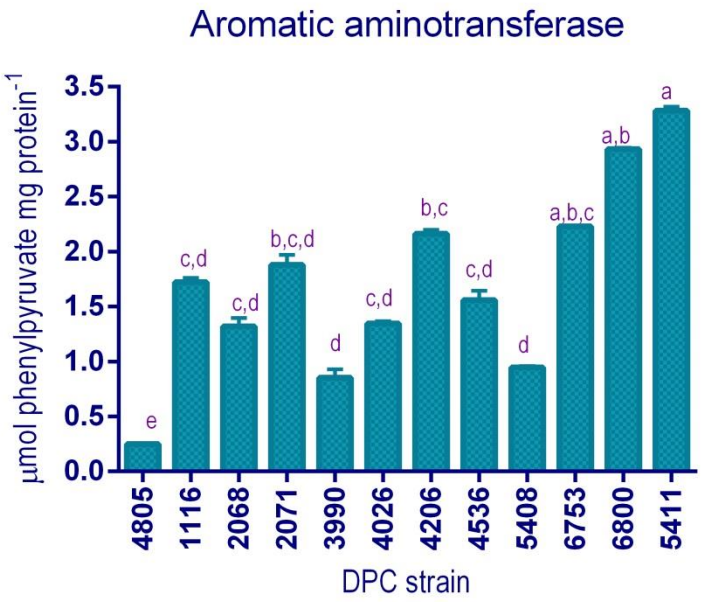


Figure 4: Aromatic aminotransferase activities of strains of *Lactobacillus casei* group determined by measuring the absorbance of phenylpyruvate, the final product of transamination between phenylalanine and α-ketoglutarate. Results are expressed as µmol of released phenylpyruvate/(min*mg of protein). Bars sharing the same letter show no significant difference according to least significant difference (LSD) test ($p<0.05$). Strains were analysed in triplicate. Error bars present standard deviation. The graph presents activities of ten representative strains, inlcuding the DPC5411 and DPC4805 with the highest and the lowest activity observed, respectively.

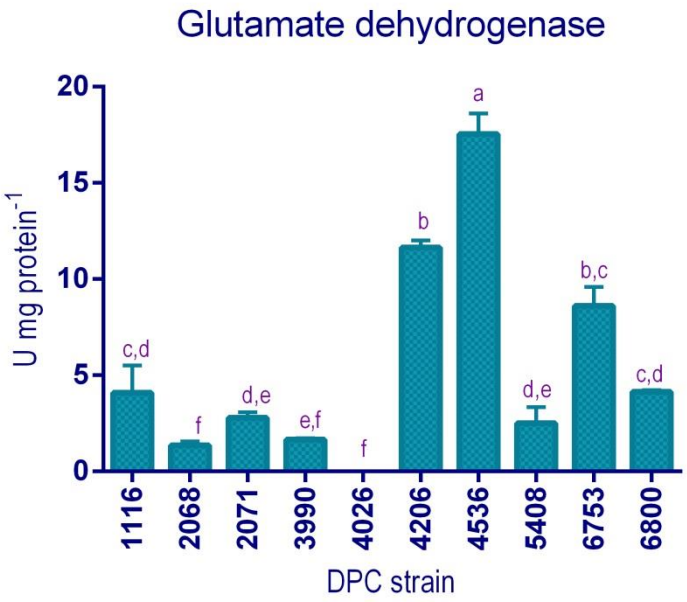


Figure 5: Glutamate dehydrogenase (GDH) activities of strains of *Lactobacillus casei* group by following change in absorbance during a reaction catalysed by GDH enzyme in which glutamic acid is converted to α-ketoglutarate in the presence of NAD^+. Results are presented as Units of enzyme activity per mg of protein, where the unit represents the amount of enzyme giving an increase of absorbance of 0.01 per 1 min. Bars sharing the same letter show no significant difference according to least significant difference (LSD) test ($p<0.05$). Strains were analysed in duplicate. Error bars present standard deviation. The graph presents activities of ten representative strains, with strain DPC4536 showing the highest, and strain DPC4026 showing the lowest GDH activity.

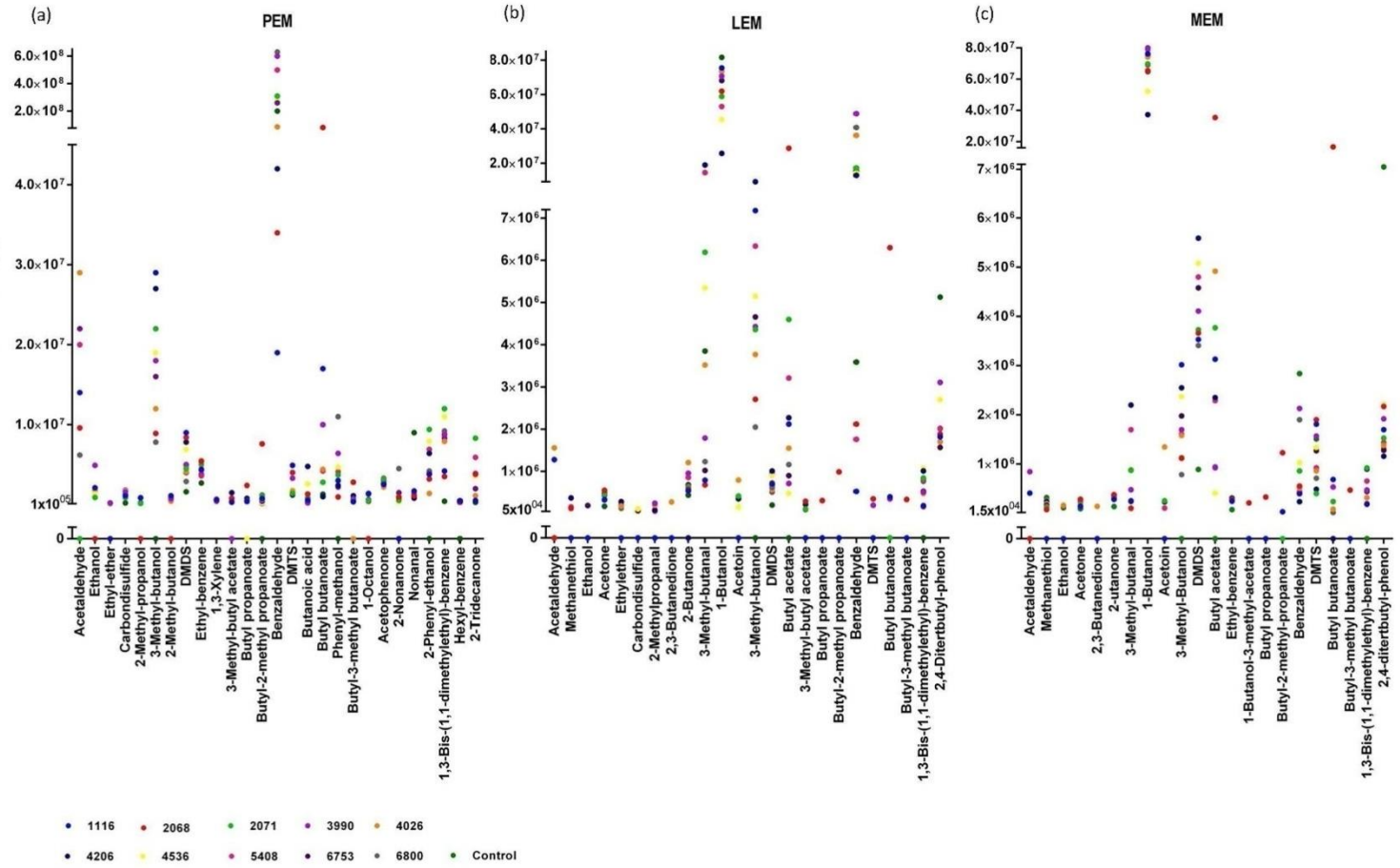


Figure 6: Abundance, in arbitrary units, of compounds for which significant differences according to least significance test (LSD, ($p<0.05$) were observed among the tested strains, including the control, in samples of ten strains of *Lactobacillus casei* group incubated for 48 h at 30 °C in three different media: phenylalanine-enhanced medium (PEM) (a), leucine-enhanced medium (LEM) (b), methionine-enhanced medium (MEM). The control consisted of un-inoculated medium incubated under the same conditions. Strains and the control were tested in triplicate in all three media.

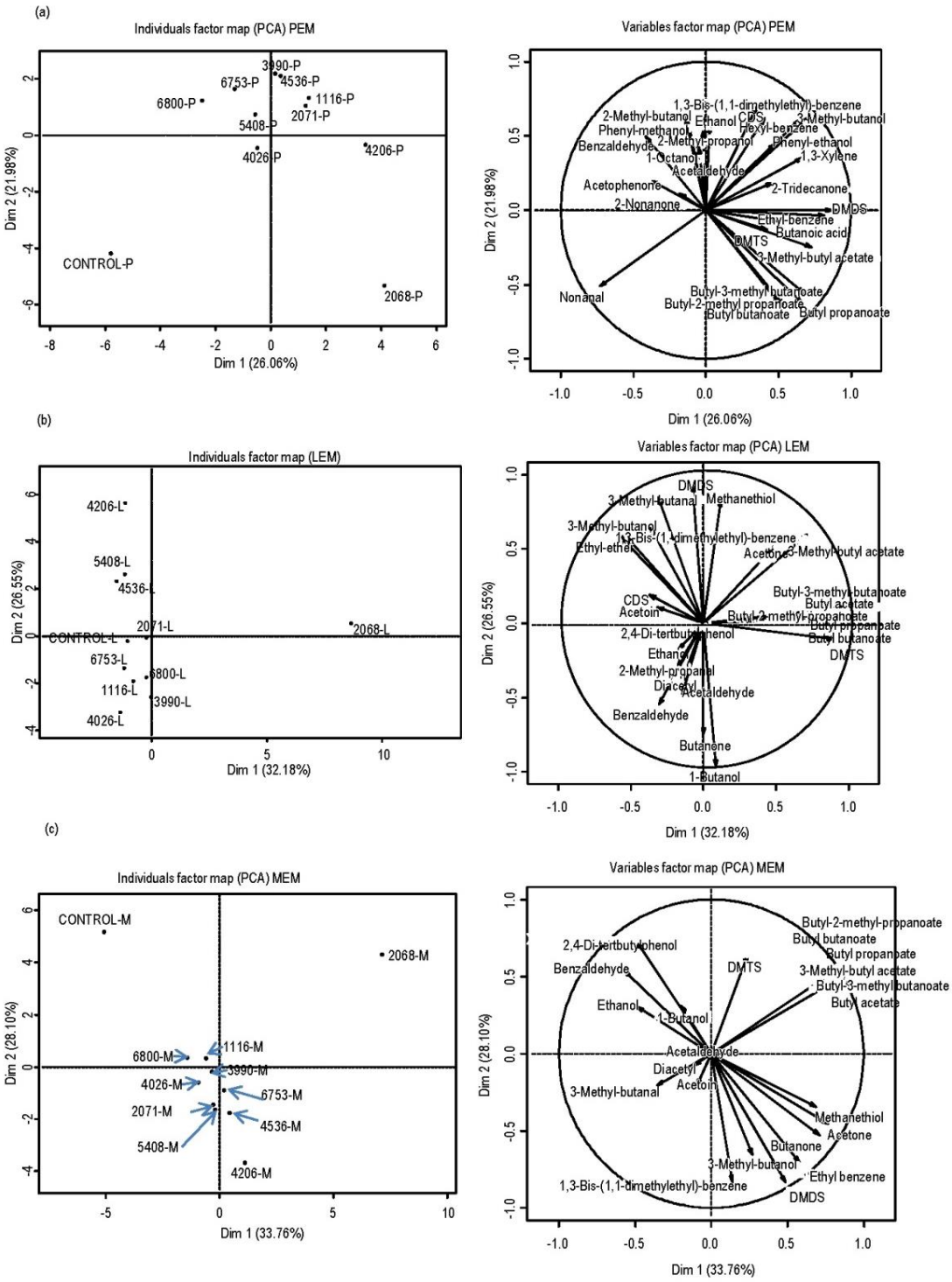


Figure 7: PCA plots of compounds for which significant differences according to least significance test (LSD), ($p<0.05$) produced by ten strains of *L. casei* group incubated for 48 h at 30 °C in three different media: phenylalanine-enhanced medium (PEM) (a), leucine-enhanced medium (LEM) (b), methionine-enhanced medium (MEM) (c) and detected as volatile compounds using HS-SPME GC-MS system. The control consisted of un-inoculated medium. Strains and the control were tested in triplicate in all three media.

Supporting Information Table 1: Results of the least significant test (LSD) ($p<0.05$) performed after Analysis of Variance (ANOVA), for the results obtained for all analyses strains of *Lactobacillus casei* group in enzymatic assays: cell envelope proteinase (CEP), aminopeptidases PepN, PepC and PepX, aromatic aminotransferase (ArAT) and glutamate dehydrogenase (GDH).

CEP (cell envelope activities):

DPC strain or sample	Fluorescence	Stat. different groups
proteinase K (2 µg/mL)	3447.33	a
2071	229.67	b
4206	222.00	bc
3990	206.33	bcd
WG2	205.33	bcde
4021	204.33	bcdef
4019	202.00	bcdefg
4076	194.33	cdefgh
5409	191.00	cdefghi
4081	189.00	cdefghij
6753	185.67	defghijk
3976	181.67	defghijkl
4023	180.67	defghijklm
4152	180.67	defghijklm
4109	180.00	defghijklmn
3971	179.00	defghijklmno
4279	178.67	defghijklmno
2435	178.33	defghijklmno

DPC strain or sample	Fluorescence	Stat. different groups
4002	177.00	defghijklmnop
5375	176.67	defghijklmnopq
5071	176.33	defghijklmnopq
4854	175.33	defghijklmnopqr
4327	172.67	defghijklmnopqrs
5408	171.33	efghijklmnopqrst
4012	170.33	fghijklmnopqrst
5411	168.67	ghijklmnopqrstu
3995	167.33	ghijklmnopqrstuv
3968	166.33	hijklmnopqrstuvw
3984	164.00	hijklmnopqrstuvwx
4140	162.33	hijklmnopqrstuvwxy
4247	161.00	hijklmnopqrstuvwxyz
4127	157.00	ijklmnopqrstuvwxyzA
4146	156.33	ijklmnopqrstuvwxyzAB
3980	155.67	jklmnopqrstuvwxyzABC
4112	155.67	jklmnopqrstuvwxyzABC
4191	155.50	jklmnopqrstuvwxyzABCD

DPC strain or sample	Fluorescence	Stat. different groups
4926	154.00	klmnopqrstuvwxyzABCD
2052	152.33	klmnopqrstuvwxyzABCD
4715	152.33	klmnopqrstuvwxyzABCD
4294	151.00	klmnopqrstuvwxyzABCD
2187	149.67	lmnopqrstuvwxyzABCDE
4139	149.00	lmnopqrstuvwxyzABCDE
4844	148.67	lmnopqrstuvwxyzABCDE
5336	148.00	lmnopqrstuvwxyzABCDEF
4077	146.33	mnopqrstuvwxyzABCDEF
2068	146.00	mnopqrstuvwxyzABCDEF
4078	145.33	nopqrstuvwxyzABCDEF
4131	145.00	opqrstuvwxyzABCDEF
3983	144.67	opqrstuvwxyzABCDEF
4474	142.67	pqrstuvwxyzABCDEFG
4680	142.00	qrstuvwxyzABCDEFGH
2186	141.00	rstuvwxyzABCDEFGHI
2049	140.33	stuvwxyzABCDEFGHI
2182	138.67	stuvwxyzABCDEFGHIJ
5412	138.67	stuvwxyzABCDEFGHIJ
4103	136.67	tuvwxyzABCDEFGHIJK
6084	135.33	uvwxyzABCDEFGHIJK
4802	133.67	vwxyzABCDEFGHIJKL
4087	132.67	vwxyzABCDEFGHIJKLM
1117	132.33	wxyzABCDEFGHIJKLM
4749	132.00	wxyzABCDEFGHIJKLM

DPC strain or sample	Fluorescence	Stat. different groups
6077	131.67	wxyzABCDEFGHIJKLM
6642	131.33	xyzABCDEFGHIJKLM
4080	130.00	xyzABCDEFGHIJKLM
1116	128.33	yzABCDEFGHIJKLM
5567	127.33	zABCDEFGHIJKLM
4176	127.00	zABCDEFGHIJKLM
4048	126.67	zABCDEFGHIJKLM
4065	125.67	ABCDEFGHIJKLM
4067	125.67	ABCDEFGHIJKLM
4815	124.00	ABCDEFGHIJKLM
6059	124.00	ABCDEFGHIJKLM
6065	122.67	ABCDEFGHIJKLMN
4257	122.33	ABCDEFGHIJKLMN
4060	121.67	BCDEFGHIJKLMN
4657	121.33	CDEFGHIJKLMN
5961	120.33	DEFGHIJKLMN
2185	119.00	DEFGHIJKLMN
5410	118.33	DEFGHIJKLMN
6800	118.33	DEFGHIJKLMN
6799	116.00	EFGHIJKLMNO
4054	115.00	EFGHIJKLMNOP
5570	113.33	FGHIJKLMNOP
4055	109.67	GHIJKLMNOP
2433	109.00	GHIJKLMNOP
4105	108.33	GHIJKLMNOP

DPC strain or sample	Fluorescence	Stat. different groups
5264	108.00	GHIJKLMNOP
5253	107.33	HIJKLMNOP
4042	107.00	IJKLMNOP
4514	106.67	IJKLMNOP
4536	106.67	IJKLMNOP
4767	104.67	JKLMNOP
4053	104.33	JKLMNOP
4819	102.00	KLMNOP

DPC strain or sample	Fluorescence	Stat. different groups
4852	99.33	LMNOP
4805	98.33	MNOP
4026	88.67	NOP
5251	88.33	NOP
4039	83.00	OP
4045	81.00	P
4764	80.33	P

PepN (Aminopeptidase N) activities:

DPC strain	nmol *p*-NA/(min*mg protein)	Stat. different groups
1116	54.21	a
2068	53.53	a
6800	52.55	a
4680	47.22	ab
2186	45.08	abc
3990	44.78	abcd
6059	44.42	abcd
4764	44.07	abcde
2071	43.30	abcdef
2433	41.47	abcdefg
4767	40.45	abcdefgh
4854	39.51	abcdefghi
4715	39.09	abcdefghij
4002	38.86	abcdefghijk
5961	37.44	abcdefghijkl
4012	37.11	abcdefghijklm
6799	36.98	abcdefghijklm
1117	36.74	abcdefghijklm
4206	36.40	abcdefghijklm
5253	34.37	bcdefghijklmn
2185	34.26	bcdefghijklmn
3976	33.83	bcdefghijklmn
4279	31.89	bcdefghijklmno
4852	31.87	bcdefghijklmno

DPC strain	nmol *p*-NA/(min*mg protein)	Stat. different groups
5412	31.58	bcdefghijklmno
4926	30.98	bcdefghijklmnop
3980	30.51	bcdefghijklmnopq
3984	30.25	bcdefghijklmnopq
5375	29.96	bcdefghijklmnopqr
4039	29.86	bcdefghijklmnopqr
2435	28.89	cdefghijklmnopqrs
6065	28.82	cdefghijklmnopqrs
5071	28.36	cdefghijklmnopqrst
4474	28.08	cdefghijklmnopqrst
4294	27.98	cdefghijklmnopqrst
4067	27.83	cdefghijklmnopqrst
4026	27.39	cdefghijklmnopqrst
4257	26.78	defghijklmnopqrst
4081	26.17	efghijklmnopqrstu
4080	26.16	efghijklmnopqrstu
6084	26.09	efghijklmnopqrstu
4819	26.05	efghijklmnopqrstu
4112	25.89	fghijklmnopqrstu
2187	25.56	fghijklmnopqrstuv
2049	25.04	ghijklmnopqrstuvw
4657	24.77	ghijklmnopqrstuvw
2182	24.64	ghijklmnopqrstuvw
4127	24.18	ghijklmnopqrstuvwx

DPC strain	nmol *p*-NA/(min*mg protein)	Stat. different groups
3971	24.12	ghijklmnopqrstuvwx
3995	24.11	ghijklmnopqrstuvwx
4078	23.98	ghijklmnopqrstuvwx
5408	23.96	ghijklmnopqrstuvwx
4749	23.61	ghijklmnopqrstuvwx
4023	23.43	hijklmnopqrstuvwx
4815	23.27	hijklmnopqrstuvwxy
5409	22.76	hijklmnopqrstuvwxy
4019	22.66	hijklmnopqrstuvwxy
4021	22.53	hijklmnopqrstuvwxy
4146	22.29	ijklmnopqrstuvwxy
2052	21.57	ijklmnopqrstuvwxyz
4327	21.08	jklmnopqrstuvwxyz
4076	21.01	klmnopqrstuvwxyz
5570	20.99	klmnopqrstuvwxyz
4536	20.61	lmnopqrstuvwxyz
4191	20.57	lmnopqrstuvwxyz
4247	20.42	lmnopqrstuvwxyz
4176	20.23	lmnopqrstuvwxyz
4139	19.33	mnopqrstuvwxyz
5410	19.26	mnopqrstuvwxyz
3968	19.09	mnopqrstuvwxyz
6642	18.36	nopqrstuvwxyz
5567	18.21	nopqrstuvwxyz
4087	18.09	nopqrstuvwxyz
6077	18.02	nopqrstuvwxyz

DPC strain	nmol *p*-NA/(min*mg protein)	Stat. different groups
4805	17.99	nopqrstuvwxyzA
4042	17.77	nopqrstuvwxyzA
4065	16.64	nopqrstuvwxyzA
4109	16.49	nopqrstuvwxyzA
4131	15.49	opqrstuvwxyzA
6753	15.05	opqrstuvwxyzA
4077	15.05	opqrstuvwxyzA
5411	15.00	opqrstuvwxyzA
4514	14.14	opqrstuvwxyzA
4105	13.89	opqrstuvwxyzA
4140	13.43	pqrstuvwxyzA
4103	13.42	pqrstuvwxyzA
4048	13.19	pqrstuvwxyzA
4055	12.86	qrstuvwxyzA
4054	12.04	rstuvwxyzA
5264	11.64	stuvwxyzA
5251	11.13	stuvwxyzA
4152	10.57	tuvwxyzA
4802	8.33	uvwxyzA
4844	7.86	vwxyzA
3983	7.10	wxyzA
4045	6.38	xyzA
4060	5.34	yzA
4053	3.60	zA
5336	0.00	A

PepC (Aminopeptidase C) activities:

DPC strain	nmol *p*-NA/(min*mg protein)	Stat. different groups
4680	50.33	a
2068	49.44	ab
1116	47.07	abc
2071	40.43	abcd
4112	40.20	abcd
2186	39.84	abcd
6800	39.54	abcd
4926	38.54	abcde
3983	38.10	abcdef
4715	37.66	abcdef
4764	37.47	abcdef
2185	36.44	abcdefg
3990	36.41	abcdefg
5961	35.69	abcdefgh
6799	34.11	abcdefghi
4767	33.73	abcdefghij
1117	31.78	bcdefghijk
4206	30.30	cdefghijkl
4657	30.19	cdefghijklm
2433	29.97	cdefghijklm
4474	29.83	cdefghijklm
4012	29.82	cdefghijklm
4294	29.50	cdefghijklm
4002	29.07	defghijklmn

DPC strain	nmol *p*-NA/(min*mg protein)	Stat. different groups
4854	28.09	defghijklmno
4080	27.75	defghijklmnop
4039	26.96	defghijklmnopq
4247	26.86	defghijklmnopq
4327	26.55	defghijklmnopq
4127	26.54	defghijklmnopq
3984	26.30	defghijklmnopq
4257	26.18	defghijklmnopq
5071	25.93	defghijklmnopqr
4536	25.62	defghijklmnopqr
6084	25.56	defghijklmnopqr
4081	25.55	defghijklmnopqr
4279	25.33	defghijklmnopqr
2435	24.81	defghijklmnopqr
2052	24.55	defghijklmnopqr
4852	24.49	defghijklmnopqr
5375	24.22	defghijklmnopqr
6059	23.88	defghijklmnopqr
5253	23.68	defghijklmnopqr
2187	23.64	defghijklmnopqr
4078	23.59	defghijklmnopqr
4042	22.72	defghijklmnopqrs
4819	20.95	efghijklmnopqrst
3976	20.88	efghijklmnopqrst

DPC strain	nmol *p*-NA/(min*mg protein)	Stat. different groups
4026	20.83	efghijklmnopqrst
4065	20.73	fghijklmnopqrst
4087	20.70	fghijklmnopqrst
3980	20.57	fghijklmnopqrst
2049	20.50	fghijklmnopqrst
5408	18.85	ghijklmnopqrst
4176	18.76	ghijklmnopqrst
2182	18.64	hijklmnopqrst
4815	18.27	hijklmnopqrstu
4067	18.08	hijklmnopqrstu
5567	17.87	ijklmnopqrstuv
4749	17.85	ijklmnopqrstuv
5570	17.47	ijklmnopqrstuvw
3968	17.33	ijklmnopqrstuvw
4023	17.08	ijklmnopqrstuvw
4019	16.90	ijklmnopqrstuvw
4191	16.38	ijklmnopqrstuvw
4805	16.33	jklmnopqrstuvw
3971	15.99	jklmnopqrstuvw
4076	15.87	klmnopqrstuvw
5409	15.07	klmnopqrstuvw
4021	15.01	klmnopqrstuvw
3995	14.78	klmnopqrstuvw
4109	14.59	klmnopqrstuvw
4077	14.26	klmnopqrstuvw
4131	14.12	klmnopqrstuvw

DPC strain	nmol *p*-NA/(min*mg protein)	Stat. different groups
6077	14.10	klmnopqrstuvw
5412	13.62	lmnopqrstuvw
4048	13.52	lmnopqrstuvw
4054	13.38	lmnopqrstuvw
6642	13.27	lmnopqrstuvw
4055	12.89	lmnopqrstuvw
6065	12.59	lmnopqrstuvw
4103	12.47	mnopqrstuvw
5264	11.66	nopqrstuvw
6753	10.53	opqrstuvw
4152	10.13	pqrstuvw
4844	9.98	qrstuvw
4514	9.75	qrstuvw
5251	8.19	rstuvw
4060	5.65	stuvw
4802	5.57	stuvw
4053	3.77	tuvw
5411	3.65	tuvw
4045	3.31	tuvw
4105	0.73	uvw
4146	0.23	vw
5336	0.18	vw
4139	0.00	w
4140	0.00	w
5410	0.00	w

PepX (Aminopeptidase X) activities:

DPC strain	nmol *p*-NA/(min*mg protein)	Stat. different groups
6800	39.24	a
2068	37.88	ab
5961	36.20	abc
4926	30.19	abcd
3984	29.34	abcde
6799	26.38	abcdef
4657	25.79	abcdefg
4206	24.49	abcdefgh
1116	24.15	abcdefgh
3980	24.07	abcdefgh
4514	23.86	abcdefgh
4715	23.52	abcdefgh
4474	23.04	abcdefgh
3990	22.51	abcdefgh
3976	21.87	bcdefghi
6077	21.85	bcdefghi
4764	21.74	bcdefghi
4247	21.53	bcdefghi
4536	20.41	cdefghij
6059	20.40	cdefghij
2071	20.02	cdefghij
3995	20.01	cdefghij
5408	19.68	cdefghij
4002	19.53	cdefghij

DPC strain	nmol *p*-NA/(min*mg protein)	Stat. different groups
3971	19.23	defghij
5411	19.20	defghij
5409	19.14	defghij
4749	19.05	defghij
1117	18.98	defghij
4109	18.85	defghij
5375	18.68	defghij
4081	18.08	defghijk
6065	17.83	defghijkl
4815	17.68	defghijkl
4078	17.55	defghijkl
2069	17.40	defghijkl
4279	17.20	defghijkl
4191	16.77	defghijklm
4076	16.30	defghijklm
4767	16.25	defghijklm
4077	16.01	defghijklm
5071	15.50	defghijklm
4819	15.33	defghijklm
4080	15.05	defghijklm
4054	14.91	defghijklm
4127	14.77	defghijklm
5410	14.49	defghijklm
4112	14.36	defghijklm

DPC strain	nmol *p*-NA/(min*mg protein)	Stat. different groups
6642	14.30	defghijklm
5251	13.93	defghijklm
2049	13.83	defghijklm
4805	13.48	defghijklm
5264	12.94	efghijklm
4087	10.91	fghijklm
4055	10.35	fghijklm
4131	9.72	fghijklm
3983	9.12	ghijklm
4176	8.05	hijklm
4152	5.37	ijklm
5412	5.28	ijklm
4802	4.22	jklm
4060	3.89	jklm
6753	1.63	klm
4048	1.19	lm
2182	0.32	m
5336	0.32	m
4146	0.27	m
5570	0.23	m
2185	0.00	m
2186	0.00	m
2187	0.00	m
2433	0.00	m
2435	0.00	m
3968	0.00	m

DPC strain	nmol *p*-NA/(min*mg protein)	Stat. different groups
4012	0.00	m
4019	0.00	m
4021	0.00	m
4023	0.00	m
4026	0.00	m
4039	0.00	m
4042	0.00	m
4045	0.00	m
4053	0.00	m
4065	0.00	m
4067	0.00	m
4103	0.00	m
4105	0.00	m
4139	0.00	m
4140	0.00	m
4257	0.00	m
4294	0.00	m
4327	0.00	m
4680	0.00	m
4844	0.00	m
4852	0.00	m
4854	0.00	m
5253	0.00	m
5567	0.00	m
6084	0.00	m

ArAT (aromatic aminotransferase) activites:

DPC strain	µmol Ph-pyruvate/mg protein	Stat. different groups
5411	3.28	a
4077	3.26	ab
6642	3.07	abc
6800	2.93	abcd
4247	2.91	abcde
4474	2.70	abcdef
4294	2.67	abcdefg
5251	2.64	abcdefgh
6059	2.60	abcdefghi
5264	2.43	abcdefghij
5410	2.40	abcdefghijk
4054	2.26	abcdefghijkl
4002	2.25	abcdefghijkl
6753	2.23	abcdefghijklm
6084	2.19	bcdefghijklmn
4749	2.18	bcdefghijklmn
4206	2.16	cdefghijklmn
4802	2.16	cdefghijklmn
6065	2.03	cdefghijklmno
4844	2.01	cdefghijklmnop
3983	1.95	defghijklmnop
3984	1.94	defghijklmnopq
5336	1.91	defghijklmnopq
3980	1.89	defghijklmnopq

DPC strain	µmol Ph-pyruvate/mg oprotein	Stat. different groups
2071	1.88	defghijklmnopq
4076	1.88	defghijklmnopq
4127	1.83	efghijklmnopqr
4131	1.78	fghijklmnopqrs
4279	1.74	fghijklmnopqrs
1116	1.72	fghijklmnopqrs
4257	1.71	fghijklmnopqrs
4715	1.70	fghijklmnopqrs
3976	1.70	fghijklmnopqrs
3971	1.69	fghijklmnopqrs
4767	1.69	fghijklmnopqrs
4819	1.64	fghijklmnopqrst
5409	1.63	fghijklmnopqrst
4764	1.61	ghijklmnopqrst
4081	1.61	ghijklmnopqrst
4536	1.56	hijklmnopqrst
5961	1.54	ijklmnopqrst
4926	1.52	ijklmnopqrstu
4078	1.52	jklmnopqrstu
4514	1.47	jklmnopqrstuv
3995	1.43	jklmnopqrstuv
4026	1.35	klmnopqrstuv
6077	1.33	klmnopqrstuvw
2068	1.32	lmnopqrstuvw

DPC strain	µmol Ph-pyruvate/mg oprotein	Stat. different groups
2049	1.24	lmnopqrstuvw
4152	1.21	lmnopqrstuvw
4176	1.21	lmnopqrstuvw
5071	1.17	lmnopqrstuvw
4112	1.16	mnopqrstuvw
1117	1.16	mnopqrstuvw
4657	1.13	nopqrstuvw
4327	1.04	opqrstuvw
4191	0.98	opqrstuvw
4680	0.97	opqrstuvw

DPC strain	µmol Ph-pyruvate/mg oprotein	Stat. different groups
4815	0.96	opqrstuvw
5408	0.95	pqrstuvw
3990	0.85	qrstuvw
4087	0.79	rstuvw
4109	0.75	rstuvw
4080	0.72	stuvw
4055	0.59	tuvw
5375	0.44	uvw
6799	0.41	vw
4805	0.25	w

GDH (glutamate dehydrogenase) activities

DPC strain	U/mg protein	Stat. different groups
4536	17.6	a
4206	11.7	b
4680	9.8	bc
4474	9.1	bcd
6753	8.6	bcde
4926	8.0	bcdef
4514	7.7	bcdefg
4327	7.3	bcdefgh
4294	6.7	bcdefghi
4112	6.4	bcdefghi
4657	5.9	cdefghi
4247	5.0	cdefghij
5409	5.0	cdefghij
6642	4.9	cdefghij
6799	4.9	cdefghij
4080	4.7	cdefghij
5375	4.5	cdefghij
5336	4.3	cdefghij
5961	4.3	cdefghij
4279	4.2	cdefghij
6800	4.1	cdefghij
1116	4.1	cdefghij
2049	4.1	defghij
6059	4.1	defghij

DPC strain	U/mg protein	Stat. different groups
6077	3.9	defghij
4819	3.8	defghij
4176	3.6	defghij
5410	3.6	defghij
4815	3.4	defghij
4805	3.3	efghij
4109	3.3	efghij
5411	3.3	efghij
4127	3.1	efghij
5264	3.1	efghij
4191	3.0	efghij
2071	2.8	fghij
3971	2.8	fghij
3980	2.7	fghij
3976	2.5	fghij
5408	2.5	fghij
4844	2.4	fghij
5071	2.4	fghij
4257	2.2	ghij
6065	2.2	ghij
1117	2.2	ghij
4715	2.2	ghij
4076	2.1	ghij
4764	2.1	ghij

DPC strain	U/mg protein	Stat. different groups
4002	2.0	ghij
4749	2.0	ghij
4054	1.9	hij
4087	1.9	hij
4131	1.9	hij
4078	1.8	hij
4081	1.8	hij
3984	1.8	hij
3990	1.7	hij
4767	1.7	hij

DPC strain	U/mg protein	Stat. different groups
5251	1.6	hij
3995	1.5	ij
2068	1.4	ij
4077	1.4	ij
4152	1.2	ij
4055	1.2	ij
3983	0.8	j
4026	0.0	j
4802	0.0	j
6084	0.0	j

Supporting Information Table 2: Cell numbers (log_{10}) of strains of *Lactobacillus casei group* in media with a predominance of a single amino acid (phenylalanine, leucine and methionine, PEM, LEM, MEM, respectively) upon inoculation (t=0 h) and after incubation at 30°C (t=48 h). Strains where significant decreases in cell numbers were observed according to the least significant differences test (LSD) ($p<0.05$) are presented in bold.

Strain of *L. casei* group	PEM		LEM		MEM	
	t=0h	t=48h	t=0h	t=48h	t=0h	t=48h
DPC1116	**9.92±0.11**	**8.85±0.20**	**9.86±0.09**	**8.42±0.12**	**9.98±0.08**	**9.11±0.13**
DPC2068	9.75±0.16	9.76±0.07	9.886±0.12	9.87±0.12	9.82±0.14	9.77±0.1
DPC2071	**10.07±0.14**	**7.96±0.29**	**10.05±0.07**	**8.05±0.19**	**10.05±0.15**	**7.85±0.12**
DPC3990	9.86±0.14	9.74±0.16	9.9±0.16	9.79±0.14	9.86±0.09	9.73±0.11
DPC4026	9.71±0.15	9.59±0.22	9.54±0.17	9.65±0.04	**9.84±0.44**	**9.61±0.16**
DPC4206	**10.18±0.35**	**8.38±0.19**	**9.99±0.16**	**8.32±0.22**	**9.92±0.11**	**8.45±0.23**
DPC4536	**9.86±0.11**	**8.37±0.19**	**9.78±0.10**	**8.82±0.14**	**9.79±0.05**	**8.12±0.13**
DPC5408	**9.76±0.18**	**8.06±0.13**	**9.72±0.152**	**8.15±0.13**	**9.73±0.14**	**8.52±0.07**
DPC6753	9.72±0.10	9.69±0.16	9.83±0.14	9.79±0.09	9.89±0.11	9.83±0.04
DPC6800	**9.88±0.13**	**8.59±0.13**	**10.07±0.1**	**8.12±0.20**	**10.1±0.14**	**8.02±0.10**

Supporting Information Table 3: Compounds generated by strains of *Lactobacillus casei* group in phenylalanine-enhanced media (PEM), leucine-enhanced media (LEM), and methionine-enhanced media (MEM) after 48 h incubation at 30°C. Compounds were detected by gas-chromatography-mass spectrometry and identified according to their linear retention indices (RI) and by comparison of mass-spectra with National Institute of Science and Technology (NIST) 2011 Mass Spectral Library. Compounds that were detected in significantly different relative abundances in samples, including the control, according to least significant differences (LSD) test ($p<0.05$), are marked with an asterix. Strains and the controls were analysed in triplicate. Controls consisted of un-inoculated media (PEM, LEM, MEM as appropriate). The information on aroma notes were obtained from "The LRI and Odour Database" at www.odour.org.uk, and publications (Curioni and Bosset 2002, Singh et al. 2003, Smit et al., 2005).

Compound	**Flavour description**	**PEM**	**LEM**	**MEM**	**RI**
alcohol					
Ethanol	Dry, dust	+*	+*	+*	<500
2-Methyl-propanol	Penetrating, alcohol, wine-like, plastic, bad	+*			617
1-Butanol	Banana-like, fruity, green, medicinal	+	+*	+*	652
3-Methyl-butanol	Fresh cheese, breath-taking, alcoholic, fruity, grainy, floral, malty	+*	+*	+*	734
2-Methyl-butanol	Malty	+*			738
2-Ethyl-hexanol	Animal, Cardboard	+			1028
Phenylmethanol (Benzyl-alcohol)	Sweet, floral, fruity, phenolic	+*			1041
1-Octanol	Waxy, green, citrus, floral, sweet, fatty, coconut	+*			1071
2-Phenylethanol	Floral, rose, dried rose	+*			1121
2,4-Di-tertbutylphenol	Phenolic	+	+*	+*	1506
acid					
Acetic acid	Vinegar, peppers, green, fruity floral, sour	+			595
Butanoic acid	Sweaty, butter, cheese, strong, acid, rancid, dirty sock	+*			791
3-methyl-butanoic acid	Cheesy, sweaty, old socks, rancid, faecal, rotten fruit, goat		+*		829
aldehyde					
Acetaldehyde	Yoghurt, green, nutty, pungent, sweet, fruity	+*	+*	+*	<500
2-Methyl-propanal	Banana, malty, chocolate-like, cocoa	+	+*		544
3-Methyl-butanal	Malty, dark chocolate, almond, cocoa, coffee	+	+*	+*	647
Benzaldehyde	Bitter almond, sweet cherry	+*	+*	+*	972
Benzeneacetaldehyde	Honey-like, rosy, violet-like, hyacinth, green	+			1052

Nonanal	Green, citrus, fatty, floral	+*			1106
3-Ethyl-benzealdehyde	?	+			1173
ketone					
2-Propanone (Acetone)	Earthy, strong fruity, wood pulp, hay	+	+*	+*	<500
2,3-Butanedione (Diacetyl)	Buttery, strong	+	+*	+*	573
2-Butanone	Buttery, sour milk, etheric	+	+*	+*	580
3-Hydroxy-butanone (Acetoin)	Buttery, sour milk, caramel	+	+*	+*	714
2-Heptanone	Blue cheese, spicy, Roquefort	+			891
1-Phenylethanone (Acetophenone)	Almond, musty, glue, orange blossom, sweet	+			1075
2-Nonanone	Malty, fruity, hot milk, smoked cheese	+*			1092
2-Undecanone	Floral, fruity, green, musty, tallow	+			1292
2-Tridecanone	Fatty, Waxy, mushroom, coconut, earthy	+*			1493
ester					
Butyl acetate	Pear, ethereal, green	+	+*	+*	814
3-Methyl-butyl acetate (Isoamyl acetate)		+*	+*	+*	876
Butyl propanoate	Earthy, sweet, rose, banana, cherry, rum	+*	+*	+*	908
Butyl-2-methyl propanoate	Sweet fruity green tropical apple banana	+*	+*	+*	953
Butyl butanoate	Pineapple, banana, sweet	+	+*	+*	996
Butyl-3-methyl butanoate (Butyl isovalerate)	Banana, sweet, pear, apple peel	+*	+*	+*	1046
Butyl hexanoate	Fruity, pineapple, waxy, green, juicy, apple	+			1188
aromatic compound					
Toluene	Nutty, bitter, almond, plastic	+*			770
Ethyl benzene	Heavy, floral	+*	+	+*	866
1,3-Xylene	Sweet, aromatic	+*			875
1,2,4,5-Tetramethylbenzene	Sweet	+			1128
1,3-Bis-(1,1-dimethylethyl)-benzene	?	+*	+*	+*	1254
Hexyl-benzene	?	+			1266
1-Methylnaphthalene	?	+			1315
sulfur compound					
Methanethiol	Rotting cabbage, cheese, vegetative, sulfur	+	+*	+*	<500
CDS	Cabbage, sulfur, fruity, burnt	+*	+*		528
Dimethyl-disulfide (DMDS)	Cabbage-like, garlic, green, sour, onion	+*	+*	+*	746
Dimethyl-trisulfide (DMTS)	Vegetable-like, sulfurous, garlic, putrid, cabbage-like	+*	+*	+*	981
other					
Ethyl ether	Sweet, ethereal	+	+*		<500

Chapter 3

Strains of the *Lactobacillus casei* group show diverse abilities for the production of flavour compounds in two model systems

All of this chapter has been published in Journal of Dairy Science (2017). In Press

DOI: 10.3168/jds.2016-12408

Ewelina Stefanovic chapter contributions:

Experimental work:

- Performed experimental work described in this chapter

Results interpretation:

- Analysis of data of work described in this manuscript

Manuscript preparation:

- Major contribution to manuscript preparation

3.1 Abstract

Cheese flavour development is directly connected with the metabolic activity of microorganisms used during its manufacture, and the selection of metabolically diverse strains represents a potential tool for the production of cheese with novel and distinct flavour characteristics. Strains of *Lactobacillus* have been proven to promote the development of important cheese flavour compounds. As cheese production and ripening are long-lasting and expensive, model systems have been developed with the purpose of rapidly screening lactic acid bacteria for their flavour potential. The biodiversity of ten strains of the *Lactobacillus casei* group was evaluated in two model systems and their volatile profiles were determined by gas chromatography-mass spectrometry (GC-MS). In Model system 1 (MS1), which represented a mixture of free amino acids, inoculated cells did not grow. In total, 66 compounds considered as flavour contributors were successfully identified, most of which were aldehydes, acids and alcohols produced *via* amino acid metabolism by selected strains. Three strains (DPC2071, DPC3990, DPC4206) had the most diverse metabolic capacities in MS1. In Model system 2 (MS2), which was based on processed cheese curd, inoculated cells increased in numbers over incubation time. A total of 47 compounds were identified, and they originated not only from proteolysis, but also from glycolytic and lipolytic processes. Tested strains produced ketones, acids and esters. Although strains produced different abundances of volatiles, diversity was less evident in MS2, and only one strain (DPC4206) was distinguished from the others. Strains identified as the most dissimilar in both of the model systems could be more useful for cheese flavour diversification.

Keywords: *Lactobacillus*, flavour, biodiversity, model system

3.2 Introduction

Formation of cheese flavour is a complex process, which results mainly from the metabolic activities of microorganisms present during cheese manufacture (Marilley and Casey, 2004, Smit et al., 2005). Lactic acid bacteria (LAB) are the most commonly found bacteria in dairy products and their metabolic diversity represents a potential tool for flavour diversification and improvement (Smit et al., 2005). Non-starter LAB (NSLAB) that originate from the cheese-making environment, dominate the cheese microbiota during ripening (Vaughan et al., 2001). The metabolic activity of NSLAB during the ripening leads to the production of compounds contributing to the flavour characteristics of cheese (Fitzsimons et al., 2001, Banks and Williams, 2004), and this effect has been shown to be highly strain-specific (Randazzo et al., 2007, Bouton et al., 2009, Pogacic et al., 2016).

The mesophilic lactobacilli dominate the NSLAB flora of cheese, as seen in a broad survey of NSLAB diversity, where 18 species of mesophilic lactobacilli were detected in 38 cheese varieties with *Lactobacillus paracasei* and *L. plantarum* as the most prevalent species. They are considered as very adaptable to the cheese environment, and along with *L. casei*, *L. curvatus* and *L. rhamnosus*, represent the core species of the non-starter microbiota (Gobbetti et al., 2015). Adjunct cultures are essentially selected strains of NSLAB that are added to cheese milk with the purpose of controlling the indigenous NSLAB population and thus, directing the development of desired cheese flavour compounds (Milesi et al., 2010, Singh and Singh, 2014). Strains of the *Lactobacillus casei* group (*L. casei, L. paracasei* and *L. rhamnosus*) have been successfully used as adjuncts, solely or in combination with

other lactobacilli in Cheddar cheese manufacture (Crow et al., 2001, Swearingen et al., 2001, Coolbear et al., 2008).

Ideally, the evaluation of the flavour-forming ability of new strains should be performed in cheese-making trials, but this is only practical as a final step as such trials are expensive, laborious and time-consuming (Milesi et al., 2007). To a certain extent, model systems mimic some aspects of the cheese ripening environment and enable rapid assessment of the development of the cheese microbiota and the resultant biochemical processes. Several types of cheese models have been developed based on miniature cheese production (Di Cagno et al., 2006, Milesi et al., 2008, Cavanagh et al., 2014), cheese slurry (Smit et al., 1995), or processed curd (Pogacic et al., 2015, Velez et al., 2015). In addition, synthetic systems that consist of solutions of a similar content to cheese could be used as model systems, such as those based on amino acid-rich media (Engels and Visser, 1996, Kieronczyk et al., 2001, Van de Bunt et al., 2014). Besides these, cheese serum extracts (Peralta et al., 2014), freeze-drying of cheese and extraction with water (Budinich et al., 2011), or lysate of cells (Sgarbi et al., 2013) were also successfully used as cheese models. A model based on miniature cheeses made from as little as 1.7 mL of milk enabled screening of flavour-forming capacities of microorganisms (Bachmann et al., 2009). In most cheese or curd based model systems, inoculated cells increased in numbers, while in synthetic medium model systems inoculated cells were not growing (Kieronczyk et al., 2001, Van de Bunt et al., 2014). Additionally, cell-free extracts have been used as a source of enzymes to investigate the flavour-forming capacity of *Lactococcus lactis* (Engels and Visser, 1996).

The aim of this study was to evaluate the diversity between strains of the *L. casei* group based on determination of their volatile profiles generated in two model

systems: a model consisting of a mixture of free amino acids and a processed curd model. Afterward, the strain diversity was mapped using a chemometric approach, which showed different abilities of strains for volatile production in the two model systems used.

3.3 Materials and methods

3.3.1 Bacterial strains

Ten strains of the *Lactobacillus casei* group of dairy origin were used in this study (DPC1116, DPC2068, DPC2071, DPC3990, DPC4026, DPC4206, DPC4536, DPC5408, DPC6753 and DPC6800). Strains used in this study were previously confirmed (by 16S rRNA PCR) to belong to species *L. casei* or *L. paracasei* and were selected based on genomic profiles (pulsed-filed gel electrophoresis) and biochemical characterization (activities of proteolytic cascade enzymes) of a set of 310 isolates obtained from the DPC Culture Collection held at the Teagasc Food Research Centre, Moorepark, Cork (Stefanovic et al., 2017). Strains were kept frozen at -80°C in de Man, Rogosa, Sharp broth (MRS, Oxoid, Basingstoke, UK) with 20 % (v/v) of glycerol, and prior to the experiment they were grown on MRS agar plates at 30°C in aerobic conditions.

3.3.2 Model system 1 (MS1): resting cells in media containing free amino acids

The MS1 consisted of a suspension of non-growing cells in a concentrated (35 % (w/v) amino acid-rich medium Bacto®Tryptone (BD, Franklin Lakes, NJ, USA) (containing minimal level of total carbohydrates (4.3 mg/g)) supplemented with 12 g/L of NaCl. This medium was chosen based on the composition of a similar model described by Van de Bunt et al. (2014). Medium for MS1 was prepared from the same batch of Bacto®Tryptone, and after addition of NaCl, it was autoclaved (121°C, 15 min). Cell manipulation was performed as described by Van de Bunt et al. (2014), with some modifications. Briefly, strains were pre-incubated for 18 h at 30°C in MRS broth, re-inoculated (1 % v/v) in 500 mL of MRS broth and incubated for 24 h at 30°C. Cells were centrifuged (4000 g, 10 min, 4°C), washed twice with 0.1 mol/L

phosphate buffer pH 6 and finally resuspended in 5 mL of the same buffer containing 15 % (v/v) glycerol and kept at -80°C until required. Thawed cell suspensions (1 mL) were added in 9 mL of the prepared amino acid-rich medium including and 10 µL of a vitamin and microelements solution, which contained 2 mg of biotin, 4.8 mg of Ca-panthotenate, 8 mg of thiamine, 8 mg of $FeSO_4$, 1.6 mg of $MgSO_4$ and 8 mg of $MnSO_4$ dissolved in 4 mL of deionised water and filter sterilised (Filtropur S syringe filter, 0.45 µm pore size, Sarstedt, Wexford, Ireland). Inoculated samples were incubated for 24 h at 30°C. For cell enumeration, samples of 100 µL were taken prior to and after incubation of the inoculated model system (at t=0 h and t=24 h) and serially diluted before plating on MRS agar followed by incubation at 30°C for 72 h. After incubation, pH values of the samples were also measured. Samples were kept at -80°C until volatile analysis was performed. The control consisted of an un-inoculated model system. Both the test strains and the un-inoculated control were evaluated in triplicate.

3.3.3 Model system 2 (MS2): growing cells in processed curd

The MS2 was prepared as previously described (Pogacic et al., 2015) with the following modifications, to achieve final concentrations of 1 g/L of lactose and 5.3 % (w/w) salt in moisture. A solution containing 1.48 g/L of peptone and 1.48 g/L of lactose and a solution of 254.25 g/L of NaCl were prepared in advance and autoclaved (121°C, 15 min). The Cheddar cheese curd (pH 5.31, NaCl 2.45 % (w/w), water activity (a_w) 0.948, moisture 38.1 % (w/w), fat 31.2 % (w/w)) and the peptone-lactose solution were mixed in a 1:2 ratio (w/w) and blended in a Waring Blender (Waring, Stamford, CT, USA), over 4 cycles for 30 s at low speed and 2 cycles for 30 s at high speed. In each tube, 10 g of the curd mixture was weighed and tubes were autoclaved (110°C, 15 min). Subsequently, 1.34 mL of the sterile NaCl

solution was added to each tube, to prevent formation of curd clogs. Strains were grown at 30°C for 24 h in MRS broth. Cultures were diluted to OD_{650nm} of 0.35-0.45 (approximately 10^7-10^8 CFU/mL), and the dilutions were used for inoculation of pre-cultures in the model system at 1 % (v/v). Pre-culture tubes (triplicate for each strain) were incubated for 24 h at 30°C after which enumeration of pre-cultures was achieved by serial dilutions and plate counting on MRS agar at 30°C for 72 h. Fresh tubes with curd were inoculated at 1 % (v/v) of pre-culture and incubated for 14 days at 30°C in anaerobic jars, after which, pH and cell counts were determined. Samples were kept at -80°C until volatile analysis was performed. The control consisted of an un-inoculated model system. As above, both the strains and the control were evaluated in triplicate.

3.3.4 Gas Chromatography-Mass Spectrometry (GC-MS) analysis of volatiles

A 2.5-mL of sample of the corresponding model system (triplicate per strain per model system) was placed in a 20 mL Perkin Elmer sealed vial. Head Space-Trap GC-MS analysis was performed using a Clarus 680 GC coupled with Clarus 600T quadrupole MS (Perkin Elmer, Courtaboeuf, France) as previously described (Pogacic et al., 2015), with modifications. Samples were warmed for 15 min to 65°C and volatiles were extracted at 207 kPa pressure maintained in the vial for 1 min with the carrier gas (helium), before being adsorbed on a Tenax® (Perkin, Elmer) trap at 35°C. The trap load was performed twice for each vial. The trap was heated at 250°C for 0.1 min and backflushed with helium at 89 kPa, leading to desorption of the volatiles. Volatiles were then separated on a Stabilwax® MS capillary column (30 m × 0.25 mm × 0.25 μm; Restek, Bellefonte, PA, USA), with helium as the mobile phase. The temperature of the oven was initially 35°C, maintained for 10 min then increased at 5°C/min up to 230°C. MS was operated in the scan mode (scan time 0.2

s, interscan delay 0.1 s) within a mass range of m/z 50 to 300. Ionization was performed by electronic impact at -70 eV. The samples were injected in a random order, with standards and blank samples (boiled deionised water) to monitor possible carryover and MS drift, as previously described (Pogacic et al., 2015).

3.3.5 Chemometric data processing and identification of compounds

Chromatographic data was processed by XCMS package of R statistical software (Smith et al., 2006) to convert GC-MS raw data to time- and mass-aligned data, providing, for each sample, the abundances for several signals (pair of mass fragment and retention time (RT). Analysis of volatiles was semi-quantitative, and results were based on abundance (peak area) only. The mean coefficient of variation of the analysis of volatile, calculated based on analysis of standards injected during GC runs, was about 17 %. Volatiles were identified by comparison of mass spectra and linear retention indices (LRI) with those of authentic standards, or tentatively identified on the basis of mass spectral data using NIST 2008 Mass Spectral Library (Scientific Instrument Services, Ringoes, NJ, USA). Where possible, in both datasets signals of same mass were used for the same compounds, in other cases, signals with the highest abundance were chosen. Some of the signals present in XCMS datasets could not be related to any compound or the percentages of identifications were considered unsatisfactory (approx. <50 %). The compounds of interest were selected according to previously published review of compounds considered as main flavour contributors in cheese (Curioni and Bosset, 2002).

3.3.6 Statistical analysis

Differences in cell counts ($\log_{10}$) and pH values before and after incubation were tested by analysis of variance (ANOVA) performed using statistical software R

(www.r-project.org). The ANOVA was also performed on selected signals to determine the presence of significant differences between the cultures. Means were compared using the least significant difference (LSD) test. Compounds with significant differences in abundances in cultures including the control, were further evaluated. Principal component analysis (PCA) was performed on selected compounds after Pareto scaling using package FactomineR of the R software.

3.4 Results

3.4.1 Resting cells in amino acid-rich medium show diverse metabolic activities

The cell enumerations in MS1 did not significantly vary during the incubation, except for DPC2071, which showed a 0.43 $\log_{10}$ unit decrease (Table 1). The pH values after incubation did not significantly differ from the pH values of the control, except for two cultures (DPC2071 and DPC3990) which showed a slight decrease (<0.2 pH units) (Table 2).

Analysis of chromatograms revealed 66 potential flavour-contributing compounds (Table 3). According to the statistical analysis (ANOVA and LSD test), 30 compounds were present in statistically different abundances in cultures, including the control ($p<0.05$). The ratio of the highest and the lowest values of abundance for a single compound between the cultures (Ratio B, Table 3) ranged between 1.5 for butyl decanoate and 111.5 for 2-ethyl-2-hexenal. It was apparent that volatile compounds were present in higher abundances in cultures than in the control, except dimethyl-trisulfide (DMTS), dimethyl-tetrasulfide and nonanal. Additionally, chromatograms of all cultures and the control showed a stretched peak of butan-1-ol and confirmed that this alcohol was present in high abundance in all MS1 samples. Several strains showed robust metabolic characteristics when incubated in MS1 (Supporting Information Table 1). Strain DPC4206 produced eight compounds in the highest relative abundances (butanal, S-methyl-thioacetate, butyl butanoate, 2-ethyl-2-hexenal, butanoic acid, 3-methyl-butanoic acid, undecan-2-ol, hexanoic acid), seven of which were significantly higher compared to abundances in all other cultures, including the control (except for hexanoic acid). Strain DPC2071 produced

3-methyl-butan-1-ol, 3-hydroxy-butan-2-one (acetoin), acetic acid, octan-1-ol and butyl decanoate in the highest abundances, and the abundance of 3-hydroxy-butan-2-one was significantly higher than in other cultures, including the control. Strain DPC3990 produced six compounds in the highest abundances (hexan-1-ol, 3-methyl-hexan-1-ol, butanedioic acid dimethyl ester, nonan-2-one, undecan-2-one, tridecan-2-one and benzeneacetic acid butyl ester), all of which except nonan-2-one and benzeneacetic acid butyl ester were produced in significantly higher abundances compared to the production by all other strains, including the control (Supporting Information Table 1).

In the PCA plot generated for MS1 using the abundance of 30 volatiles across all cultures and the control (Fig. 1), the first two axes accounted for 73.5 % of the total variability. Dimension 1 (PC1), describing 60.3 % variability was related to the abundance of the majority of flavour compounds. The variables factor map shows 14 variables that were the best represented in dimensions 1 and 2. The variables were positively associated with PC1, except for DMTS, dimethyl-tetrasulfide and nonanal, which were negatively associated with PC1. Dimension 2 (PC2), describing 13.2 % variability was positively related to butanal, butanoic acid, butyl butanoate, 2-ethyl-2-hexenal, and dimethyl-tetrasulfide, while other variables showed poor correlation with PC2. The control appeared in the left quadrant, and was negatively associated with most variables. PC1 was positively related with most of the strains, except for strains DPC6753 and DPC4026, and PC2 was positively related to DPC4206 and DPC4536, and negatively to DPC6800 and DPC2071.

Based on the PCA plot, three strains were distinguished from others. The DPC4206 strain was positioned in the upper right quadrant and was associated with the production of butanal, butanoic acid, butyl butanoate, and 2-ethyl-2-hexenal,

whereas DPC2071 appeared to be associated with the highest production of acetic acid, butyl decanoate, 3-hydroxy-butan-2-one, octan-1-ol and 3-methyl-butan-1-ol, in the lower right quadrant. The DPC3990 strain was equally separated as two fore mentioned strains in PC1, and associated with the highest abundances of nonan-2-one, hexan-1-ol, undecan-2-one, tridecan-2-one and benzeneacetic acid butyl ester.

3.4.2 Growing cells in processed curd show limited diversity in volatiles production

In MS2, cell numbers significantly increased in all cultures (Table 1), and the mean increase was 1.3 $\log_{10}$ units, while measured pH values after incubation showed a significant decrease for all cultures compared to the pH value of the control, with mean of decrease of 0.22 (Table 2).

The volatile profiles of cultures revealed a total of 47 potential flavour compounds and the abundances of ten of these showed significant differences between the cultures, including the control ($p<0.05$) (Table 4). The ratio of the highest and the lowest values of abundance between the cultures for a single compound (Ratio B, Table 4) ranged between 1.3 for 2-phenylethanol to 3.2 for 2,3-butanedione (diacetyl). Of the ten compounds for which significant differences in signal abundances were observed, aldehydes were present in lower abundances in cultures than in the control, while compounds present in higher abundances in the cultures included acetic acid, 2-phenylethanol, ethyl acetate, 2,3-butanedione and 1-hydroxy-propane-2-one. Although mean comparison showed that in the case of 2-methyl-propanal, hexanal and 2-phenylethanol there were significant differences in the relative abundances between the cultures and the control, no significant differences in relative abundances between the cultures was observed (Supporting Information

Table 2). The DPC4206 strain produced the highest abundance of 2,3-butanedione, DPC3990 produced the highest abundance of acetic acid, while DPC2068 produced the highest abundance of ethyl acetate.

In the PCA plot made for MS2 using the abundances of ten volatiles across all cultures and the control (Fig. 2), the first two axes accounted for 91 % of the total variability. Dimension 1 (PC1) described 85.6 % of the variability. Five variables were negatively associated with PC1 (2,3-butanedione, 1-hydroxy-propan-2-one, ethyl acetate, acetic acid and 2-phenylethanol), while all aldehydes were positively associated with PC1. Dimension 2 (PC2), describing 5.8 % of the variability was mainly related to 2,3-butanedione. The control appeared in the right quadrant, and was associated with the aldehydes. PC1 was negatively associated with all the cultures. Conversely, PC2 was positively associated with DPC4206, DPC2068 and DPC6800, and negatively with DPC2071 and DPC4026. The compound that contributed the most to differentiation was 2,3-butanedione. PCA showed that cultures were separated along PC2 according to the production of 2,3-butanedione with DPC4206 and DPC2071 containing significantly the highest and the lowest amounts, respectively.

3.5 Discussion

Strains of *L. casei* group represent one of the most frequently isolated NSLAB in cheese. Recently, it has been shown that NSLAB play the most pivotal role in flavour development of fermented dairy products, and it is considered that differences in their metabolic characteristics and activities could be a crucial factor for flavour diversification (Coolbear et al., 2008, Gobbetti et al., 2015). In this study, our objective was to analyse the metabolic biodiversity of ten strains of the *L. casei* group, belonging to *L. casei or L. paracasei* species by assessing their abilities to produce flavour-contributing compounds. To this end, two model systems were employed which differ in their constituents and incubation conditions. Model system 1 presents a highly concentrated mixture of amino acids at pH 7. In this model, our aim was to estimate the capability of strains to metabolise amino acids in optimal conditions for amino acid converting enzyme activity (pH approx. 7). In Model system 2, we aimed to mimic the cheese environment providing different types of substrates (proteins, sugar, lipids) and follow the metabolic activity of growing cells and cells in the stationary phase, as it occurs during cheese ripening.

Cell enumeration confirmed the expected behaviour of strains in both model systems. In MS1, we did not observe any significant change in cell numbers, except for DPC2071, or changes in pH during incubation, except for two cultures, DPC2071 and DPC3990. As this model system contains a highly concentrated solution of amino acids, and very little of other nutrients that would support cell growth were present, cells did not grow. In contrast, MS2, provided the whole range of nutrients (sugar, proteins and lipids), and cells numbers increased significantly ($p<0.05$) while pH decreased, compared to the un-inoculated control. Similar results were described

in other studies where model systems were used. For example, in a synthetic model, Kieronczyk et al. (2001) showed only a slight decrease in the numbers of lactobacilli over six days of incubation. In cheese-based models, the number of *L. plantarum* increased over 30 days of incubation (Milesi et al., 2008), but a conflicting result was reported by Di Cagno et al. (2006) where numbers of mesophilic lactobacilli decreased by one log unit after 36 days of incubation. In the study of Pogacic et al. (2015), cell numbers of lactobacilli in a curd-based medium increased in the first 24 h of incubation, but after five weeks the numbers of *L. paracasei* slightly decreased, reaching 8.14 $\log_{10}$ units. The lower cell numbers reached in this study, using the same model system, could be due to the different incubation conditions (temperature and time).

In MS1, the aim was to determine the diversity of strains by their ability to metabolise amino acids, since the products of amino acid catabolism are generally seen as highly important food flavour contributors (Yvon and Rijnen, 2001). This model was developed on the basis of a model published by Van de Bunt et al. (2014), which provided a rapid way to evaluate the flavour formation capacity of strains. In our approach, we used pancreatic digest of casein, because it brings amino acids in proportion similar to that of ripened cheese. In MS1, the inoculated strains produced volatiles such as short-chain aldehydes, alcohols and acids which correspond to the intensive amino acid catabolism (McSweeney and Sousa, 2000). Butanal is known for its pungent flavour (Singh et al., 2003), while 3-methyl-butan-1-ol, a product of leucine metabolism, has a fruity, alcohol, grainy flavour (Singh et al., 2003). Aromatic alcohols, such as 2-phenylethanol arising from phenylalanine, have a rose flavour (Singh et al., 2003). Long-chain alcohols, such as undecan-2-ol, are produced in the reduction of 2-methyl-ketones, and hexan-1-ol and octan-1-ol

probably originated from the reduction of fatty acids. Long-chain ketones (undecan-2-one and tridecan-2-one) are products of lipid metabolism and have fruity, nutty notes (Collins et al., 2003). The aldehyde 2-ethyl-2-hexanal most probably originated during lipid oxidation. As inoculated cells were in the stationary phase of growth, some of fatty acids released from the cell membranes may have been the source of these compounds, as lipids were not present in this model system. Acids, such as butanoic and hexanoic acid, are characterised by rancid and goaty flavour, respectively (Curioni and Bosset, 2002), and have a lipid source, probably from the cell membranes although hexanoic acid may also originate from lysine (Peralta et al., 2014). Branched-chain 3-methyl-butanoic acid, originated from leucine and has rancid, cheesy and sweaty notes (Curioni and Bosset, 2002). Among the esters detected, butyl esters dominated, as a result of the secondary reaction of acid esterification due to the high abundance of butan-1-ol observed in this model system. This alcohol was present in the media itself, since we detected it in the un-inoculated control. Esters in general contribute to fruity flavour notes (Curioni and Bosset, 2002). Sulfur compounds that arise from sulfur amino acid (methionine, cysteine) metabolism contribute to garlic and onion flavours (Singh et al., 2003). Of all the sulfur compounds detected, the production of S-methyl-thioacetate showed the highest variations among cultures. This molecule is generated in the reaction of acetyl-CoA and methanethiol, a metabolite of methionine, and has cooked cauliflower flavour (Arfi et al., 2002). DMTS and dimethyl-tetrasulfide were present in the highest concentration in the control. Their presence in the control may be a result of methionine degradation during the medium manufacture process, or its sterilisation. Lower concentrations in the cultures in comparison to the control could also suggest that either DMTS and dimethyl-tetrasulfide, either one of their

precursors, such as methanethiol, may have been metabolised by the strains. In addition, while pyrazine derivatives were present in the control presumably originating due to the sterilisation of the media, significant differences between cultures were observed for two pyrazines (2,3,5-trimethyl-6-ethyl-pyrazine and 2,6-dimethyl-3-*sec*-butyl-pyrazine) and it appears that the cultures may be producing these compounds, which contribute to earthy, roasty and potato flavours (Curioni and Bosset, 2002, Singh et al., 2003). Some bacteria and yeast can produce pyrazines (Schulz and Dickschat, 2007, Rajini et al., 2011), although an enzymatic pathway involved in pyrazine synthesis in lactobacilli has not been described. Pyrazines may arise in non-enzymatic reactions between metabolites of amino acids, such as α-aminoketones and α-dicarbonyl compounds (Rajini et al., 2011). A low level of total carbohydrates was present in the MS1 and, as expected, we observed a limited number of sugar metabolites, including ethanol, acetic acid, which has typical vinegar flavour, and 3-hydroxy-butan-2-one which is important for its buttery notes (Singh et al., 2003). However, both 3-hydroxy-butan-2-one and acetic acid could have also originated through amino acid metabolism (Skeie et al., 2008, Peralta et al., 2014). Nevertheless, we did not detect 2,3-butanedione, a compound produced from pyruvate, an intermediate molecule in carbohydrate metabolism (Jyoti et al., 2003, Liu, 2003, Bachmann et al. 2009). Additionally, in MS1, we identified 4-propylbenzaldehyde (most probably metabolite of phenylalanine) and 2-acetylthiazole (most probably originating from methionine, or cysteine (Law, 1997)). These compounds are inevitably produced in the amino acid-rich environment, and while some of them have flavour potential (2-acetylthiazole (Burdock, 2016)), they are not usually, if at all, associated with cheese flavour.

In MS1, strains showed considerable metabolic differences, with three strains, DPC2071, DPC3990 and DPC4206 producing the highest amounts of flavour compounds, often significantly higher compared to the production by other strains (LSD test, Supporting Information Table 1). This observation confirms the biodiversity of *L. casei* strains in their ability to metabolise amino acid and produce a variety of volatile compounds.

In MS2, we examined the biodiversity of strains not only in the presence of amino acids, but also in the presence of other substrates available in the processed curd or added during model preparation (lactose), to investigate their glycolytic and lipolytic activities. The main metabolic product of strains was 2,3-butanedione, which is considered as a major flavour contributor to buttery and cheesy notes (Curioni and Bosset, 2002) arising from lactose or citrate metabolism (Bachmann et al., 2009). Moreover, we observed the highest variability among tested cultures in the production of 2,3-butanedione, with DPC4206 and DPC2071 producing the highest and the lowest abundance, respectively. The aromatic alcohol 2-phenylethanol, a product of phenylalanine metabolism known for its rose flower notes (Curioni and Bosset, 2002), was also detected, with all the strains producing similar amounts of this alcohol. The strains also produced acetic acid from amino acid or carbohydrate sources (Singh et al., 2003), and ethyl acetate, which gives fruity notes. Esters originated from esterification of the acids and alcohols formed from carbohydrate and amino acid metabolism. Although many acids were detected in cultures, only two esters were identified. The reason for this observation could be the lower level of alcohols available or the dominance of the reverse reaction over the course of incubation time. The chromatograms were abundant in long-chain methyl-ketones and acids, but there was no significant difference between the abundances observed

between the cultures and the control. These molecules most probably arose from lipid hydrolysis and the metabolism of starter cultures present in non-processed cheese curd (McSweeney and Sousa, 2000, Singh et al., 2003) and lipid oxidation probably occurred independently of the cultures metabolic activities, during sterilisation, which contributed to the equal amounts of these compounds in all cultures and the control. All aldehydes, for which significant differences were observed, were present in lower concentrations in cultures than in the control, and probably were reduced to alcohols during incubation. The initial presence of aldehydes in the control could be connected to the metabolic activity of starter culture present in the fresh curd.

Although cultures showed different abilities to metabolise substrates in MS2 and we observed differences in compound abundances in cultures compared to the control, the diversity of microbial volatiles among cultures was lower than observed using MS1, as only a few compounds were produced in significantly different abundances across the strains. 2,3-butanedione was the molecule that contributed to the highest level of differentiation, as the ratio between the highest and the lowest abundance among the cultures was the highest for this compound. The DPC4206 strain was shown to be the most differentiated from the other strains, producing the highest amount of 2,3-butanedione, followed by DPC6800 and DPC2068 (Supporting Information Table 2). Other compounds have also contributed to differentiation, but their effect was modest, as differences in abundances were lower. Although in PCA plot DPC2071 appears differentiated, its position was mainly due to a low level of aldehydes and 2,3-butanedione in comparison to the other cultures.

As an outcome of the diversity studies, a comparison of the two model systems was possible. Firstly, in both model systems, we observed a difference between the

control and the cultures, which suggests that all the strains were metabolically active in both environments. MS1 enabled detection of more flavour-related compounds compared to the MS2, but not all were directly associated with cheese flavour. MS1 is rich in amino acids, and as expected, this was the major pathway that could be investigated in this study with that model. MS1 provides a rapid approach for estimation of strains ability to metabolise amino acids in ratios present in final stages of cheese ripening. On the other hand, MS2, based on curd, enabled different flavour pathway development and also evaluated growing strains in conditions simulating cheese ripening (NaCl and presence of other cheese substrates in corresponding amounts and ratios). This model allowed determination of volatiles produced by both growing cells (first 24-48 h of incubation) and cells in stationary phase (until the end of incubation). The profiles of MS1 were abundant in sulfur compounds that arose from methionine metabolism. However, although some of these are seen as flavour contributors, they are not often observed in cheese. Conversely, in MS2, only one sulfur compound (dimethyl-disulfide) was detected. Compounds like 2,3-butanedione and propanoic acid originate from sugar and amino acid metabolism and were not present in MS1, but we identified them in MS2. Esterification was much more efficient in MS1 with butyl esters dominant, due to the extremely high abundance of butan-1-ol in the substrate. Conversely, only two ethyl esters were identified in MS2. However, in both model systems we confirmed that tested strains of the *L. casei* group have different abilities to metabolise substrates and produce a variety of compounds with potential to contribute to cheese flavour.

3.6 Conclusion

In this study, we aimed to investigate diversity of ten strains of *L. casei* group based on their flavour-contributing potential. The metabolic variability of the strains was evaluated in two model systems. The results obtained in MS1 demonstrated that tested strains have different abilities to metabolise amino acids to flavour compounds, with strains DPC2071, DPC3990 and DPC4206 displaying the most diverse metabolic profiles. In MS2, strains used various metabolic pathways, and apart from volatiles produced through amino acid catabolism, metabolites originating from glycolysis and lipolysis were also identified, but differences between the strains were less evident and only strain DPC4206 was slightly different from the other strains.

Taking all these results into account, we can conclude that strains of *L. casei* group express diverse metabolic potential in the two model systems. The use of model systems gave an insight into the strains' metabolic characteristics and flavour development potential. The differences observed in volatile production can serve as guidance for selection of strains with the potential to diversify cheese flavour. It is envisaged that strain-to-strain diversity in volatile profiles will reflect in variations in flavour of manufactured cheese. The screening of volatile profiles of strains in model systems prior cheese manufacture could help in selection of strains with potential to diversify cheese flavour.

3.7 Acknowledgments

Ewelina Stefanovic and Andrea Bertuzzi are in a receipt of Teagasc Walsh Fellowships. This collaboration was supported by a Short Term Overseas Training Award 2015, funded by Teagasc.

3.8 References

Arfi, K., Spinnler, H. E., Tache, R. & Bonnarme, P. 2002. Production of volatile compounds by cheese-ripening yeasts: requirement for a methanethiol donor for S-methyl thioacetate synthesis by *Kluyveromyces lactis*. *Appl Microbiol Biotechnol,* 58**,** 503-510.

Bachmann, H., Kruijswijk, Z., Molenaar, D., Kleerebezem, M. & van Hylckama Vlieg, J. E. 2009. A high-throughput cheese manufacturing model for effective cheese starter culture screening. *J Dairy Sci,* 92**,** 5868-5882.

Banks, J. M. & Williams, A. G. 2004. The role of the nonstarter lactic acid bacteria in Cheddar cheese ripening. *Int J Dairy Technol,* 57**,** 145-152.

Bouton, Y., Buchin, S., Duboz, G., Pochet, S. & Beuvier, E. 2009. Effect of mesophilic lactobacilli and enterococci adjunct cultures on the final characteristics of a microfiltered milk Swiss-type cheese. *Food Microbiol,* 26**,** 183-191.

Budinich, M. F., Perez-Diaz, I., Cai, H., Rankin, S. A., Broadbent, J. R. & Steele, J. L. 2011. Growth of *Lactobacillus paracasei* ATCC 334 in a cheese model system: a biochemical approach. *J Dairy Sci,* 94**,** 5263-5277.

Burdock, G. A. 2016. *Fenaroli's handbook of flavor ingredients,* CRC press.

Cavanagh, D., Kilcawley, K. N., O'Sullivan, M. G., Fitzgerald, G. F. & 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.

Collins, Y. F., McSweeney, P. L. H. & Wilkinson, M. G. 2003. Lipolysis and free fatty acid catabolism in cheese: a review of current knowledge. *Int Dairy J,* 13**,** 841-866.

Coolbear, T., Crow, V., Harnett, J., Harvey, S., Holland, R. & Martley, F. 2008. Developments in cheese microbiology in New Zealand - Use of starter and non-starter lactic acid bacteria and their enzymes in determining flavour. *Int Dairy J,* 18**,** 705-713.

Crow, V., Curry, B. & Hayes, M. 2001. The ecology of non-starter lactic acid bacteria (NSLAB) and their use as adjuncts in New Zealand Cheddar. *Int Dairy J,* 11**,** 275-283.

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.

Di Cagno, R., Quinto, M., Corsetti, A., Minervini, F. & Gobbetti, M. 2006. Assessing the proteolytic and lipolytic activities of single strains of mesophilic lactobacilli as adjunct cultures using a Caciotta cheese model system. *Int Dairy J,* 16**,** 119-130.

Engels, W. J. M. & Visser, S. 1996. Development of cheese flavour from peptides and amino acids by cell-free extracts of *Lactococcus lactis* subsp *cremoris* B78 in a model system. *Neth Milk Dairy J,* 50**,** 3-17.

Fitzsimons, N. A., Cogan, T. M., Condon, S. & Beresford, T. 2001. Spatial and temporal distribution of non-starter lactic acid bacteria in Cheddar cheese. *J Appl Microbiol,* 90**,** 600-608.

Gobbetti, M., De Angelis, M., Di Cagno, R., Mancini, L. & Fox, P. F. 2015. Pros and cons for using non-starter lactic acid bacteria (NSLAB) as secondary/adjunct starters for cheese ripening. *Trends Food Sci Tech,* 45**,** 167-178.

Jyoti, B. D., Suresh, A. K. & Venkatesh, K. V. 2004. Effect of preculturing conditions on growth of *Lactobacillus rhamnosus* on medium containing glucose and citrate. *Microbiol Res,* 159**,** 35-42.

Kieronczyk, A., Skeie, S., Olsen, K. & Langsrud, T. 2001. Metabolism of amino acids by resting cells of non-starter lactobacilli in relation to flavour development in cheese. *Int Dairy J,* 11**,** 217-224.

Law, B. A. 2012. *Microbiology and biochemistry of cheese and fermented milk,* Blackie Academic&Professional.

Liu, S. 2003. Practical implications of lactate and pyruvate metabolism by lactic acid bacteria in food and beverage fermentations. *Int J Food Microbiol,* 83**,** 115-131.

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

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

Milesi, M. M., Candioti, M. & Hynes, E. 2007. Mini soft cheese as a simple model for biochemical studies on cheese-making and ripening. *LWT-Food Sci Technol,* 40**,** 1427-1433.

Milesi, M. M., McSweeney, P. L. & Hynes, E. R. 2008. Viability and contribution to proteolysis of an adjunct culture of *Lactobacillus plantarum* in two model cheese systems: cheddar cheese-type and soft-cheese type. *J Appl Microbiol,* 105**,** 884-892.

Milesi, M. M., Wolf, I. V., Bergamini, C. V. & Hynes, E. R. 2010. Two strains of nonstarter lactobacilli increased the production of flavor compounds in soft cheeses. *J Dairy Sci,* 93**,** 5020-5031.

Peralta, G. H., Wolf, I. V., Bergamini, C. V., Perotti, M. C. & Hynes, E. R. 2014. Evaluation of volatile compounds produced by *Lactobacillus paracasei* I90 in a hard-cooked cheese model using solid-phase microextraction. *Dairy Sci Technol,* 94**,** 73-81.

Pogacic, T., Maillard, M. B., Leclerc, A., Herve, C., Chuat, V., Valence, F. & Thierry, A. 2016. *Lactobacillus* and *Leuconostoc* volatilomes in cheese conditions. *Appl Microbiol Biotechnol,* 100**,** 2335-2346.

Pogacic, T., Maillard, M. B., Leclerc, A., Herve, C., Chuat, V., Yee, A. L., Valence, F. & Thierry, A. 2015. A methodological approach to screen diverse cheese-related bacteria for their ability to produce aroma compounds. *Food Microbiol,* 46**,** 145-153.

Rajini, K. S., Aparna, P., Sasikala, C. & Ramana Ch, V. 2011. Microbial metabolism of pyrazines. *Crit Rev Microbiol,* 37**,** 99-112.

Randazzo, C. L., De Luca, S., Todaro, A., Restuccia, C., Lanza, C. M., Spagna, G. & Caggia, C. 2007. Preliminary characterization of wild lactic acid bacteria and their

abilities to produce flavour compounds in ripened model cheese system. *J Appl Microbiol,* 103**,** 427-435.

Schulz, S. & Dickschat, J. S. 2007. Bacterial volatiles: the smell of small organisms. *Nat Prod Rep,* 24**,** 814-842.

Sgarbi, E., Lazzi, C., Tabanelli, G., Gatti, M., Neviani, E. & Gardini, F. 2013. Nonstarter lactic acid bacteria volatilomes produced using cheese components. *J Dairy Sci,* 96**,** 4223-4234.

Singh, S. & Singh, R. 2014. Phenotypic and genotypic characterization of non-starter *Lactobacillus* species diversity in Indian Cheddar cheese. *LWT-Food Sci Technol,* 55**,** 415-420.

Singh, T., Drake, M. & Cadwallader, K. 2003. Flavor of Cheddar cheese: a chemical and sensory perspective. *Compr Rev Food Sci Food Saf,* 2**,** 166-189.

Skeie, S., Kieronczyk, A., Næss, R. M. & Østlie, H. 2008. *Lactobacillus* adjuncts in cheese: Their influence on the degradation of citrate and serine during ripening of a washed curd cheese. *Int Dairy J,* 18**,** 158-168.

Smit, G., Braber, A., Van Spronsen, W., Van den Berg, G. & Exterkate, F. A. 1995. Ch-easy model: A cheese-based model to study cheese ripening. *In*: P. Étiévant, and P. Schreier, (eds.) *Bioflavour 95***,** INRA.

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

Smith, C. A., Want, E. J., O'Maille, G., Abagyan, R. & Siuzdak, G. 2006. XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Anal Chem,* 78**,** 779-787.

Stefanovic, E., Kilcawley, K. N., Rea, M. C., Fitzgerald, G. F. & McAuliffe, O. 2017. Genetic, enzymatic and metabolite profiling of the *Lactobacillus casei* group reveals strain biodiversity and potential applications for flavour diversification. *J Appl Microbiol***,** 122**,** 1245-1261.

Swearingen, P. A., O'Sullivan, D. J. & Warthesen, J. J. 2001. Isolation, characterization, and influence of native, nonstarter lactic acid bacteria on Cheddar cheese quality. *J Dairy Sci,* 84**,** 50-59.

Van de Bunt, B., Bron, P. A., Sijtsma, L., de Vos, W. M. & Hugenholtz, J. 2014. Use of non-growing *Lactococcus lactis* cell suspensions for production of volatile metabolites with direct relevance for flavour formation during dairy fermentations. *Microb Cell Fact,* 13**,** 1-9.

Vaughan, C., Curry, B. & Hayes, M. 2001. The ecology of non-starter lactic acid bacteria (NSLAB) and their use as adjuncts in New Zealand Cheddar. *Int Dairy J,* 11**,** 215-283.

Velez, M. A., Perotti, M. C., Rebechi, S. R. & Hynes, E. R. 2015. Short communication: A new minicurd model system for hard cooked cheeses. *J Dairy Sci,* 98**,** 3679-3683.

Yvon, M. & Rijnen, L. 2001. Cheese flavour formation by amino acid catabolism. *Int Dairy J,* 11**,** 185-201.

Table 1: Cell enumeration in two model systems prior to and after incubation. All strains belong to *Lactobacillus casei* group. Results are presented as mean values ± standard deviation of triplicate experiments. Strains which showed a significant difference in cell numbers after incubation according to LSD test are presented in bold ($p<0.05$).

Strain	Model system 1		Model system 2	
	t=0 h	t=24 h	t=0 h	t=14 d
DPC1116	9.96±0.21	9.875±0.09	**6.71±0.05**	**8.35±0.16**
DPC2068	9.49±0.06	9.62±0.08	**6.71±0.05**	**7.79±0.17**
DPC2071	**10.22±0.21**	**9.79±0.09**	**6.67±0.07**	**8.13±0.18**
DPC3990	10.05±0.11	10.02±0.07	**6.64±0.06**	**8.10±0.14**
DPC4026	9.84±0.04	9.68±0.17	**6.65±0.18**	**7.92±0.05**
DPC4206	10.02±0.04	9.95±0.06	**6.67±0.09**	**7.63±0.33**
DPC4536	9.97±0.07	9.90±0.10	**6.35±0.19**	**7.47±0.41**
DPC5408	9.80±0.08	9.79±0.13	**5.87±0.73**	**7.92±0.22**
DPC6753	9.99±0.13	9.85±0.20	**6.20±0.55**	**7.69±0.13**
DPC6800	10.1±0.04	10.06±0.10	**6.75±0.05**	**7.33±0.17**

Table 2: pH values of cultures and control measured at the end of incubation of strains of *Lactobacillus casei* group in the two models. Results are presented as mean values ± standard deviation of triplicate experiments. The control consisted of un-inoculated model systems. Results of pH values sharing the same letter in the column corresponding to Model system 1, or Model system 2 do not significantly differ according to LSD test ($p<0.05$).

Strain	Model system 1		Model system 2	
DPC1116	6.95±0.02	ab	5.12±0.03	b
DPC2068	6.95±0.03	ab	5.01±0.04	f
DPC2071	6.82±0.11	c	5.07±0.01	bcde
DPC3990	6.87±0.04	bc	5.01±0.02	ef
DPC4026	7.05±0.03	a	5.08±0.02	bcd
DPC4206	6.93±0.08	abc	5.05±0.03	cdef
DPC4536	6.96±0.14	ab	5.09±0.08	bc
DPC5408	6.93±0.13	abc	5.03±0.01	def
DPC6753	7.01±0.04	a	5.06±0.02	cde
DPC6800	7.01±0.04	a	5.05±0.00	cdef
Control	7.01±0.01	a	5.28±0.02	a

Table 3: Compounds identified in Model system 1 along with linear retention indices (LRI) used for compounds identification. The control was an un-inoculated model under the same conditions. Cultures and control were tested in triplicate. Compounds that exhibited significant differences in abundances in cultures, including the control ($p<0.05$), are presented in bold.

Chemical group	Compound	Mass fragment used for analysis	LRI	Ratio A[1]	Ratio B[2]
aldehyde	**Butanal**	**44**	**867**	**17.8 cult>C**	**16.1**
	3-Methyl-butanal	58	909		
	2-Ethyl-2-hexenal	**97**	**1291**	**111.5 cult>C**	**111.5**
	Nonanal	**68**	**1388**	**7.6 cult<C**	**2.7**
	Benzaldehyde	77	1518		
	Benzeneacetaldehyde	91	1642		
	4-propylbenzaldehyde	91	1805		
ketone	Propan-2-one (Acetone)	58	-		
	Butan-2-one	72	896		
	3-Hydroxy-butan-2-one (Acetoin)	**43**	**1278**	**80.3 cult>C**	**11.8**
	4-methyl-pentan-2-one	100	1003		
	Heptan-2-one	43	1180		
	Nonan-2-one	**58**	**1383**	**6.0 cult~C**	**6.0**
	Undecan-2-one	**58**	**1594**	**19.7 cult>C**	**12.2**
	Tridecan-2-one	**58**	**1807**	**17.9 cult>C**	**7.2**
	1-Phenylethanone (Acetophenone)	77	1646		
	Benzyl-methyl-ketone	91	1727		
alcohol	Ethanol	46	912		
	Butan-1-ol	56	1169		
	3-Methyl-butan-1-ol	**70**	**1216**	**205.6 cult>C**	**1.9**
	Hexan-1-ol	**69**	**1358**	**38.3 cult>C**	**9.7**
	3-Methyl-hexan-1-ol	**70**	**1357**	**44.2 cult>C**	**44.2**
	2-Ethyl-hexan-1-ol	57	1495		
	Octan-1-ol	**68**	**1583**	**17.1 cult>C**	**3.4**
	Undecan-2-ol	**69**	**1722**	**19.3 cult>C**	**10.1**
	Phenol	66	-		
	Phenylmethanol (Benzyl-alcohol)	79	-		
	2-Phenylethanol (Phenyl-ethyl alcohol)	**91**	**-**	**11.0 cult>C**	**2.2**
	2,4-Di-tertbutylphenol	191	-		
	2-Furanmethanol	98	1659		
ester	**Ethyl hexanoate**	**88**	**1227**	**11.6 cult>C**	**6.9**
	Butyl acetate	**43**	**1094**	**53.4 cult>C**	**2.6**
	Butyl butanoate	**71**	**1273**	**15.2 cult>C**	**9.0**
	Butyl hexanoate	99	1399		
	Butyl octanoate	101	1602		
	Butyl decanoate	**116**	**-**	**10.3 cult>C**	**1.5**
	Butanedioic acid, dimethyl ester	**115**	**1591**	**2.4 cult>C**	**2.2**
	Benzoic acid, 1-methylpropyl ester	123	-		
	Benzeneacetic acid, butyl ester	**91**	**1915**	**9.4 cult>C**	**5.1**

acid	**Acetic acid**	**43**	**1450**	**7.3 cult>C**	**2.5**
	Butanoic acid	**60**	**1627**	**5.6 cult>C**	**5.6**
	3-Methyl-butanoic acid (Isovaleric acid)	**60**	**1669**	**2.3 cult>C**	**2.6**
	Hexanoic acid	**60**	**-**	**2.2 cult>C**	**2.2**
	Octanoic acid	60	-		
S compounds	Dimethyl-disulfide (DMDS)	94	1085		
	Methyl-sec-butyl-disulfide	80	1269		
	Dimethyl-trisulfide (DMTS)	**126**	**1353**	**43.4 cult<C**	**23.9**
	Dimethyl-tetrasulfide	**158**	**1759**	**321.4 cult<C**	**53.6**
	S-methyl-thioacetate	**90**	**1055**	**365.6 cult>C**	**12.8**
	Thiazole	85	1251		
	2-Acetylthiazole	**99**	**1642**	**3.2 cult>C**	**3.2**
	3-(Methylthio)-propan-1-ol	**106**	**1724**	**18.4 cult>C**	**2.8**
	3-Phenyl-thiophene	160	-		
pyrazine	2-Methyl-pyrazine	94	1158		
	2,5-Dimethyl-pyrazine	108	1312		
	2,3,5-Trimethyl-pyrazine	123	1383		
	2-Ethyl-5-methyl-pyrazine	121	1382		
	3-Ethyl-2,5-dimethyl-pyrazine	135	1436		
	3,5-Diethyl-2-methyl-pyrazine	149	1485		
	2,3,5-Trimethyl-6-ethyl-pyrazine	**177**	**1504**	**12.7 cult>C**	**9.7**
	2-Isopropyl-pyrazine	107	1346		
	2-Methyl-3-isopropyl-pyrazine	121	1391		
	2,6-Dimethyl-3-*sec*-butyl-pyrazine	**134**	**1464**	**2.9 cult>C**	**2.9**
	2,5-dimethyl-3-(3-methylbutyl)-pyrazine	163	1651		
N comp.	Benzonitrile	103	1601		
	Indole	117	-		

[1] Ratio A presents maximal ratio of abundance of a compound, between the cultures and the control: (abundance in cultures) / (abundance in control), if cult>C, or (abundance in control) / (abundance in cultures), if cult<C, or cult~C, if abundance of compound in the control was higher than in some cultures, but lower than in others.

[2] Ratio B presents maximal ratio of abundance of a compound between the cultures.

Table 4: Compounds identified in Model system 2 along with linear retention indices (LRI) used for compound identification. The control was an un-inoculated model under the same conditions. Cultures and control were tested in triplicate. Compounds that exhibited significant differences in abundances in cultures, including control ($p<0.05$), are presented in bold.

Chemical group	Compound	Mass fragment used for analysis	LRI	Ratio A[1]	Ratio B[2]
aldehyde	**2-Methyl-propanal**	**72**	**-**	**30.9 cult<C**	**2.3**
	3-Methyl-butanal	**58**	**876**	**6.18 cult<C**	**2.8**
	3-Methyl-2-butenal	84	1202		
	Hexanal	**56**	**1106**	**30.5 cult<C**	**1.6**
	Benzaldehyde	**77**	**1518**	**4.4 cult<C**	**1.9**
	3-Methyl-benzaldehyde	119	1618		
	Benzeneacetaldehyde	91	1637		
	Furfural	**96**	**1463**	**19.3 cult<C**	**2.9**
ketone	Propan-2-one (Acetone)	58	-		
	1-Hydroxy-propan-2-one	**31**	**1298**	**2.7 cult>C**	**1.8**
	Butan-2-one	72	850		
	3-Hydroxy-butan-2-one (Acetoin)	45	1284		
	2,3-Butanedione (Diacetyl)	**43**	**985**	**3.2 cult>C**	**3.2**
	Pentan-2-one	43	969		
	2-Hydroxy-pentan-3-one	100	1356		
	Hexan-2-one	100	1105		
	Heptan-2-one	43	1189		
	Nonan-2-one	58	1384		
	Undecan-2-one	58	1591		
	Dodecan-2-one	156	1488		
	Tridecan-2-one	58	1802		
	Pentadecan-2-one	71	-		
	1-Phenylethanone (Acetophenone)	77	1646		
alcohol	Ethanol	46	968		
	2-Methyl-propan-1-ol	33	1129		
	Butan-1-ol	56	1168		
	3-Methyl-butan-1-ol	70	1220		
	Pentan-1-ol	42	1261		
	Heptan-2-ol	98	1324		
	Phenol	66	-		
	2-Phenylethanol (Phenyl-ethyl alcohol)	**91**	**-**	**15.2 cult>C**	**1.3**
	2,4-Di-tertbutylphenol	191	-		
ester	**Ethyl acetate**	**61**	**825**	**6.2 cult>C**	**1.9**
	Ethyl butanoate	101	1084		
acid	**Acetic acid**	**43**	**1448**	**7.2 cult>C**	**1.5**
	Propanoic acid	74	1538		
	2,2-Dimethyl-propanoic acid	102	1575		
	Butanoic acid	60	1623		
	3-Methyl-butanoic acid (Isovaleric acid)	60	1666		

	Hexanoic acid	60	-
	Heptanoic acid	116	-
	Octanoic acid	60	-
	Nonanoic acid	60	-
	Decanoic acid	60	-
	Benzoic acid	105	-
S comp.	Dimethyl-disulfide (DMDS)	94	1092
Pyrazine	2,5-Dimethyl-pyrazine	108	1315

[1]Ratio A presents maximal ratio of abundance of a compound, between the cultures and the control: (abundance in cultures) / (abundance in control), if cult>C, or (abundance in control) / (abundance in cultures), if cult<C.

[2] Ratio B presents maximal ratio of abundance of a compound between the cultures.

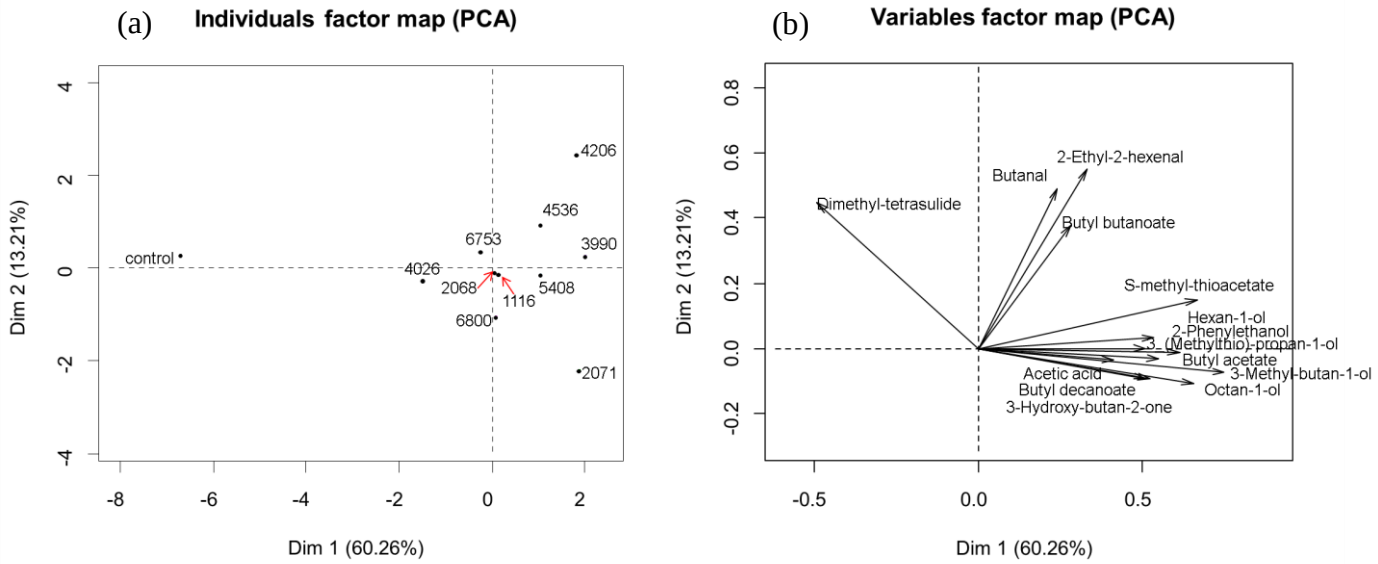


Figure 1: Individual factor map (a) and variable factor map (b) of principal component analysis (PCA) on 30 volatile compounds produced by ten strains of *Lactobacillus casei* group in Model system 1 incubated for 24 h at 30 °C. The control was a un-inoculated model system incubated under the same conditions. Cultures and control were tested in triplicate. The variables poorly represented in this plot (square cosinus limit below 0.8) are not shown, and only the 14 variables that are the best represented in dimension 1 (Dim1) and dimension 2 (Dim2) are shown. The DPC prefix has been removed from the strains name to avoid potential illegibility of the figure.

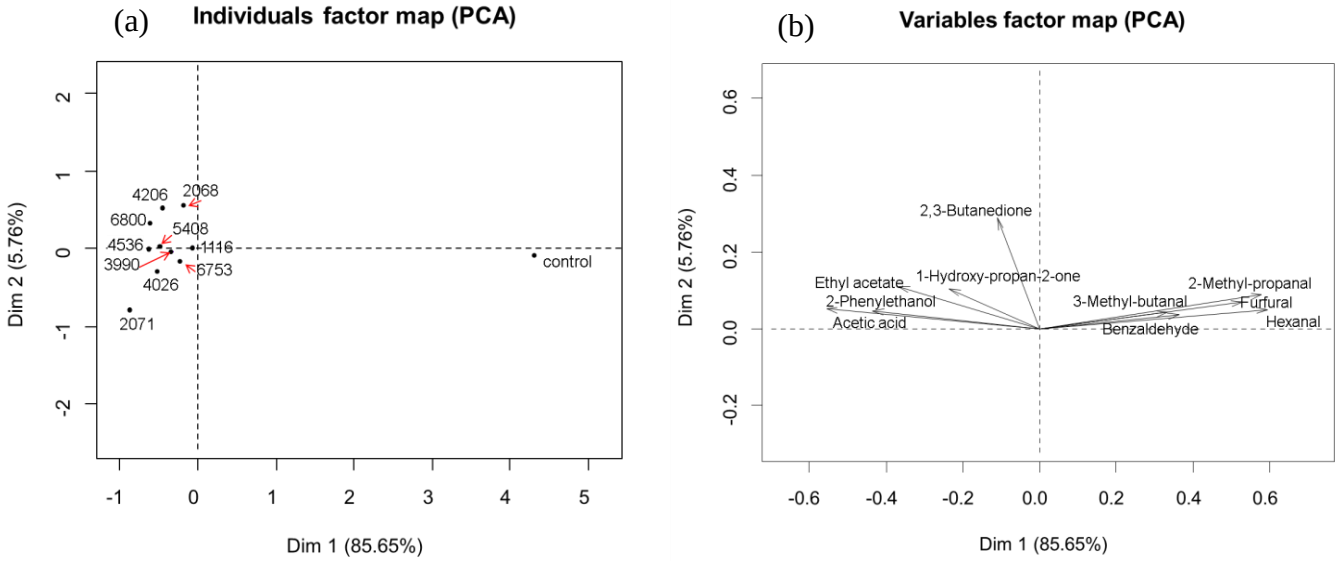


Figure 2: Individual factor map (a) and variable factor map (b) of principal component analysis (PCA) on 10 volatile compounds produced by ten strains of *Lactobacillus casei* group in Model system 2 incubated for 14 days at 30 °C. The control was an un-inoculated model system incubated under the same conditions. Cultures and control were tested in triplicate. Dim1- dimension 1, Dim2- dimension 2. The DPC prefix has been removed from the strains name to avoid potential illegibility of the figure.

Supporting Information Table 1: Abundances, in arbitrary units, standard deviations and least significant test of mean comparison for 30 compounds detected in significanlty different abundances ($p<0.05$) in chromatograms obtained by incubation of ten strains of *Lactobacillus casei* group for 48 h at 30° in Model system 1.

Butanal	Sample	DPC4206	DPC4536	DPC3990	DPC2068	DPC6753	DPC1116	DPC5408	DPC6800	DPC2071	DPC4026	Control
	abundance, arbitrary units	6.32E+09	1.95E+09	1.37E+09	9.62E+08	8.96E+08	7.71E+08	7.43E+08	4.51E+08	4.33E+08	3.93E+08	3.55E+08
	st. deviation	3.52E+09	3.54E+08	4.52E+08	3.92E+08	3.99E+08	1.65E+08	1.06E+08	9.57E+06	3.51E+08	4.94E+07	2.17E+07
	LSD test	a	b	b	b	b	b	b	b	b	b	b
S-methyl-thioacetate	Sample	DPC4206	DPC5408	DPC4536	DPC6753	DPC3990	DPC4026	DPC6800	DPC2071	DPC1116	DPC2068	Control
	abundance, arbitrary units	4.12E+08	1.94E+08	1.09E+08	1.06E+08	9.19E+07	8.80E+07	6.72E+07	6.49E+07	3.52E+07	3.23E+07	1.13E+06
	st. deviation	2.93E+08	7.11E+07	5.36E+07	2.04E+07	2.02E+07	5.01E+06	7.45E+06	3.31E+07	1.12E+07	1.11E+07	6.83E+05
	LSD test	a	b	bc	bc	bc	bc	bc	bc	c	c	c
Butyl acetate	Sample	DPC2068	DPC3990	DPC4206	DPC1116	DPC2071	DPC4536	DPC5408	DPC6800	DPC6753	DPC4026	Control
	abundance, arbitrary units	1.61E+10	1.50E+10	1.15E+10	1.09E+10	9.85E+09	9.71E+09	8.71E+09	8.60E+09	7.77E+09	6.32E+09	3.02E+08
	st. deviation	4.22E+09	4.19E+09	1.52E+09	6.23E+09	1.70E+09	2.88E+09	1.77E+09	3.09E+09	9.54E+08	2.74E+09	7.11E+07
	LSD test	a	ab	abc	abc	bc	bc	c	c	c	c	d
3-Methyl-butan-1-ol	Sample	DPC2071	DPC3990	DPC5408	DPC4536	DPC4206	DPC1116	DPC6800	DPC6753	DPC4026	DPC2068	Control
	abundance, arbitrary units	1.44E+10	1.35E+10	1.30E+10	1.19E+10	1.13E+10	1.11E+10	1.04E+10	9.67E+09	9.64E+09	7.52E+09	6.99E+07
	st. deviation	9.04E+08	9.58E+08	1.61E+09	4.09E+08	9.03E+08	1.99E+09	7.68E+08	1.60E+09	1.30E+09	1.32E+09	5.20E+07
	LSD test	a	ab	abc	bcd	cde	cde	de	e	e	f	g
Butyl butanoate	Sample	DPC4206	DPC4536	DPC6753	DPC1116	DPC2068	DPC3990	DPC5408	DPC2071	DPC4026	DPC6800	Control
	abundance, arbitrary units	3.10E+09	1.46E+09	6.13E+08	6.04E+08	6.00E+08	4.21E+08	3.98E+08	3.83E+08	3.80E+08	3.46E+08	2.04E+08
	st. deviation	1.41E+09	4.96E+08	1.81E+08	1.07E+08	1.66E+08	1.27E+08	2.42E+08	2.04E+08	1.51E+08	1.45E+08	1.10E+08
	LSD test	a	b	c	c	c	c	c	c	c	c	c
Ethyl hexanoate	Sample	DPC4536	DPC6753	DPC2068	DPC3990	DPC1116	DPC2071	DPC4206	DPC5408	DPC4026	DPC6800	Control
	abundance, arbitrary units	8.01E+06	5.41E+06	5.13E+06	5.12E+06	4.64E+06	3.82E+06	3.72E+06	1.96E+06	1.24E+06	1.17E+06	6.88E+05
	st. deviation	2.44E+06	3.23E+06	1.07E+06	1.49E+06	1.91E+06	9.24E+05	2.34E+06	3.71E+05	7.54E+05	1.88E+05	2.83E+05
	LSD test	a	ab	b	b	bc	bcd	bcd	cde	de	de	e
3-Hydroxy-butan-2-one (Acetoin)	Sample	DPC2071	DPC3990	DPC5408	DPC4536	DPC4206	DPC2068	DPC1116	DPC6800	DPC6753	DPC4026	Control
	abundance, arbitrary units	1.48E+09	9.92E+08	7.37E+08	5.01E+08	4.81E+08	3.79E+08	3.55E+08	2.07E+08	1.99E+08	1.26E+08	1.85E+07
	st. deviation	3.23E+08	4.05E+08	3.24E+08	3.57E+07	1.21E+08	1.01E+08	6.62E+07	3.67E+07	5.21E+07	3.82E+07	5.00E+06
	LSD test	a	b	bc	cd	cd	de	de	def	def	ef	f

2-Ethyl-2-hexenal	Sample	DPC4206	DPC3990	DPC4536	DPC6753	DPC2068	DPC1116	DPC5408	DPC4026	DPC2071	Control	DPC6800
	abundance, arbitrary units	2.37E+07	5.38E+06	3.00E+06	1.02E+06	9.70E+05	8.25E+05	6.98E+05	6.12E+05	5.45E+05	4.31E+05	2.12E+05
	st. deviation	2.12E+07	4.19E+06	1.50E+06	2.09E+04	5.16E+05	5.12E+05	6.12E+05	3.87E+05	4.89E+05	4.23E+05	1.48E+05
	LSD test	a	b	b	b	b	b	b	b	b	b	b
Hexan-1-ol	Sample	DPC3990	DPC5408	DPC4536	DPC2071	DPC4206	DPC1116	DPC6753	DPC6800	DPC2068	DPC4026	Control
	abundance, arbitrary units	1.30E+08	7.79E+07	6.30E+07	4.92E+07	4.03E+07	3.49E+07	3.35E+07	2.44E+07	1.99E+07	1.35E+07	3.40E+06
	st. deviation	4.73E+07	3.45E+07	1.31E+06	2.87E+07	2.27E+07	2.47E+07	1.90E+07	1.31E+07	6.98E+06	8.67E+06	6.49E+05
	LSD test	a	b	bc	bcd	bcde	cde	cde	cde	de	de	e
Dimethyl-trisulfide (DMTS)	Sample	Control	DPC6753	DPC4026	DPC5408	DPC1116	DPC6800	DPC2068	DPC3990	DPC4536	DPC4206	DPC2071
	abundance, arbitrary units	3.50E+09	1.93E+09	1.63E+09	1.56E+09	1.37E+09	1.07E+09	1.04E+09	9.88E+08	9.29E+08	6.23E+08	8.07E+07
	st. deviation	2.67E+09	7.69E+08	6.86E+08	5.56E+08	3.14E+08	3.25E+08	1.83E+08	3.08E+08	6.16E+08	1.84E+08	2.14E+07
	LSD test	a	b	b	bc	bc	bc	bc	bc	bc	bc	c
Nonan-2-one	Sample	DPC3990	DPC2071	DPC4206	Control	DPC4026	DPC6800	DPC6753	DPC5408	DPC4536	DPC2068	DPC1116
	abundance, arbitrary units	7.28E+07	6.46E+07	4.28E+07	4.19E+07	3.85E+07	3.46E+07	3.46E+07	3.22E+07	2.43E+07	2.27E+07	1.21E+07
	st. deviation	4.77E+06	1.28E+07	9.17E+06	1.86E+07	9.63E+05	3.95E+06	6.98E+06	1.49E+06	2.75E+06	1.59E+06	1.11E+06
	LSD test	a	a	b	b	b	bc	bc	bc	cd	cd	d
Nonanal	Sample	Control	DPC2068	DPC5408	DPC4536	DPC4206	DPC1116	DPC3990	DPC6800	DPC6753	DPC4026	DPC2071
	abundance, arbitrary units	1.38E+07	4.83E+06	4.62E+06	3.95E+06	3.94E+06	3.85E+06	3.32E+06	3.29E+06	3.02E+06	2.74E+06	1.81E+06
	st. deviation	1.05E+07	3.11E+06	2.99E+05	1.04E+06	1.16E+06	5.08E+05	1.01E+06	3.37E+05	1.85E+05	8.03E+05	7.93E+05
	LSD test	a	b	b	b	b	b	b	b	b	b	b
3-Methyl-hexan-1-ol	Sample	DPC3990	DPC1116	DPC2071	DPC5408	DPC2068	DPC4536	DPC4206	DPC6800	Control	DPC4026	DPC6753
	abundance, arbitrary units	1.64E+08	7.90E+07	7.74E+07	5.93E+07	5.18E+07	4.70E+07	3.32E+07	1.80E+07	4.10E+06	3.79E+06	3.70E+06
	st. deviation	3.51E+07	5.11E+07	1.25E+07	1.43E+07	9.64E+06	8.31E+06	1.63E+06	3.24E+06	2.14E+06	3.61E+05	5.63E+05
	LSD test	a	b	b	bc	bcd	bcd	cde	de	e	e	e
Acetic acid	Sample	DPC2071	DPC4206	DPC3990	DPC5408	DPC1116	DPC6753	DPC4026	DPC4536	DPC2068	DPC6800	Control
	abundance, arbitrary units	5.21E+09	4.19E+09	3.92E+09	2.74E+09	2.48E+09	2.30E+09	2.27E+09	2.24E+09	2.15E+09	2.09E+09	7.11E+08
	st. deviation	1.61E+09	8.65E+08	1.20E+09	8.43E+08	8.87E+08	6.11E+08	3.92E+08	4.71E+08	4.61E+08	5.71E+08	4.00E+08
	LSD test	a	a	ab	bc	c	c	c	c	c	cd	d
2,6-Dimethyl-3-sec-butyl-pyrazine	Sample	DPC5408	DPC4536	DPC2068	DPC1116	DPC3990	DPC6800	DPC2071	DPC6753	DPC4206	Control	DPC4026
	abundance, arbitrary units	8.44E+06	6.29E+06	6.19E+06	6.05E+06	5.61E+06	5.26E+06	5.09E+06	5.05E+06	4.93E+06	3.64E+06	2.96E+06
	st. deviation	2.09E+06	1.81E+06	9.57E+05	6.38E+05	1.31E+06	1.47E+06	2.67E+06	1.85E+06	1.24E+06	1.71E+06	5.20E+05
	LSD test	a	ab	ab	ab	bc	bc	bc	bc	bc	bc	c
2,3,5,-Trimethyl-6-ethylpyrazine	Sample	DPC5408	DPC2068	DPC1116	DPC4536	DPC4206	DPC2071	DPC3990	DPC6800	DPC6753	DPC4026	Control
	abundance, arbitrary units	4.35E+06	3.47E+06	3.37E+06	2.65E+06	2.51E+06	1.83E+06	1.82E+06	1.60E+06	1.43E+06	4.49E+05	3.43E+05

	st. deviation	2.13E+06	4.54E+05	1.42E+06	1.41E+06	1.76E+06	8.56E+05	1.12E+06	1.10E+06	1.19E+06	2.59E+05	1.35E+05
	LSD test	a	ab	ab	ab	abc	bcd	bcd	bcd	bcd	cd	d
Octan-1-ol	Sample	DPC2071	DPC4536	DPC5408	DPC4206	DPC6800	DPC3990	DPC6753	DPC1116	DPC2068	DPC4026	Control
	abundance, arbitrary units	4.03E+07	3.59E+07	3.19E+07	2.09E+07	2.03E+07	1.88E+07	1.77E+07	1.76E+07	1.51E+07	1.18E+07	2.35E+06
	st. deviation	1.47E+07	7.67E+06	1.52E+07	4.58E+05	9.24E+06	6.40E+06	6.63E+06	5.75E+06	4.94E+06	1.26E+07	3.18E+06
	LSD test	a	ab	b	c	c	cd	cd	cd	de	e	f
Butanedioic acid, dimethyl ester	Sample	DPC3990	DPC6800	DPC2071	DPC4206	DPC2068	DPC5408	DPC4536	DPC1116	DPC4026	DPC6753	Control
	abundance, arbitrary units	1.49E+07	1.43E+07	1.37E+07	1.37E+07	1.26E+07	9.84E+06	9.61E+06	9.08E+06	7.35E+06	6.65E+06	6.08E+06
	st. deviation	1.67E+06	3.20E+06	2.67E+06	2.73E+06	8.54E+05	4.52E+06	3.83E+06	1.66E+06	4.20E+06	3.07E+06	2.70E+06
	LSD test	a	ab	abc	abc	abc	abcd	bcd	cd	d	d	d
Undecan-2-one	Sample	DPC3990	DPC4206	DPC6753	DPC2071	DPC6800	DPC5408	DPC4026	DPC2068	DPC4536	DPC1116	Control
	abundance, arbitrary units	1.50E+08	6.71E+07	5.37E+07	5.13E+07	3.97E+07	3.49E+07	2.07E+07	1.74E+07	1.60E+07	1.22E+07	7.57E+06
	st. deviation	6.14E+07	3.35E+07	2.65E+07	1.18E+07	5.49E+06	4.02E+06	1.77E+06	4.92E+06	1.67E+06	4.23E+06	2.71E+06
	LSD test	a	b	bc	bc	bcd	bcd	cd	cd	cd	d	d
Butanoic acid	Sample	DPC4206	DPC4536	DPC2071	DPC1116	DPC3990	Control	DPC4026	DPC6753	DPC5408	DPC2068	DPC6800
	abundance, arbitrary units	6.67E+09	2.20E+09	1.82E+09	1.48E+09	1.44E+09	1.44E+09	1.43E+09	1.39E+09	1.35E+09	1.25E+09	1.19E+09
	st. deviation	1.11E+09	5.66E+08	9.24E+08	2.87E+08	2.02E+08	5.79E+08	3.87E+08	4.27E+08	4.49E+08	2.17E+08	1.21E+08
	LSD test	a	b	bc	bc	bc	bc	bc	bc	bc	c	c
3-Methyl-butanoic acid	Sample	DPC4206	DPC4536	DPC3990	DPC2071	DPC6800	Control	DPC5408	DPC1116	DPC2068	DPC4026	DPC6753
	abundance, arbitrary units	3.62E+08	2.15E+08	2.09E+08	2.08E+08	2.05E+08	1.84E+08	1.73E+08	1.70E+08	1.58E+08	1.56E+08	1.38E+08
	st. deviation	6.32E+07	2.08E+07	3.50E+07	9.92E+07	1.54E+07	8.22E+07	5.54E+07	4.12E+07	4.27E+07	2.60E+07	3.17E+07
	LSD test	a	b	b	b	b	b	b	b	b	b	b
Undecan-2-ol	Sample	DPC4206	DPC5408	DPC2071	DPC3990	DPC6753	DPC4536	DPC6800	DPC1116	DPC4026	DPC2068	Control
	abundance, arbitrary units	5.54E+07	2.44E+07	2.26E+07	2.14E+07	1.63E+07	1.15E+07	1.02E+07	6.19E+06	6.09E+06	5.48E+06	2.87E+06
	st. deviation	4.35E+07	3.97E+06	4.95E+06	1.14E+07	9.06E+06	5.57E+06	4.38E+06	2.27E+06	4.20E+06	2.53E+06	6.51E+05
	LSD test	a	b	b	b	b	b	b	b	b	b	b
3-(Methylthio)-propan-1-ol	Sample	DPC1116	DPC6800	DPC5408	DPC3990	DPC4206	DPC2071	DPC2068	DPC4536	DPC6753	DPC4026	Control
	abundance, arbitrary units	7.42E+07	6.83E+07	6.71E+07	6.48E+07	6.44E+07	6.15E+07	5.90E+07	5.63E+07	5.02E+07	2.63E+07	4.03E+06
	st. deviation	1.73E+07	5.80E+06	6.87E+06	3.35E+07	1.30E+07	3.09E+07	8.81E+06	1.05E+07	1.91E+07	4.67E+06	2.58E+06
	LSD test	a	a	a	a	a	a	a	a	ab	bc	c
Dimethyl-tetrasulfide	Sample	Control	DPC4206	DPC4026	DPC6753	DPC5408	DPC4536	DPC2068	DPC3990	DPC1116	DPC6800	DPC2071
	abundance, arbitrary units	3.17E+07	5.28E+06	4.53E+06	3.94E+06	3.70E+06	2.75E+06	2.29E+06	2.22E+06	1.99E+06	1.25E+06	9.85E+04
	st. deviation	1.72E+07	5.43E+06	1.97E+06	1.85E+06	1.49E+06	2.73E+06	1.09E+06	4.98E+05	8.39E+05	5.79E+05	6.99E+04
	LSD test	a	b	b	b	b	b	b	b	b	b	b

Tridecan-2-one	Sample	DPC3990	DPC4206	DPC1116	DPC6800	DPC6753	DPC2071	DPC5408	DPC4536	DPC2068	DPC4026	Control
	abundance, arbitrary units	3.70E+08	1.37E+08	1.35E+08	1.32E+08	1.20E+08	1.12E+08	1.06E+08	7.93E+07	6.90E+07	5.14E+07	2.07E+07
	st. deviation	1.43E+08	1.47E+07	6.67E+07	4.99E+06	5.60E+07	3.71E+07	3.24E+07	6.69E+06	1.43E+07	1.84E+07	1.74E+07
	LSD test	a	b	b	b	b	b	bc	bc	bc	bc	c
Butyl decanoate	Sample	DPC2071	DPC5408	DPC2068	DPC6800	DPC3990	DPC4536	DPC1116	DPC4206	DPC6753	DPC4026	Control
	abundance, arbitrary units	9.35E+06	9.00E+06	8.58E+06	8.32E+06	8.01E+06	8.01E+06	7.68E+06	6.65E+06	6.44E+06	6.23E+06	9.09E+05
	st. deviation	3.37E+06	1.49E+06	1.79E+06	1.67E+06	3.36E+06	3.67E+06	1.24E+06	2.58E+06	2.39E+06	2.45E+06	8.00E+05
	LSD test	a	a	a	a	a	a	a	a	a	a	b
Hexanoic acid	Sample	DPC4206	DPC4536	DPC6800	DPC1116	Control	DPC2068	DPC4026	DPC5408	DPC2071	DPC3990	DPC6753
	abundance, arbitrary units	1.25E+09	9.10E+08	7.86E+08	7.47E+08	7.23E+08	7.09E+08	6.94E+08	6.68E+08	6.64E+08	5.69E+08	5.60E+08
	st. deviation	1.90E+08	1.35E+08	6.26E+07	2.32E+08	3.53E+08	1.31E+08	2.02E+08	1.72E+08	3.79E+08	1.17E+08	1.56E+08
	LSD test	a	ab	b	b	b	b	b	b	b	b	b
2-Phenylethanol	Sample	DPC4536	DPC4206	DPC6800	DPC5408	DPC2071	DPC4026	DPC3990	DPC1116	DPC2068	DPC6753	Control
	abundance, arbitrary units	5.68E+08	5.61E+08	4.95E+08	4.66E+08	4.35E+08	4.33E+08	4.01E+08	3.07E+08	2.66E+08	2.54E+08	5.18E+07
	st. deviation	1.05E+08	4.34E+07	1.36E+08	1.56E+08	1.61E+08	1.34E+08	5.70E+07	8.49E+07	3.20E+07	5.17E+07	5.00E+07
	LSD test	a	a	a	ab	abc	abc	abcd	bcd	cd	d	e
Benzeneacetic acid, butyl ester	Sample	DPC3990	DPC2071	DPC6800	DPC4206	DPC5408	DPC4536	DPC2068	DPC4026	DPC1116	DPC6753	Control
	abundance, arbitrary units	3.73E+07	2.55E+07	2.25E+07	1.84E+07	1.57E+07	1.36E+07	1.31E+07	9.40E+06	8.04E+06	7.38E+06	3.97E+06
	st. deviation	1.61E+07	1.34E+07	5.25E+06	5.52E+06	4.29E+06	4.50E+05	3.97E+06	1.59E+06	1.70E+06	1.37E+06	2.21E+06
	LSD test	a	ab	bc	bcd	bcde	cde	cde	de	de	de	e
2-Acetylthiazole	Sample	DPC6753	DPC4026	DPC5408	DPC4536	DPC2068	DPC4206	DPC6800	DPC2071	Control	DPC1116	DPC3990
	abundance, arbitrary units	8.32E+06	6.28E+06	5.98E+06	5.69E+06	4.46E+06	4.32E+06	3.56E+06	3.54E+06	3.38E+06	3.03E+06	2.59E+06
	st. deviation	3.97E+06	1.74E+06	5.97E+05	8.02E+05	5.97E+05	8.35E+05	3.18E+05	4.78E+05	1.83E+06	2.65E+05	1.60E+05
	LSD test	a	ab	abc	bcd	bcde	bcde	cde	cde	de	e	e

Supporting Information Table 2: Abundances, in arbitrary units, standard deviations and least significant test of mean comparison for 10 compounds detected in significantly different abundances ($p<0.05$) in chromatograms obtained by incubation of ten strains of *Lactobacillus casei* group for 14 days at 30° in Model system 2.

2-Methyl-propanal	Sample	Control	DPC2068	DPC4206	DPC3990	DPC6800	DPC5408	DPC4536	DPC1116	DPC6753	DPC4026	DPC2071
	abundance, arbitrary units	1.31E+08	9.82E+06	9.32E+06	7.51E+06	7.34E+06	7.24E+06	6.89E+06	6.51E+06	6.42E+06	5.36E+06	4.23E+06
	st. deviation	1.07E+07	2.33E+06	5.29E+06	1.62E+06	2.86E+06	2.85E+05	5.96E+06	1.99E+06	4.05E+06	1.18E+06	1.77E+06
	LSD test	a	b	b	b	b	b	b	b	b	b	b
Ethyl acetate	Sample	DPC2068	DPC4026	DPC3990	DPC6800	DPC4536	DPC6753	DPC5408	DPC4206	DPC1116	DPC2071	Control
	abundance, arbitrary units	1.82E+07	1.57E+07	1.35E+07	1.33E+07	1.27E+07	1.11E+07	1.09E+07	1.08E+07	1.00E+07	9.39E+06	2.91E+06
	st. deviation	3.70E+06	1.12E+07	1.41E+06	9.69E+05	3.90E+06	6.82E+05	1.79E+06	2.47E+06	1.78E+06	3.59E+05	5.15E+05
	LSD test	a	ab	ab	ab	ab	ab	b	b	b	bc	c
3-Methyl-butanal	Sample	Control	DPC1116	DPC3990	DPC4206	DPC4536	DPC6753	DPC4026	DPC6800	DPC5408	DPC2068	DPC2071
	abundance, arbitrary units	2.16E+09	9.88E+08	6.74E+08	6.62E+08	6.57E+08	5.17E+08	5.15E+08	4.77E+08	4.70E+08	3.90E+08	3.49E+08
	st. deviation	2.67E+07	2.84E+08	5.54E+07	1.44E+08	1.87E+08	7.99E+07	2.41E+08	6.08E+07	4.14E+07	4.44E+07	5.87E+07
	LSD test	a	b	bc	bc	bc	c	c	c	c	c	c
2,3-Butanedione	Sample	DPC4206	DPC6800	DPC2068	DPC4536	DPC5408	DPC1116	DPC4026	DPC6753	DPC3990	Control	DPC2071
	abundance, arbitrary units	8.72E+09	6.50E+09	5.78E+09	5.61E+09	5.46E+09	4.71E+09	4.26E+09	3.65E+09	3.53E+09	3.18E+09	2.73E+09
	st. deviation	3.09E+09	2.15E+09	7.23E+08	2.33E+09	1.35E+09	3.07E+08	1.07E+09	9.37E+08	6.81E+07	1.75E+08	4.79E+08
	LSD test	a	ab	bc	bcd	bcd	bcde	bcde	bcde	cde	de	e
Hexanal	Sample	Control	DPC2068	DPC6753	DPC1116	DPC6800	DPC5408	DPC3990	DPC4026	DPC4206	DPC4536	DPC2071
	abundance, arbitrary units	1.11E+08	9.32E+06	6.17E+06	5.75E+06	5.70E+06	5.44E+06	5.37E+06	5.21E+06	5.13E+06	4.24E+06	3.63E+06
	st. deviation	1.77E+07	2.13E+06	2.94E+06	1.81E+06	2.49E+06	1.10E+06	1.71E+06	8.04E+05	2.37E+06	1.18E+06	1.07E+06
	LSD test	a	b	b	b	b	b	b	b	b	b	b

1-Hydroxy-propan-2-one	Sample	DPC2068	DPC4206	DPC3990	DPC6800	DPC5408	DPC2071	DPC6753	DPC1116	DPC4536	DPC4026	Control
	abundance, arbitrary units	4.44E+07	3.98E+07	3.96E+07	3.73E+07	3.68E+07	3.63E+07	3.32E+07	3.09E+07	3.09E+07	2.54E+07	1.65E+07
	st. deviation	1.95E+06	9.65E+06	6.29E+06	5.18E+06	3.55E+06	3.05E+06	1.76E+06	4.71E+06	1.30E+07	9.46E+06	9.92E+05
	LSD test	a	ab	ab	ab	ab	abc	abc	bc	bc	cd	d
Acetic acid	Sample	DPC3990	DPC6800	DPC6753	DPC4536	DPC1116	DPC2071	DPC2068	DPC4206	DPC4026	DPC5408	Control
	abundance, arbitrary units	8.55E+09	8.22E+09	7.49E+09	6.44E+09	6.32E+09	6.28E+09	6.06E+09	6.05E+09	5.88E+09	5.75E+09	1.20E+09
	st. deviation	2.15E+09	1.62E+09	2.96E+09	1.35E+09	1.95E+09	1.77E+09	6.55E+08	7.51E+08	2.56E+09	5.47E+08	6.48E+08
	LSD test	a	ab	ab	ab	ab	ab	ab	ab	ab	b	c
Furfural	Sample	Control	DPC2068	DPC6753	DPC1116	DPC3990	DPC4536	DPC6800	DPC4206	DPC4026	DPC5408	DPC2071
	abundance, arbitrary units	4.78E+07	7.07E+06	6.39E+06	5.62E+06	4.83E+06	4.19E+06	3.79E+06	3.61E+06	3.40E+06	2.59E+06	2.47E+06
	st. deviation	5.45E+06	1.10E+06	2.86E+06	2.96E+06	9.49E+05	4.94E+05	3.34E+05	1.74E+06	1.50E+06	4.57E+05	9.06E+05
	LSD test	a	b	bc	bc	bc	bc	bc	bc	bc	c	c
Benzaldehyde	Sample	Control	DPC6753	DPC1116	DPC3990	DPC2068	DPC4206	DPC5408	DPC6800	DPC4026	DPC4536	DPC2071
	abundance, arbitrary units	3.46E+08	1.49E+08	1.45E+08	1.32E+08	1.22E+08	1.12E+08	1.03E+08	9.85E+07	9.82E+07	8.46E+07	7.86E+07
	st. deviation	4.48E+07	3.29E+04	6.54E+06	2.09E+07	1.83E+07	1.77E+07	4.02E+07	6.32E+06	2.96E+07	5.82E+07	1.10E+07
	LSD test	a	b	b	bc	bcd	bcd	bcd	bcd	bcd	cd	d
2-Phenylethanol	Sample	DPC4536	DPC4206	DPC6800	DPC4026	DPC2071	DPC3990	DPC6753	DPC2068	DPC1116	DPC5408	Control
	abundance, arbitrary units	2.45E+07	2.33E+07	2.28E+07	2.17E+07	2.16E+07	2.13E+07	2.12E+07	2.09E+07	2.07E+07	1.88E+07	1.61E+06
	st. deviation	4.50E+06	2.16E+06	1.75E+06	6.78E+06	3.98E+05	6.70E+05	3.34E+06	2.40E+06	4.04E+06	4.41E+06	7.07E+05
	LSD test	a	a	a	a	a	a	a	a	a	a	b

Chapter 4

Variation of volatile profiles resulting from the choice of extraction and separation techniques - examples from a cheese model system

Ewelina Stefanovic chapter contributions:

Experimental work:

- Performed experimental work described in this chapter

Results interpretation:

- Analysis of all data of work described in this manuscript

Manuscript preparation:

- Major contribution to manuscript preparation

4.1 Abstract

The most commonly used methods for volatile analysis in food include headspace (HS) based volatile extraction followed by separation and identification using gas chromatography (GC) and mass spectrometry (MS). This study examined the inter-laboratory variation of volatile analysis with methods based on different extraction and GC methods on the identification of volatiles generated by ten strains of *Lactobacillus paracasei* in a cheese model system. Method A consisted of HS-Solid Phase Microextraction (HS-SPME) using a non-polar GC column, while Method B consisted of HS-Trap extraction using a polar column. Two methods had similar values for Limits of Detection (LOD) and Quantification (LOQ). Higher numbers of alcohols, esters, and acids were detected using Method A, while Method B detected more short-chain aldehydes and ketones, pyrazine derivatives, and specific sulfur-containing compounds. The variations in volatile profiles led to differences in discrimination of the most different samples in the analysed set, suggesting the importance of the choice of HS GC-MS method.

Keywords: extraction, GC-MS, *Lactobacillus*, SPME, headspace

4.2 Introduction

In the overall experience of eating, the first aspect of food flavour that a consumer encounters is aroma, which consists of volatile compounds (Tunick, 2014), followed by taste (consisting of non-volatile compounds) and mouth feel perception of texture (Laing and Jinks, 1996). Understanding and improving the formation of food flavour requires comprehensive analytical approaches to identify flavour contributing compounds (Thomsen et al., 2014). The most commonly used method to analyse aroma volatiles in numerous fields, including the food domain is gas chromatography (GC) coupled to mass spectrometry (MS) (Lehotay and Hajšlová, 2002). Prior to separation by GC, compounds are extracted from the sample, using various techniques based on their volatility and/or their polarity, i.e. distillation, solvent extraction, or headspace-based techniques. Headspace (HS) represents the atmosphere above the sample in which volatiles diffuse (Soria et al., 2015). HS-related techniques have become very popular and are widely used for volatile extraction from food samples, since they are relatively simple and automatable, require little sample preparation, and extract most volatile compounds. Additionally, data obtained by analysis of HS extracts is considered to be closely related to results of descriptive sensory analysis (Bosset et al., 1995, Lawlor et al., 2002). The obtained extracts are "cleaner" and they do not contain traces of solvent or artefacts (Plutowska and Wardencki, 2007). Concerning dairy products, which are abundant in fat and proteins, HS sampling is a method of choice, as it prevents the adsorption of non-volatile compounds that may interfere with the analysis (Marsili, 2011).

HS extraction can be performed in a static or dynamic manner. In static headspace (SHS) extraction, only a portion of the headspace is analysed (Soria et al., 2015).

The fundamentals of SHS extraction have been the subject of a number of reviews (Snow and Bullock, 2010, Soria et al., 2015). SHS extraction is routinely used in a wide range of disciplines including food sciences. It is simple, reliable, easily automated, but only extracts the most abundant and most volatile compounds (Tunick, 2014). In dynamic headspace (DHS) extraction techniques, the volatiles are purged by an inert gas and transferred from the headspace to a trap containing a solid sorbent, most often Tenax® (Idris et al., 2010), on which they are pre-concentrated (Snow and Slack, 2002, Tunick, 2014, Soria et al., 2015). They are then thermally desorbed and transferred into the GC injection port. DHS techniques are very sensitive, however they are often time-consuming with lower reproducibility (Kilcawley, 2017). In between these two approaches, many other HS-related techniques have been developed. The most widely used, especially in food analysis, is headspace solid-phase microextraction (HS-SPME) (Jelen et al., 2012). This method is based on using a fibre that adsorbs volatiles from the headspace and desorbs them into a GC port (Tunick, 2014). Headspace-Trap (HS-Trap) is another more recently developed technique, also used in the food domain (Schulz et al., 2007, Nikfardjam and Maier, 2011, Aberl and Coelhan, 2012, Pogacic et al., 2015, Bosse Nee Danz et al., 2017). HS-Trap differs from DHS methods as it is not based on a continuous flow of gas, but uses a carrier gas to reach the desired pressure in the vial with sample, before the pressurised headspace is sent to the trap (Barani et al., 2006). This step of gas injection and headspace removal can be repeated up to four times (Barani et al., 2006). Both HS-SPME and HS-Trap can be carried out using automated sampling devices coupled to GC-MS.

The effect of the extraction method on the volatile profiles of various food samples has been the subject of several studies (Elmore et al., 1997, Contarini and Povolo,

2002, Povolo and Contarini, 2003, Mallia et al., 2005, Liu et al., 2007, Murat et al., 2012). These studies compared SPME and DHS-based methods and showed that different volatile profiles were obtained for the same samples, depending on the extraction method. In addition, some studies showed marked differences even between DHS techniques (Barron et al., 2005), or different SPME fibers (Mondello et al., 2005, Feng et al., 2016).

Apart from the effect of sorbent used for the extraction, factors such as the temperature of thermostating and time of extraction, "salting out", pH, type of GC column and characteristics of the mass spectrometer also contribute to differences in the volatiles detected by different methods, with lower or higher impact. The effects of these factors are mainly discussed in optimization of a method for analysis of a certain type of sample, for example heated rapeseed oil (Sghaier et al., 2016), but hard to elucidate in a direct comparison of different analytical approaches.

In bacterial ripened cheeses, aroma development mainly results from the metabolic activity of bacteria present in cheese during ripening (McSweeney and Sousa, 2000). Starter, non-starter and adjunct bacteria possess numerous enzymes that are able to degrade available substrates to compounds that are perceived as aroma. The main metabolic reactions that occur during milk fermentation and contribute to aroma development include the catabolism of lactose and citrate, lipolysis, and proteolysis. These metabolic processes generate volatile organic molecules of different chemical groups (alcohols, aldehydes, ketones, acids, esters, sulfur compounds) that contribute to the characteristic aromatic notes of cheese (McSweeney and Sousa, 2000, Marilley and Casey, 2004).

In this study, we aimed to explore the effect of an analytical method on the identification of cheese aroma contributing volatiles and on the differentiation between the samples analysed. For this, the samples (cultures of strains of the bacterium *Lactobacillus paracasei* inoculated in a cheese model system) were analysed with two methods. The methods differed in headspace extraction steps (HS-SPME vs. HS-Trap) and used different types of columns (non-polar vs. polar). To our knowledge, the methods based on these two extraction techniques have been compared once so far (Sghaier et al., 2016).

4.3 Materials and methods

4.3.1 Standards for LOD and LOQ determination

A set of ten standards were prepared to determine the limits of detection (LOD) and of quantification (LOQ) of the two methods. These standards consisted of ethyl acetate, ethyl propanoate, ethyl butanoate, ethyl hexanoate, 3-methyl-butanal, benzaldehyde, 2-heptanone, 2,3-butanedione, dimethyl-disulfide (DMDS), and 3-methyl-butan-1-ol, as previously described (Pogacic et al., 2015). The concentration of compounds ranged from 5 to 1800 ng/g, with the exception of 3-methyl-butan-1-ol where concentration ranged from 300 to 66300 ng/g, The standards were analysed by both Method A and Method B, (for Method B, results are reported in Pogacic et al., 2015 and confirmed in the actual run of samples), with blank samples run after each standard.

4.3.2 Cheese model system and sample preparation

The preparation of the cheese model system and samples were performed as previously described (Stefanovic et al., 2017). Briefly, ten *Lactobacillus paracasei* strains from the Teagasc Food Research Centre, Moorepark DPC Culture collection (DPC1116, DPC2068, DPC2071, DPC3990, DPC4026, DPC4206, DPC4536, DPC5408, DPC6753, DPC6800) were inoculated (1 % v/v) in a concentrated amino acid-rich medium Bacto®Tryptone (BD, Oxford, UK) (35 % w/v) supplemented with 12 g/L of NaCl and incubated for 24 h at 30°C. All samples (cultures of inoculated *L. paracasei* strains and the control, which was an un-inoculated model system) were prepared in triplicate, and they were kept at -80°C until GC-MS analysis was performed.

4.3.3 Method A: HS-SPME GC-MS

For each sample, 2 g were placed in a 20 mL screw capped HS-SPME vial with a silicone/PTFE septum vial (Apex Scientific, Maynooth, Ireland). Samples in vials were equilibrated to 40°C for 10 min with pulsed agitation of 5 s at 500 rpm. A Shimadzu AOC 5000 plus auto-sampler was used for sample introduction (Mason Technology, Dublin, Ireland). A single 50/30 μm Carboxen®/ divinylbenzene/ polydimethylsiloxane (CAR/DVB/PDMS) fibre was used to perform HS-SPME (Agilent Technologies, Cork, Ireland). The SPME fibre was exposed to the headspace above the samples for 20 minutes at 40°C. After extraction, the fibre was injected into the GC inlet and desorbed for 2 min at 250 °C. The fibre was pre-conditioned using a bakeout station in a nitrogen flow at 270°C for 3 min between samples to ensure no carry over occurred between injections. Injections were made on Shimadzu 2010 Plus GC (Mason Technology) with an DB-5 (60 m × 0.25 mm × 0.25 μm) column (Agilent Technologies) using a split/splitless injector (split mode was in 1:10). Helium was used as a carrier gas at a fixed pressure at 23 psi. The temperature of the column oven was set at 35°C, held for 5 min, increased at 6.5°C/min to 230°C then increased at 15°C/min to 320°C. The mass spectrometer detector Shimadzu TQ8030 was run in single quad mode (Mason Technology). The ion source temperature was 230°C, the interface temperature were set at 280°C, and the MS mode was electronic ionization (-70 eV) with the mass range m/z scanned between 35 and 250.

All samples were analysed in a single GC run. A set of external standards (standard mix) was also run at the start and the end of the sample set to ensure that both the HS-SPME extraction and MS detection were within specifications. The external standard contained dimethyl-sulfide (DMS), benzaldehyde, cyclohexanone, butyl

acetate, acetone, and ethanol in water in concentrations of 10 µg/g. Blanks (empty vials) were injected regularly to ensure no carry over occurred.

4.3.4 Method B: HS-Trap GC-MS

Method B was reported in the previous study (Stefanovic et al., 2017). A 2.5 mL of each sample was placed in a sealed vial. HS-Trap gas chromatography was performed using a Clarus 680 gas chromatograph coupled with Clarus 600T quadrupole mass spectrometer (Perkin Elmer, Courtaboeuf, France) as previously described (Pogacic et al., 2015). Briefly, the samples were warmed for 15 min to 65°C and volatiles adsorbed on a Tenax® trap at 35°C. The trap load was performed twice for each vial trap. Volatiles were separated on a Stalbilwax® MS capillary column (30 m × 0.25 mm × 0.25 µm; Restek, Bellefonte, PA, USA), with helium as the mobile phase. The temperature of the oven was initially 35°C, maintained for 10 min then increased at 5°C/min up to 230°C. The mass spectrometer was operated in the scan mode (scan time 0.3 s, interscan delay 0.03 s) within a mass range of m/z 29-206. Ionization was performed by electronic impact at -70 eV. All samples were run in the same GC run, with external standards (previously described (Pogacic et al., 2015)) and blank samples (boiled deionised water) injected regularly to confirm the absence of carryover.

4.3.5 Data processing, identification of volatile compounds and statistical analysis

The chromatographic data were processed by converting raw data to time- and mass-aligned chromatographic peak areas using the open source XCMS package implemented with the R statistical software (Smith et al., 2006). The signal presented pair of mass fragment and retention time (RT) on which this fragment occurred.

Analysis of the volatiles was based on abundance (peak area only). The mean coefficient of variation based on the analysis of external standards injected during GC runs was 7 % for Method A and 17 % for Method B. Compounds were identified based on mass spectra and linear retention indices (LRI), or tentatively identified on the basis of mass spectral data using NIST 2008 Mass Spectral Library (Scientific Instrument Services, Ringoes, NJ, USA). The compounds considered as main flavour contributors in cheese were selected according to previously published report (Curioni and Bosset, 2002). Some signals present in XCMS datasets could not be attributed to any compound because the percentage of identification was considered low (<50 %). The abundance of one mass fragment per identified compound was retained for further analysis if it fulfilled one or both of these conditions: the mass fragment possessed the highest abundance of all fragments present at the specific RT and/or the mass fragment was common in both datasets from Methods A and B for a given compound.

Statistics were performed using R statistical software (www.r-project.org). Analysis of variance (ANOVA) was performed on abundances to determine if they significantly differed ($p<0.05$) between cultures. Means were compared using the least significant difference test (LSD). Principal component analysis (PCA) was performed for each data set on the abundance of selected compounds that showed significant differences in ANOVA.

4.4 Results

4.4.1 The analysed methods differ in linearity ranges, LOD and LOQ

Table 1 presents the values of LOD and LOQ, along with the linearity range for each component of the standards analysed using each method. In general, linearity ranges were higher in Method A. Both methods showed a large linear dynamic range for most of the tested compounds, and gave similar results for seven of the ten compounds. Some differences were observed for 2,3-butanedione and ethyl acetate, which were better detected using Method A (LOD 10 to 100 times lower), and 3-methyl-butan-1-ol, which had a high LOD using both methods, but especially in Method B (Table 1).

4.4.2 Method A and Method B generated different volatile profiles from the same set of samples

A total of 1788 and 2073 signals were extracted from the XCMS analysis of chromatograms obtained from the cheese model system samples (cultures of inoculated strains and un-inoculated control) using Method A and B, respectively. In the chromatograms obtained for both methods, the peak of butan-1-ol covered extended regions. In chromatograms generated by Method A, the peak covered region between 5.5 and 6.5 min, while in chromatograms generated by Method B, the peak stretched between 14 and 16 min. A total of 94 compounds considered as potential flavour contributors were identified in both datasets, 70 and 66 of which were in Methods A and B, respectively, with 42 compounds common to both datasets. Moreover, 51 and 25 substances could not be identified based on the signals detected in Methods A and B, respectively, due to the low percentage of identification with the NIST database. Some compounds were identified but not

retained for further analysis, such as air pollutants or contaminants, or polysiloxane compounds originating from the GC column (16 and 42 in Methods A, and B, respectively).

Tables 2 and 3 present the common and the specific compounds identified in each dataset, along with their experimental linear retention index (LRI), molecular weight (MW) and boiling point (bp). Compounds of all chemical classes were detected by both methods (Table 2). However, more long-chain acids (MW>150 g/mol, bp>250°C), alcohols (MW>140 g/mol, bp>210°C), ketones (MW>130 g/mol, bp>170°C), and esters (MW>180 g/mol, bp>170°C) were identified using Method A, while more pyrazine derivatives and aldehydes were detected using Method B. Method A was also more efficient in detection of low boiling point sulfur compounds (MW<80 g/mol, bp<80°C), i.e. dimethyl-sulfide (DMS) and carbon-disulfide (CDS), whereas Method B was better in detecting low bp aldehydes (MW<80 g/mol, bp<100°C) and ketones (MW<100 g/mol, bp<100°C) (Table 3).

4.4.3 Quantitative comparison of detected compounds in Methods A and B

The results of ANOVA for each compound of both datasets showed that in total, 53 compounds significantly varied in concentration between the samples (Tables 2 and 3). Among the common compounds, 16 were detected as significantly different (SD) in samples of both datasets ($p<0.05$), seven detected as SD only in dataset A and three detected as SD only in dataset B (Table 2). The range of variation (i.e. the ratio between minimal and maximal concentrations in the set of samples studied, including the control) is indicated in the ANOVA column in Table 2. The range of variation of the 16 common compounds detected as SD in both datasets showed in general similar values, and the highest variations of abundance were observed in the

case of dimethyl-tetrasulfide for both methods (>300) (Table 2). The ten common compounds detected as SD in only one of the two datasets were the ones with the lowest ranges of variation (<8). For the 11 common compounds detected as SD in both datasets with a range of variation >10 between samples, the coefficient of correlation (r) between the values of abundance observed using both methods was calculated. The reason for this approach was to avoid the compounds where the variation in abundance was low, as it would not be useful in quantitative comparison of methods. All correlations were statistically significant, with r ranging from 0.44 for dimethyl-trisulfide (DMTS) to 0.95 for butyl butanoate. The stronger correlations ($r>0.7$) were observed in general for alcohols and ketones, whereas weaker correlations ($0.4<r<0.7$) were observed in case of sulfur compounds and the majority of esters (Table 2). In regard to specific compounds, i.e. detected only in one dataset, 16 and 11 compounds were detected as SD ($p<0.05$) in datasets A and B, respectively (Table 3).

Both methods of analysis were compared based on the relative abundance of chemical groups by sum of peaks identified in all samples (Fig. 1). Alcohols were the most abundant compounds in the volatile profiles, with the volatile fractions consisting of 70 % and 55 % of alcohols for Methods A and B, respectively. However, methods differed in volatile fractions of aldehydes (5 % and 16 % in Methods A and B, respectively) and pyrazines (0 % and 4 % in Methods A and B, respectively), while fractions of esters, acids, ketones and sulfur compounds were similar in both methods. The fractions of nitrogen compounds (benzonitrile, indole) were negligibly low in both datasets and were not presented in Figure 1.

4.4.4 Global variations among the samples differed after analysis by the two methods

A PCA was performed for each dataset using the compounds detected as SD ($p<0.05$) as variables (39 and 30 for datasets A and B respectively). On the plot built from dataset A, PC1 and PC2 described 37 % and 19 % of the total variation, respectively (Fig. 2a). PC1 separated the control from all the other samples, with the control associated with nonanal, dimethyl-tetrasulfide, 2-butyl-3,5-dimethyl-pyrazine and 6,10-dimethyl-undecan-2-one. Three strains were positioned separately from others: DPC3990, DPC4206 and, to a lesser extent, DPC2071. Strain DPC4206 was associated with undecan-2-ol, nonan-2-ol, butanoic acid, and butan-2-one, 3-methyl-butan-1-ol, DMDS, DMS, 5-decen-1-ol acetate, 2-phenylethanol, butyl butanoate, 2-ethyl-2-butenal. The position of strain DPC3990 was associated with nonan-2-one, undecan-2-one, tridecan-2-one, benzaldehyde, methional, pentadecan-2-one, 3-methyl-hexan-1-ol, butyl acetate, decanoic acid, benzeneacetaldehyde, benzeneacetic acid butyl ester. The position of strain DPC2071 was associated with butan-1-ol, 3-methyl-butan-1-ol, benzaldehyde, octan-1-ol, butyl hexanoate, butyl octanoate and butyl decanoate (Fig. 2a, confirmed by the results of LSD tests, not shown).

On the plot related to Method B, PC1 and PC2 described 42 % and 22 % of variation, respectively (Fig. 2b). PC1 separated the control from all the other samples. The control was negatively correlated with PC1 and mainly associated with higher amounts of DMTS and dimethyl-tetrasulfide. Three strains, DPC2071, DPC3990, and DPC4206, showed higher scores on PC1 compared to the other cultures and the control. Strain DPC2071 was associated with 3-methyl-butan-1-ol, 3-hydroxy-butan-2-one, acetic acid, octan-1-ol, and butyl decanoate, while strain DPC3990 was associated with hexan-1-ol, nonan-2-one, 3-methyl-hexan-1-ol, acetic

acid, butanedioic acid dimethyl ester, undecan-2-one, and benzeneacetic acid butyl ester. Strain DPC4206 was also positively correlated with PC2, and associated with high amounts of butanoic acid, butyl butanoate, S-methyl-thioacetate, butanal, undecan-2-ol, acetic acid, 3-methyl-butanoic acid, hexanoic acid, 2-phenylethanol, and 2-ethyl-2-hexenal (Fig. 2b, confirmed by the results of LSD tests, not shown).

4.5 Discussion

The most important volatile odorant compounds that contribute to cheese aroma belong to several chemical classes of compounds, including alcohols, aldehydes, ketones, esters, acids, sulfur compounds, and compounds with an aromatic ring (Yvon, 2006, Ardö and Varming, 2010). However, not all compounds have the same importance towards the final aroma (Curioni and Bosset, 2002, Dunkel et al., 2014), because their concentrations and their perception thresholds vary markedly. In this study, we contrasted two diverse approaches for the detection of volatiles generated in a cheese model system by *L. paracasei* strains. The results were obtained in the present study (Method A) and in a previous study (Stefanovic et al., 2017) (Method B).

Both methods were compared firstly based on their linearity ranges and values of LOD and LOQ for a set of standards. Factors that influence LOD and LOQ values are the nature of sample, the equilibration conditions, and the type of the extraction adsorbent. In the comparison of HS-SPME and HS-Trap based methods for analysis of rapeseed off-flavours, HS-Trap had higher values of LOD and LOQ, but within the same order of magnitude as HS-SPME (Sghaier et al., 2016). In agreement with this, the results from the present study illustrated similar sensitivities of the evaluated methods, even if they differed for the extraction of certain low boiling point volatiles, such as 2,3-butanedione and ethyl acetate.

The two methods detected common compounds (present in both datasets) and specific compounds (present in only one of the two datasets). Important differences in detection of common compounds were observed. For instance, Method A was able to detect higher number of common compounds in SD concentrations. The

calculated coefficients of correlation (Table 3) showed that compounds such as ketones and alcohols were analysed in the same manner by both methods, whereas the results obtained for sulfur compounds were poorly correlated. This is due to the high variability in detection of the latter, as shown by the high CV% observed for some compounds, especially dimethyl-tetrasulfide (Table 2). However, it is worth to mention that in general, sulfur compounds are highly reactive and therefore difficult to quantify by GC-MS, and pulsed-flame photometric detection has been suggested as more suitable for quantification of these compounds (Burbank and Qian, 2005, Heroult et al., 2008).

The differences observed in the detection of common compounds can result from several factors. The first of them is the affinity of the sorbent for the adsorption/absorption of the compound. Besides this, the surface of sorbent available for the adsorption can impact on the sensitivity of the method. The smaller surface of the SPME fibre could limit the extraction of some compounds with a lower affinity, which are better extracted by HS-Trap, with a bigger adsorption area. A known disadvantage of SPME is volatile displacement during exposure to the headspace where very volatile compounds with a lower affinity to the fibre get displaced by less volatile compounds with a greater affinity for the fibre over time or until headspace equilibrium is reached (Mondello et al., 2005).

The presence of specific compounds in the two datasets results mainly from different specificities of the sorbents used for certain groups of compounds, and/or from the competition for the binding sites. The specificity of CAR/DVB/PDMS fibre was higher for extraction of different acids, especially those of higher MW, while Tenax® sorbent did not extract any specific acids. This result is consistent to previous studies, which compared fingerprints of volatiles in three cheese varieties (Mallia et

al., 2005) and butter (Povolo and Contarini, 2003), where the same SPME fibre as used in this study showed higher selectivity towards ketones and acids, including those with higher molecular mass. Conversely, DHS with Tenax® sorbent was more effective in extraction of alcohols and aldehydes. The specificity of Tenax® sorbent in the present study was to extract more pyrazine derivatives, as well as some specific sulfur compounds (acetylthiazol, thiophene). Similar findings were reported in a study on goat meat, where Tenax® extracted pyrazines, pyrroles, pyridines, and DMS in higher concentrations compared to SPME with CAR/PDMS, and also extracted specific sulfur compounds such as thiophenes, alicyclic sulfides, and thiazoles (Madruga et al., 2009).

Apart from sorbent characteristics, MW is recognised as another factor influencing the volatiles extraction (Povolo and Contarini, 2003). In the present study, two techniques showed different efficacies in extracting molecules based on molecular mass and boiling points, and in general, Method A demonstrated better detection of compounds with higher MW and bp, while Method B detected better compounds with lower MW and bp. This corresponds to the results of the study on butter samples where the same SPME sorbent as in the present study (DVB/CAR/PDMS) was more efficient in the extraction of high MW molecules, while Tenax® preferably extracted compounds with lower MW (Povolo and Contarini, 2003). This was due to the porous characteristics of DVB component of SPME fibre, which is known to capture larger and less volatile compounds (Mondello et al., 2005).

The characteristics of the GC column influence the efficient separation of compounds depending on their polarity and affinity towards the column, sometimes resulting in coeluted compounds and difficulty for quantification and/or even accurate identification. In this study, different columns were used for GC separation

and the type of column used may also have influenced the nature of the compounds identified in the two datasets. In Method A, the non-polar column facilitated a better separation of the less polar compounds that were possibly extracted by Tenax®, but due to the column characteristics were not detected in Method B. This could possibly be the case for long acids, such as nonanoic and decanoic, which were detected in Method A, but not in Method B. Conversely, in Method B, a polar column was used and some low MW aldehydes (butanal, 3-methyl-butanal), phenol and pyrazine derivatives were detected, and although they might have been extracted by the SPME fibre, due to the column characteristics they would not be detected in Method A. In a study on virgin olive oil volatiles, SPME was coupled to GC with different columns, and it was shown that when a non-polar column was used, some polar compounds (propanoic acid, pentanoic acid) were not detected, but their presence in samples was confirmed when a polar column was used (Vichi et al., 2003). However, both types of columns are equally used in aroma volatile analysis and neither of them has a specific characteristic that makes it superior to the other type. In spite of the differences potentially caused by the column selection, a study of bacterial volatiles showed that the main factor contributing to variations was the selection of the extraction method and not the column (Tait et al., 2014).

To analyse the detection of compounds by chemical classes, sums of peak areas detected in each method were compared (Fig. 1). This result is consistent with the specific compounds reported in Table 3, where higher numbers of specific alcohols were detected by Method A, while higher numbers of specific aldehydes were detected in Method B. A similar conclusion could be made for pyrazines, which more expansively detected by Method B. In the case of sulfur compounds and esters,

the relative proportions of compounds were similar, but qualitatively they differed, confirming the importance of the analysis conditions.

Depending on the study and the experimental design, the aim of volatile analysis is often to differentiate the samples within a given set. In a previous comparative study, both SPME and DHS based methods resulted in a perfect differentiation of cheese samples belonging to three varieties (Mallia et al., 2005), and the same trends of distribution of butter samples were observed (Povolo and Contarini, 2003). However, contrasting results were obtained in a study that compared three DHS methods performed on the samples of same cheeses (Barron et al., 2005), where significantly different volatile profiles were obtained by the three methods. In the present study, the discrimination of samples by two methods differed in relation to the most discriminated strain (DPC3990 for Method A and DPC4206 for Method B), but both methods identified the same samples (cultures of strains DPC4206, DPC3990, and, to a lesser extent, culture of strain DPC2071) as the most diverse from all the other cultures, and also clearly differentiated the un-inoculated control. The differences between both methods relate to the variation in the presence of compounds that determine cultures positions, some of which are specific for the method itself or are present in SD abundances only in one dataset. These differences arise mainly due to the selectivity of the sorbent and to a lesser extent due to the nature of the extraction and the polarity of the column, as discussed above.

4.6 Conclusion

Different volatile fingerprints of the same samples were obtained by HS-SPME and HS-Trap methods. One of the factors contributing to the differences was the extraction step due to the type of sorbent, its surface availability and affinity towards specific aroma classes. Additionally, column polarity also contributed to differences, however apparently to a lesser extent. Chemical family, molecular weight, and boiling points defined the molecules detected by both methods. The level of variation amongst samples differed in both methods, and different samples were detected as having the most diverse volatile profiles. This is an important finding and suggests the importance of the extraction and GC method conditions.

4.7 Acknowledgments

Ewelina Stefanovic is in a receipt of Teagasc Walsh Fellowship. This collaboration was supported by Short Term Overseas Training Award 2015, funded by Teagasc.

4.8 References

Aberl, A. & Coelhan, M. 2012. Determination of volatile compounds in different hop varieties by headspace-trap GC/MS - in comparison with conventional hop essential oil analysis. *J Agric Food Chem,* 60**,** 2785-2792.

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

Barani, F., Dell'Amico, N., Griffone, L., Santoro, M. & Tarabella, C. 2006. Determination of volatile organic compounds by headspace trap. *J Chromatogr Sci*, 44**,** 625-630.

Barron, L. J. R., Redondo, Y., Ortigosa, M. A., Carbonell, M. A., Garcia, S., Torre, P., de Renobales, M. & Fernandez-Garcia, E. 2005. Comparison of dynamic headspace methods used for the analysis of the volatile composition of Spanish PDO ewe's raw milk cheeses. *Lait,* 85**,** 491-513.

Bosse Nee Danz, R., Wirth, M., Konstanz, A., Becker, T., Weiss, J. & Gibis, M. 2017. Determination of volatile marker compounds in raw ham using headspace-trap gas chromatography. *Food Chem,* 219**,** 249-259.

Bosset, J.-O., Gauch, R., Mariaca, R. & Klein, B. 1995. Comparison of various sample treatments for the analysis of volatile compounds by GC-MS: Application to the Swiss Emmental cheese. *Mitt Gebiete Lebensm Hyg*, 86**,** 672-698.

Burbank, H. M. & Qian, M. C. 2005. Volatile sulfur compounds in Cheddar cheese determined by headspace solid-phase microextraction and gas chromatograph-pulsed flame photometric detection. *J Chromatogr A,* 1066**,** 149-157.

Contarini, G. & Povolo, M. 2002. Volatile fraction of milk: Comparison between purge and trap and solid phase microextraction techniques. *J Agric Food Chem,* 50**,** 7350-7355.

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.

Dunkel, A., Steinhaus, M., Kotthoff, M., Nowak, B., Krautwurst, D., Schieberle, P. & Hofmann, T. 2014. Nature's chemical signatures in human olfaction: a foodborne perspective for future biotechnology. *Angew Chem Int Ed Engl*, 53**,** 7124-7143.

Elmore, J. S., Erbahadir, M. A. & Mottram, D. S. 1997. Comparison of dynamic headspace concentration on Tenax with solid phase microextraction for the analysis of aroma volatiles. *J Agr Food Chem*, 45**,** 2638-2641.

Feng, Y., Su, G., Sun-Waterhouse, D., Cai, Y., Zhao, H., Cui, C. & Zhao, M. 2016. Optimization of Headspace Solid-Phase Micro-extraction (HS-SPME) for Analyzing Soy Sauce Aroma Compounds via Coupling with Direct GC-Olfactometry (D-GC-O) and Gas Chromatography-Mass Spectrometry (GC-MS). *Food Anal Methods*, 10**,** 713-726

Heroult, J., Bueno, M., Potin-Gautier, M. & Lespes, G. 2008. Organotin speciation in French brandies and wines by solid-phase microextraction and gas chromatography-pulsed flame photometric detection. *J Chromatogr A*, 1180**,** 122-130.

Idris, S. A., Robertson, C., Morris, M. A. & Gibson, L. T. 2010. A comparative study of selected sorbents for sampling of aromatic VOCs from indoor air. *Anal Methods*, 2**,** 1803.

Jelen, H. H., Majcher, M. & Dziadas, M. 2012. Microextraction techniques in the analysis of food flavor compounds: A review. *Anal Chim Acta*, 738, 13-26.

Kilcawley, K. N. 2017. Cheese flavour. *In*: Fox, P.F., Guinee, T.P., Cogan, T.M., McSweeney, P.L.H., (eds.) *Fundamentals of cheese science*. Springer.

Laing, D. G. & Jinks, A. 1996. Flavour perception mechanisms. *Trends Food Sci Technol*, 7**,** 387-389.

Lawlor, J. B., Delahunty, C. M., Wilkinson, M. G. & Sheehan, J. 2002. Relationships between the gross, non-volatile and volatile compositions and the sensory attributes of eight hard-type cheeses. *Int Dairy J***,** 12, 493-509.

Lehotay, S. J. & Hajšlová, J. 2002. Application of gas chromatography in food analysis. *Trends Anal Chem*, 21**,** 686-697.

Liu, Y., Xu, X. L. & Zhou, G. H. 2007. Comparative study of volatile compounds in traditional Chinese Nanjing marinated duck by different extraction techniques. *Int J Food Sci Tech*, 42**,** 543-550.

Madruga, M. S., Elmore, J. S., Dodson, A. T. & Mottram, D. S. 2009. Volatile flavour profile of goat meat extracted by three widely used techniques. *Food Chem*, 115**,** 1081-1087.

Mallia, S., Fernández-García, E. & Olivier Bosset, J. 2005. Comparison of purge and trap and solid phase microextraction techniques for studying the volatile aroma compounds of three European PDO hard cheeses. *Int Dairy J***,** 15, 741-758.

Marilley, L. & Casey, M. G. 2004. Flavours of cheese products: metabolic pathways, analytical tools and identification of producing strains. *Int J Food Microbiol*, 90**,** 139-59.

Marsili, R. 2011. Flavors and off-flavors in dairy foods. *In*: Fuquay, J. W. (ed.) *Encyclopedia of Dairy Sciences* (Second Edition). Academic Press.

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

Mondello, L., Costa, R., Tranchida, P. Q., Chiofalo, B., Zumbo, A., Dugo, P. & Dugo, G. 2005. Determination of flavor components in Sicilian goat cheese by automated HS-SPME-GC. *Flavour Frag J*, 20**,** 659-665.

Murat, C., Gourrat, K., Jerosch, H. & Cayot, N. 2012. Analytical comparison and sensory representativity of SAFE, SPME, and Purge and Trap extracts of volatile compounds from pea flour. *Food Chem*, 135**,** 913-920.

Nikfardjam, M. P. & Maier, D. 2011. Development of a headspace trap HRGC/MS method for the assessment of the relevance of certain aroma compounds on the sensorial characteristics of commercial apple juice. *Food Chem*, 126**,** 1926-1933.

Plutowska, B. & Wardencki, W. 2007. Aromagrams-Aromatic profiles in the appreciation of food quality. *Food Chem*, 101**,** 845-872.

Pogacic, T., Maillard, M. B., Leclerc, A., Herve, C., Chuat, V., Yee, A. L., Valence, F. & Thierry, A. 2015. A methodological approach to screen diverse cheese-related bacteria for their ability to produce aroma compounds. *Food Microbiol*, 46**,** 145-153.

Povolo, M. & Contarini, G. 2003. Comparison of solid-phase microextraction and purge-and-trap methods for the analysis of the volatile fraction of butter. *J Chromatogr A*, 985**,** 117-125.

Schulz, K., Dreßler, J., Sohnius, E.-M. & Lachenmeier, D. W. 2007. Determination of volatile constituents in spirits using headspace trap technology. *J Chromatogr A*, 1145**,** 204-209.

Sghaier, L., Vial, J., Sassiat, P., Thiebaut, D., Watiez, M., Breton, S., Rutledge, D. N. & Cordella, C. B. Y. 2016. Analysis of target volatile compounds related to fishy off-flavor in heated rapeseed oil: A comparative study of different headspace techniques. *Eur J Lipid Sci Technol*, 118**,** 906-918.

Smith, C. A., Want, E. J., O'Maille, G., Abagyan, R. & Siuzdak, G. 2006. XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Anal Chem*, 78**,** 779-787.

Snow, N. H. & Bullock, G. P. 2010. Novel techniques for enhancing sensitivity in static headspace extraction-gas chromatography. *J Chromatogr A*, 1217**,** 2726-2735.

Snow, N. H. & Slack, G. C. 2002. Head-space analysis in modern gas chromatography. *Trends Anal Chem*, 21**,** 608-617.

Soria, A. C., Garcia-Sarrio, M. J. & Sanz, M. L. 2015. Volatile sampling by headspace techniques. *Trends Anal Chem*, 71, 85-99.

Stefanovic, E., Thierry, A., Maillard, M. B., Bertuzzi, A., Rea, M. C., Fitzgerald, G. F., McAuliffe, O. & Kilcawley, K. N. 2017. Strains of the *Lactobacillus casei* group show diverse abilities for the production of flavor compounds in 2 model systems. *J Dairy Sci*, 100**,** 6918-6929.

Tait, E., Perry, J. D., Stanforth, S. P. & Dean, J. R. 2014. Identification of volatile organic compounds produced by bacteria using HS-SPME-GC-MS. *J Chromatogr Sci,* 52**,** 363-373.

Thomsen, M., Gourrat, K., Thomas-Danguin, T. & Guichard, E. 2014. Multivariate approach to reveal relationships between sensory perception of cheeses and aroma profile obtained with different extraction methods. *Food Res Int,* 62**,** 561-571.

Tunick, M. H. 2014. Analyzing volatile compounds in dairy products. *J Sci Food Agric,* 94**,** 1701-1705.

Vichi, S., Castellote, A. I., Pizzale, L., Conte, L. S., Buxaderas, S. & Lopez-Tamames, E. 2003. Analysis of virgin olive oil volatile compounds by headspace solid-phase microextraction coupled to gas chromatography with mass spectrometric and flame ionization detection. *J Chromatogr A,* 983**,** 19-33.

Yvon, M. 2006. Key enzymes for flavour formation by lactic acid bacteria. *Aust J Dairy Technol,* 61**,** 88-96.

Table 1: Linearity ranges, limits of detection (LOD), limits of quantification (LOQ), and coefficient of determination (r^2) of the analysis of aqueous solutions of mixed standard compounds analysed by Method A (HS-SPME)- this study, and Method B (HS-Trap)- Pogacic et al., 2015.

Compound	MW (g/mol)	Bp (°C)	Mass fragment of quantification		Linearity range (ng/g)		r^2		LOD (ng/g)		LOQ (ng/g)	
			A	B	A	B	A	B	A	B	A	B
2,3-butanedione	86	88	86	86	0-1600	0-1000	0.872	0.995	506	5	1517	16
Ethyl acetate	88	77	43	88	0-800	0-900	0.983	0.990	62	5	186	16
Ethyl propanoate	102	99	57	102	0-1520	0-1000	0.953	0.974	7	3	20	10
Ethyl butanoate	116	122	43	116	0-760	0-900	0.977	0.924	9	3	28	9
Ethyl hexanoate	144	168	99	99	0-1450	0-160	0.992	0.933	5	5	15	17
3-methyl-butanal	86	92	58	58	0-960	0-350	0.984	0.942	16	4	50	12
Benzaldehyde	106	179	105	105	0-1800	0-1000	0.997	0.967	1.2	4	3.6	12
Heptan-2-one	114	151	58	58	0-100	0-100	0.988	0.960	0.5	8	1.4	26
Dimethyl-disulfide	94	110	94	94	0-850	0-250	0.986	0.890	2	2	7	7
3-methyl-butan-1-ol	88	131	55	70	0-50000	0-4000	0.984	0.962	88	586	263	1758

MW-molecular weight, Bp- boiling point.

Table 2: Common compounds identified by Methods A and B.

Chemical group	Compound	LRI Method A	LRI Method B	MW (g/mol)	Bp (°C)	ANOVA Method A	ANOVA Method B	r	r^2	CV% Method A	CV% Method B
aldehydes	Benzaldehyde	1000	1517	106	179	7.4					
	2-Phenylacetaldehyde(Benzeneacetaldehyde)	1089	1642	120	195	1.9					
	Nonanal	1149	1388	142	191	4.4	7.6				
ketones	Propan-2-one (Acetone)	-	-	58	56						
	Butan-2-one	-	896	72	80	3.8					
	Heptan-2one	915	1180	114	151						
	1-Phenylethanone (Acetophenone)	1114	1646	120	202						
	Nonan-2-one	1133	1383	142	195	16.3	6.0				
	Undecan-2-one	1357	1594	170	232	35.6	19.7	0.935	0.8744	25	28
	Tridecan-2-one	1497	1807	198	263	74.7	17.9	0.934	0.8723	19	33
alcohols	Ethanol	-	912	46	78						
	Butan-1-ol	675	1169	74	118	1.4					
	3-Methyl-butan-1-ol	751	1216	88	132	339.6	205.6	0.833	0.4784	13	17
	2-Furanmethanol	883	1673	98	171						
	Phenyl-methanol (Benzyl-alcohol)	1076	-	108	205						
	3-Methyl-hexan-1-ol	962	1357	116	143	97.0	44.2	0.907	0.8226	18	24
	2-Phenylethanol (phenyl-ethyl-alcohol)	1165	-	122	218	92.2	11.0	0.704	0.4966	18	30
	Octan-1-ol	1109	1583	130	195	19.8	17.1	0.637	0.4060	27	48
	2-Ethyl-hexan-1-ol	1064	-	130	184						
	Undecan-2-ol	1367	1722	172	233	7.2	19.3				
	2,4-Di-tert-butylphenol	1507	-	206	265						
esters	Butyl acetate	832	1094	116	126	71.5	53.4	0.488	0.2385	23	28
	Butyl butanoate	1027	1273	144	167	27.6	15.2	0.955	0.9122	41	39
	Butyl hexanoate	1203	1399	172	208	7.4					
	Benzeneacetic acid butyl ester	1465	1915	192	258	43.3	9.4				
	Butyl octanoate	1433	1602	200	240						
	Butyl decanoate	1585	-	228	274	10.8	10.3	0.551	0.3043	21	36
acids	Acetic acid	-	1450	60	118		7.3				
	Butanoic acid	808	1627	88	164	6.8	5.6				
	3-Methyl-butanoic acid	858	1669	102	177		2.6				
	Hexanoic acid	1000	-	116	205		2.2				

	Octanoic acid	1208	-	144	239						
S compounds	Dimethyl-disulfide (DMDS)	765	1085	94	110	7.7					
	Dimethyl-trisulfide (DMTS)	1011	1366	126	170	25.6	43.4	0.436	0.1904	30	38
	Methyl-sec-butyl-disulfide	1036	1353	136	177						
	Dimethyl-tetrasulfide	1294	1759	158	243	8059.7	321.4	0.635	0.4044	71	56
pyrazines	2,5-Dimethyl-pyrazine	943	1312	108	155						
	2-Ethyl-5-methyl-pyrazine	1041	1382	122	169						
	3-Ethyl-2,5-dimethyl-pyrazine	1120	1436	136	180						
	2-Methyl-3-isopropyl-pyrazine	1097	1391	136	188						
N compounds	Benzonitrile	1024	1601	103	190						
	Indole	1371	-	117	254	6.8					

LRI- linear retention index, MW- molecular weight; Bp- boiling point, CV%- coefficient of variation.

ANOVA- for signals where significant differences ($p<0.05$) were observed after performing the least significant differences test (LSD), fold between the highest and the lowest value is indicated. For compounds with no significant difference after ANOVA, blank fields are left.

Correlations coefficients (r), and determination coefficients (r^2) of extraction of common compounds are calculated for compounds that fulfilled two conditions: there was a significant difference in abundances of a compound between the samples in Method A and Method B datasets, and the folds between the highest and the lowest abundance were higher than 10 in both datasets.

Table 3: Specific compounds identified by Methods A and B.

	Method A	LRI	MW (g/mol)	Bp (°C)	ANOVA	Method B	LRI	MW (g/mol)	Bp (°C)	ANOVA
aldehydes	2-ethyl-2-butenal	839	98	136	217.4	Butanal	867	72	75	17.8
						3-Methyl-butanal	909	86	92	
						2-Ethyl-2-hexenal	1291	126	175	111.5
						4-Propylbenzaldehyde		148	240	
ketones	Octan-2-one	1022	130	173		3-Hydroxy-butan-2-one (Acetoin)	1278	88	148	80.3
	Octan-3-one	1018	130	167		4-Methyl-pentan-2-one	1003	100	117	
	6,10-dimethyl-undecan-2-one	1245	198	245	5.6	1-Phenyl-propan-2-one	1727	134		
	Pentadecan-2-one	1814	226	293	4.8					
alcohols	Nonan-2-ol	1144	144	194	35.3	Phenol	-	94	182	
	Nonan-1-ol	1222	144	213	29.5	Hexan-1-ol	1358	102	157	38.3
	2-Ethyl-5-propylphenol	1231	164	290						
	2,6-dimethyl-7-octen-2-ol	1287	156	188						
	Tridecan-2-ol	1502	200	257	90.4					
esters	Azepan-2-one (Caprolactam)	1327	113	270		Ethyl hexanoate	1227	144	167	11.6
	Butyl propanoate	932	130	146		Butanedioic acid dimethyl ester	1591	146	195	2.4
	2-Methylpropanoic acid, butyl ester	980	144	157	4.1	Benzoic acid 1-methylpropyl ester	-	178	280	
	3-Methylbutanoic acid, butyl ester ester	1083	186	175	108.2					
	3-Decen-1-ol-acetate-(Z)	1484	198	256	7.1					
	5-Decen-1-ol-acetate-(E)	1492	198	210	10.8					
	Hexyl-hexanoate	1242	200	245						
acids	2-Methyl-butanoic acid	870	102	177						
	4-Hydoxy-butanoic acid	948	104	180						
	Nonanoic acid	1315	158	254						
	Decanoic acid	1414	172	269	5.2					
S compounds	Dimethyl-sulfide (DMS)	-	62	37	8.8	1,3-thiazole	1251	85	117	
	Carbon-disulfide	-	76	46	5.5	S-methyl-thioacetate	1055	90	96	365.6
	3-methylsulfanylpropanal (Methional)	936	104	165	20.5	3-(Methylthio)-propan-1-ol	1724	106	195	18.4
	2-Thiophenemethanamine	1232	113	195	6.3	2-Acetylthiazole	1642	127	89	3.2
	Benzyl-methyl-disulfide	1448	170	85		3-Phenyl-thiophene	-	160	229	
nes	2-Butyl-3,5-dimethyl-pyrazine	1383	164	230	7.4	2-Methyl-pyrazine	1158	94	136	
	2,3-Dimethyl-5-(1-methylpropyl)-pyrazine	1250	164	225		2-Isopropyl-pyrazine	1344	122	170	
						2,3,5-Trimethyl-pyrazine	1393	122	172	

3,5-Diethyl-2-methyl-pyrazine	1485	150	205	
2,3,5-Trimethyl-6-ethyl-pyrazine	1504	150	206	12.7
3,5-Dimethyl-2-(1-methylpropyl)-pyrazine	1646	164	224	2.8
2,5-Dimethyl-3-(3-methylbutyl)-pyrazine	1651	178	242	

LRI- linear retention index, MW- molecular weight, Bp- boiling point.

ANOVA- for signals where significant differences ($p<0.05$) were observed after performing the least significant differences test (LSD), fold between the highest and the lowest value is indicated. For compounds with no significant difference after ANOVA, blank fields are left.

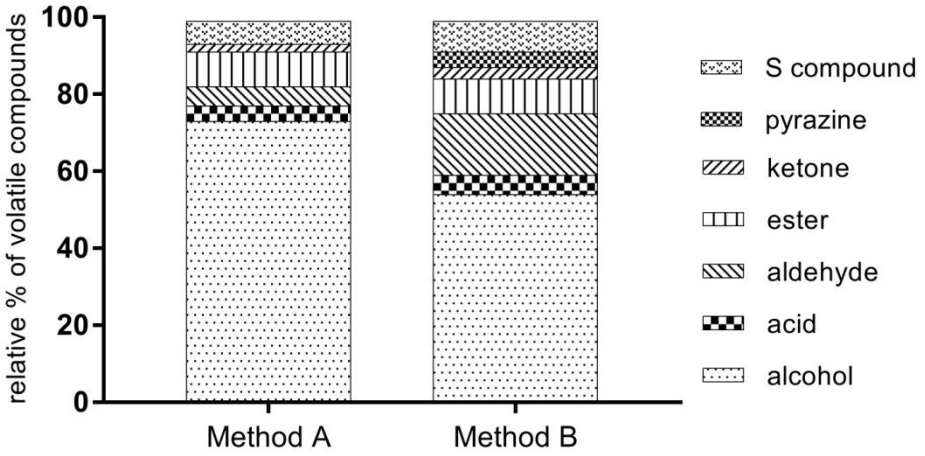
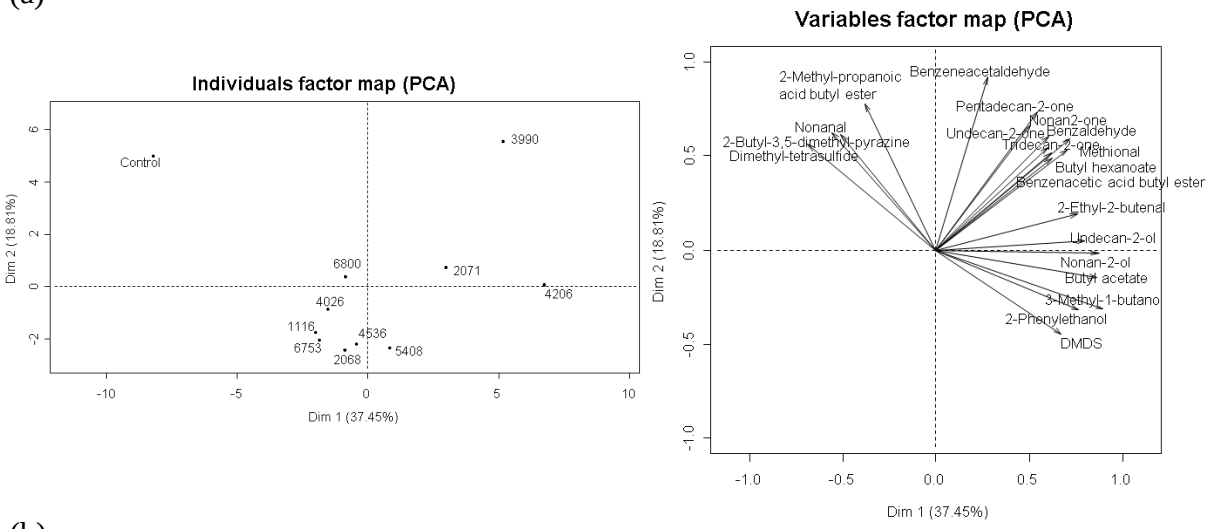


Figure 1: Comparison of volatile fractions of cheese model system inoculated with strains *Lactobacillus paracasei* obtained by Methods A and B. Bars represent the relative percentages of the sums of chromatographic peaks areas of the compounds identified in all samples (cultures and un-inoculated control), grouped by chemical class.

(a)



(b)

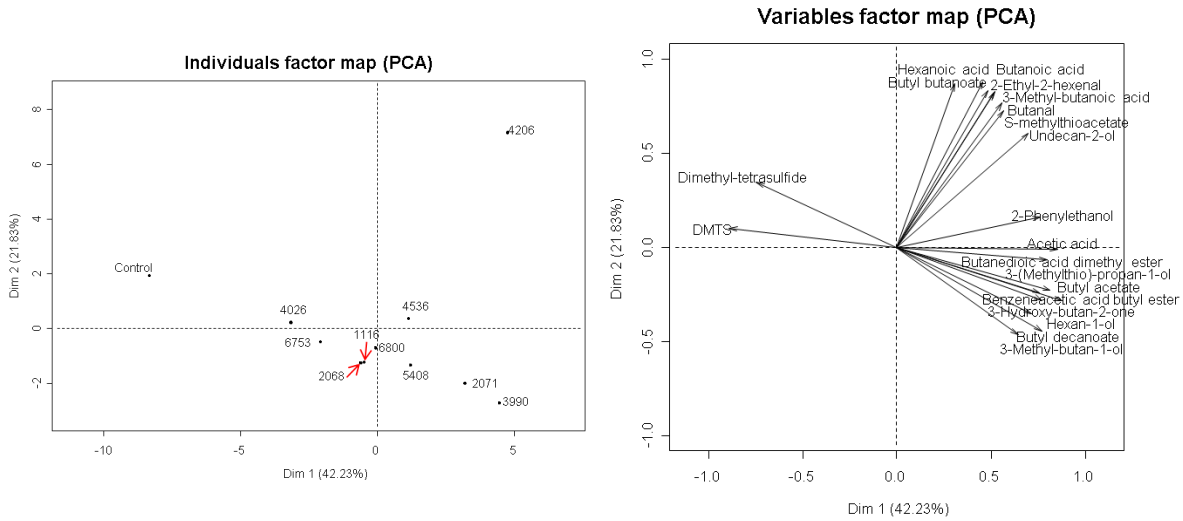


Figure 2: Individual factor map and variable factor map of principal component analysis (PCA) on volatile compounds detected in significantly different abundances ($p<0.05$) produced by ten strains of *Lactobacillus paracasei* in model system incubated for 24 h at 30 °C, analysed by Method A (a) and Method B (b). The control was an un-inoculated model system incubated under the same conditions. All samples and the control were tested in triplicate. The variables poorly represented in this plot (square cosinus limit below 0.6) are not shown and the DPC prefix has been removed from the strains name, to improve the clarity of the figure.

Chapter 5

Comparative genomic and metabolic analysis of three *Lactobacillus paracasei* cheese isolates reveals considerable genomic differences in strains from the same niche

5.1 Abstract

Strains of *Lactobacillus paracasei* are present in many diverse environments, including dairy and plant material and the intestinal tract of humans and animals. Their adaptation to various niches is correlated to intra-species diversity at the genomic and metabolic level. In this study, we elucidated and compared the genome sequences of three *L. paracasei* strains isolated from mature Cheddar cheeses, two of which (DPC4206 and DPC4536) shared the same genomic fingerprint by PFGE, but demonstrated varying metabolic capabilities. Genome sizes varied from 2.9 Mbp for DPC2071, to 3.09 Mbp for DPC4206 and 3.08 Mpb for DPC4536. The presence of plasmids was a distinguishing feature between the strains with strain DPC2071 possessing an unusually high number of plasmids (11), while DPC4206 had one plasmid and DPC4536 harboured no plasmids. Each of the strains possessed specific genes not present in the other two analysed strains. The three strains differed in their abundance of carbohydrate-converting genes, and in the types of carbohydrates that could be used as energy sources. Genes involved in the metabolism of carbohydrates not usually connected with the dairy niche, such as *myo*-inositol and pullulan were also detected, but strains did not utilise these sugars. The genetic content of the three strains also differed in regard to specific genes for arginine and sulfur-containing amino acid metabolism, genes contributing to resistance to heavy metal ions and oxidative stress, and genes involved in regulation of metabolic processes. In addition, variability in the presence of phage remnants and phage protection systems was evident. These findings confirm a considerable level of heterogeneity of *Lactobacillus paracasei* strains, even between strains isolated from the same niche.

5.2 Introduction

The genus *Lactobacillus* consists of more than 200 species and subspecies (Sun et al., 2015) present in various environments such as plants, fermented food products (dairy, meat, wine), and both the human and animal gastrointestinal and reproductive tracts (Pfeiler and Klaenhammer, 2007, Schroeter and Klaenhammer, 2009). One of the most studied groups of this genus is the *Lactobacillus casei* group, which includes the species *Lactobacillus casei*, *Lactobacillus paracasei* and *Lactobacillus rhamnosus*. Strains of this group show remarkable ecological adaptability and have been isolated from all the typical habitats of lactobacilli (Cai et al., 2009, Toh et al., 2013). Such a diverse range of sources facilitated a broad spectrum of applications of strains of this species in dairy production (adjunct cultures), health-related (probiotics, bacteriocins) and biotechnological fields. The characteristics and potential applications make the species of the *L. casei* group one of the best explored within the *Lactobacillus* genus.

To date, the genome sequences of 104 *L. casei* and *L. paracasei* strains are publicly available, 16 of which are complete genome sequences (www.ncbi.nlm.nih.gov, last accessed in May 2017). The comparative genomic analysis of *L. casei and L. paracasei* genomes has revealed that, as in other *Lactobacillales*, there is an evolutionary trend towards minimisation of genome size through the decay of genes coding for functions not required for strains inhabiting specific niches. This loss of redundant genes has been shown to be followed by the acquisition of genes by horizontal gene transfer (HGT) as a response to niche adaptation (Makarova et al., 2006). The rich pool of available genome sequences enables the definition of the gene sets that are common to all strains (core genome), the genes present only some

of the strains (dispensable genome), or genes that are unique for a single strain (unique genes). Insights into the common and unique genes enable correlation of gene variations among different strains to the presence or absence of phenotypic traits (Smokvina et al., 2013). The pangenome (or supragenome) comprises the union of all genes present within a selected genome set (species, genera or higher taxonomic group) (Medini et al., 2005). *L. casei* and *L. paracasei* pangenome studies have confirmed the wide range of ecological niches that can be inhabited by strains of the *L. casei* group (Broadbent et al., 2012, Smokvina et al., 2013, Toh et al., 2013), arising from the variability of genes supporting utilisation of numerous energy sources and other specific genes contributing to the efficient survival in habitats with differing environmental conditions.

The dairy niche represents a nutritionally rich habitat, and niche specialisation in dairy strains has led to substantial gene decay, which has limited their survival in more nutritionally scarce environments (Cai et al., 2009). As a consequence, genomes of dairy isolates possess a higher ratio of pseudogenes, compared to non-dairy isolates (O'Sullivan et al., 2009). Conversely, genomes of dairy specialists are abundant in sugar transportation, proteolytic and amino acid transportation-encoding genes that enable uptake of nutrients present in the dairy environment (Makarova et al., 2006). However, the isolation source does not necessarily correspond to the usual habitat of a strain, as strains can change their habitats due to their adaptability. This is evident from genome content, where often unusual genes that are not expected for strains of a certain isolation source are present, suggesting that a strain may have transferred from one niche to another (Ceapa et al., 2015).

The aim of this study was to compare the genomic and metabolic characteristics of three *L. paracasei* strains that were isolated from mature Cheddar cheese.

Previously, these strains were selected based on the activity of the key enzymes in flavour production and their volatile profiles in cheese model systems (Stefanovic et al., 2017b, Stefanovic et al., 2017c). Genomic fingerprinting established that two of the strains (DPC4206 and DPC4536) showed identical PFGE profiles, despite demonstrating considerable differences in selected enzyme activities, such as cell envelope proteinase, aminopeptidases, aromatic aminotransferase and glutamate dehydrogenase (Stefanovic et al., 2017b). Similarly, these two strains exhibited distinct differences when examined for the production of volatile flavour compounds in two cheese model systems (Stefanovic et al., 2017c). The third strain (DPC2071), which differed considerably in terms of PFGE profile, possessed high levels of activity of enzymes of the proteolytic system, especially cell envelope proteinase, and exhibited one of the most distinct volatile profiles in cheese model systems, as shown in the previous studies (Stefanovic et al., 2017b, Stefanovic et al., 2017c). It was proposed that elucidation and comparison of the genomes of these three strains would enable our understanding of the genetic basis of their different phenotypic characteristics.

5.3 Materials and methods

5.3.1 Bacterial strains for comparative analysis

The three *L. paracasei* strains examined in this study were isolated from the non-starter flora of Cheddar cheese, and deposited in the DPC Culture Collection. The genomes of all three strains are available from public databases (accession numbers: NCSN01000000, NCSO01000000 and NCSP01000000, for strains DPC2071, DPC4206 and DPC4536, respectively). The details of genome sequencing and assembly are reported in Stefanovic et al., 2017a. Contig mapping was performed using Mauve, with the genomes of *L. paracasei* ATCC 334, *L. casei* BDII and *L. casei* 12A as references for strains DPC2071, DPC4206 and DPC4536, respectively (Darling et al., 2004).

5.3.2 Identification of strain-specific genes in each of the input genomes

Whole genome comparisons were undertaken using BLAST Ring Image Generator (Alikhan et al., 2011), and progressiveMauve alignments (Darling et al., 2004), in order to identify unique genomic regions belonging to each of the strains.

Clustered regularly interspaced short palindromic repeat (CRISPR) regions in each genome were identified using an online tool CRISPRfinder (Grissa et al., 2007). Viable and cryptic prophages within each of the genomes were detected using PHASTER tool (Arndt et al., 2016). Contigs representing plasmid sequences were predicted based on the presence of typical plasmid-associated genes, such as replication and mobilisation genes, or based on similarity to published plasmids (www.ncbi.nlm.nih.gov).

5.3.3 Plasmid profiles

Plasmid DNA was isolated using a method described by O'Sullivan and Klaenhammer (1993). Plasmid DNA was run on a 0.7 % (w/v) agarose gel, and visualised by staining with ethidium bromide.

5.3.4 Minimal media to assess carbohydrate fermentation

Chemically defined MRS broth (CDMRS) was made by the omission of beef extract and any other additional sugar source and was subsequently used as a medium to examine the growth of three strains in the presence of different carbohydrate substrates. CDMRS contained the following: bacteriological peptone (Oxoid, Basingstoke, UK) 10 g, yeast extract (Merck, Darmstadt, Germany) 10 g, Tween® 80 (SigmaAldrich, St. Louis, MO, USA) 1 g, ammonium citrate 2 g, CH_3COONa 5 g, $MgSO_4$ 0.1 g, $MnSO_4$ 0.05 g, Na_2HPO_4 2 g (all products of SigmaAldrich) per 1 L of the medium. The pH of the media was adjusted to 6.4 and sterilised by autoclaving at 121°C for 15 min.

5.3.5 Carbohydrate fermentation

Initial screening of carbohydrate fermentation was performed using the commercial API50® kit (Biomerieux, Basingstoke, UK) following the manufacturer's instructions. Additionally, growth measurements in the presence of twelve selected carbohydrates (D-tagatose, L-sorbose, *myo*-inositol, D-lactose, D-saccharose, D-maltose, D-lyxose, pullulan, starch (all products of SigmaAldrich), amygdaline, inulin, L-arabitol (all products of AlphaAesar, Ward Hill, MA, USA) for each of the strains were performed by monitoring $OD600_{nm}$ using a Synergy HT plate reader (BioTek Instruments, Winsooski, VT, USA). Carbohydrate solutions were prepared by the addition of the carbohydrate of interest (1 % w/v) to the RMRS followed by

filter sterilisation (0.45 μm filter, Sarstedt, Wexford, Ireland). 500 μL of supplemented CDMRS was inoculated with 1 % (v/v) of an overnight bacterial culture grown in MRS at 30°C. The inoculated samples were grown at 30°C and OD_{600nm} readings were taken after 48 h, by placing 200 μL of a culture in triplicate in 96 well plate. Significance of differences in growth was tested by One-way Analysis of Variance (ANOVA), followed by Least Significant Test (LSD), performed in R statistical software (www.r-project.org).

5.3.6 Growth in the presence of heavy metal salts

Insensitivity to cobalt, cadmium and arsenic ions was determined by measuring OD_{600nm} in a 96-well microplate. MRS was supplemented with increasing concentrations of $CoCl_2$, $CdCl_2$, and Na_2HAsO_4 (all products of SigmaAldrich) from 0.25 to 6 mM and autoclaved at 121°C for 15 min. Following inoculation at 1 % (v/v) with cultures grown in the absence of heavy metal salts, growth was determined in triplicate for each concentration of heavy metal salt after 24 h of incubation. Significance of differences in growth was tested by One-way Analysis of Variance (ANOVA), followed by Least Significant Test (LSD), performed in R statistical software.

5.3.7 Putrescine production

To determine if the strains produce putrescine, strains were grown in Moller Decarboxylase broth (Moller, 1954). Briefly, the broth contained bacteriological peptone (Oxoid) 5 g, meat extract (Merck) 5 g, glucose 0.5 g, bromcresol purple 0.01 g, cresol red 0.005 g, pyridoxal-5'-phosphate 0.005 g, and L-arginine 10 g (all products of SigmaAldrich) per 1 L of medium. The final pH was set to 6.0±0.2, and the medium was autoclaved at 121°C for 15 min. The strains were inoculated in the

medium at 1 % (v/v) and incubated at 30°C for 24 h. A yellow colour indicated a negative reaction, and a purple colour indicated a positive reaction (i.e. putrescine production).

5.3.8 Exopolysaccharide (EPS) production

EPS production was determined by plating strains on reconstituted MRS plates. The specific agar contained the following: bacteriological peptone (Oxoid) 10 g, yeast extract (Merck) 10 g, meat extract (Merck) 10 g, Tween®80 (SigmaAldrich) 1 g, ammonium citrate 2 g, CH_3COONa 5 g, $MgSO_4$ 0.1 g, $MnSO_4$ 0.05 g, Na_2HPO_4 2 g (all products of SigmaAldrich), agar (Oxoid) 15 g, and glucose or saccharose (SigmaAldrich), 20 g per 1 L of media. Strains were inoculated on the prepared agar plates, and incubated for 48 hours at 30°C. EPS production was tested by examination of colonies for a ropy phenotype. Additionally, EPS production was determined on ruthenium agar plates, prepared as described by Mora et al. (2002) and Amina et al. (2014). Dark pink colonies represent EPS-producing strains. In both assays, strain DPC1116, previously confirmed to be an EPS producer, was used as a positive control.

5.4 Results and discussion

5.4.1 Genome characteristics of DPC2071, DPC4206 and DPC4536

Each of the three strains that were the subject of this study were previously designated as *Lactobacillus paracasei*, according to the results of 16S rRNA PCR (Stefanovic et al., 2017b) and current nomenclature rules (Tindall, 2008). The main features of their genomes are reported in Table 1. All three genomes had a GC content of 46.3 % and genome size of approximately 3 Mbp, typically observed in *L. paracasei*.

In pairwise comparisons of the genomes using the Mauve alignment tool, genes specific for each of the strains were identified. In Figure 1a, regions specific for strain DPC2071 correspond mainly to plasmid content, and a Type II CRISPR system, while specific regions in DPC4206 and DPC4536 code for phage remnants and a Type I CRISPR system (Fig. 1b and 1c). When the genomes of DPC4206 and DPC4536, strains with the same PFGE fingerprint, were aligned by BLASTn, it was shown, as expected, that the level of identity was very high (99 %, Table 1). However, the genome of DPC4206 is slightly larger, and, unlike DPC4536, it carries a single plasmid (Fig. 1b, 2). Although they shared the majority of their content, specific genes not present in the other strain were detected in both of the genomes.

5.4.2 Plasmid-encoded markers suggest a more complex evolutionary route for DPC2071

Plasmids often encode genes of technological importance, such as lactose utilisation, bacteriocin production and phage resistance (Wang and Lee, 1997). However, plasmids of *L. paracasei* encode a limited number of functional genes with a high prevalence of hypothetical proteins (Smokvina et al., 2013). Previous reports have

cited that, in general, strains of *L. paracasei* harbour up to four (strain NFBC338) (Desmond et al., 2005) or perhaps even six plasmids (strain Lpp120) (Smokvina et al., 2013). However, the plasmid profile of DPC2071 suggests that this strain possesses a total of 11 plasmids (Fig. 2). This large number of plasmids was confirmed upon genome analysis, with many contigs encoding plasmid specific features, such as plasmid replication or plasmid mobilisation genes. Many of the predicted proteins identified on these contigs were designated as hypothetical, but certain proteins with assigned functions, such as pullulanase, thiol disulfide isomerase, collagen adhesion protein, cation transporting ATPase, pyridine-nucleotide disulfide oxidoreductase, were also identified. Apart from similarity to plasmids *of L. paracasei*, many of the plasmid-associated contigs displayed similarity to plasmids of closely-related *L. rhamnosus* (Contig 38), to plasmids of the dairy species *L. helveticus* (Contig 14), *L. plantarum* (Contig 30) or more distantly related lactobacilli, such as *Lactobacillus hokkaidoensis* (Contig 13) and *Lactobacillus backii* (Contig 34). *L. hokkaidoensis* is a psychrophilic obligate heterofermentative LAB isolated from plant material or silage (Tanizawa et al., 2015), while *L. backii* has been isolated from spoiled beer (Tohno et al., 2013, Geissler et al., 2016). Additionally, Contig 14 (plasmid) was abundant in genes encoding hypothetical proteins belonging to other genera, such as *Pediococcus*, or other unrelated lactobacilli (*L. diolivorans, L. parakefiri, L. brevis, L. suebicus*). Again, some of these species are directly connected to fermenting plant material, such as *L. suebicus* isolated from cider (Ibarburu et al., 2015) and *L. diolivorans* isolated from spoiled cider juice (Martinez Viedma et al., 2009) or maize silage (Krooneman et al., 2002). Similarly, Contig 26 (not a plasmid contig) was shown to encode a large number of proteins with low level of query covers and low levels of

identity with other known proteins (50 %). These proteins have been shown to be mainly involved in EPS synthesis and corresponded to other lactobacilli (*L. plantarum, L. crispatus, L. rhamnosus*) or *Oenococcus oeni*. Such a high number of plasmids and an unusual genetic content of diverse origin in certain genome regions points to potential interactions of DPC2071 with varying environments and the organisms therein during the evolution of this strain. It is plausible that this strain changed environments and took part in numerous genetic exchange events, which contributed to its heterogeneous gene content.

5.4.3 Specificities of carbohydrate utilisation of three cheese isolates

It is believed that *Lactobacillus* species that are cheese specialists have lost numerous genes for various carbohydrate utilisation and transcriptional regulation of carbohydrate utilisation, as the dairy niche has a very limited spectrum of available carbohydrates with lactose predominating (Smokvina et al., 2013). In a study by Broadbent et al. (2012), the most restrictive sugar utilisation profiles were detected among cheese isolates, compared to plant and human isolates, which were able to use a greater variety of sugars that are available in the constantly changing habitat of these isolates. Moreover, sugar utilisation profiles and gene content can indirectly indicate an organism's previous habitats or potential interaction with strains from different ecological niches.

5.4.3.1 Diverse carbohydrate utilisation profiles

In order to determine sugar utilisation profiles, two approaches were used: an initial screening with API50® kit and followed by monitoring of growth in presence of twelve selected sugars. In the API assay, it was shown that strains differed in the utilisation of certain sugars. Unlike DPC2071, strains DPC4206 and DPC4536 were

able to use L-sorbose, D-maltose, inulin, D-tagatose and 5-ketogluconate, while in cases of *myo*-inositol and D-lyxose change of colour was small. However, DPC2071 was able to utilise amygdaline, a plant glucoside, and grew better in the presence of L-arabitol, compared to two other strains (colour change was more apparent). These results were confirmed in the subsequent analysis of growth in the presence of the selected sugars. Growth of *L. paracasei* strains in CDMRS did not exceed an OD_{600nm} of 0.45 after 48 h. Strain DPC2071 showed OD_{600nm} of 0.45 or less in the case of D-tagatose, L-sorbose, *myo*-inositol, D-maltose and inulin, and just slight growth of OD_{600nm} 0.64 in the presence of D-saccharose (Fig. 3). Indeed, genome comparison indicated that all genes for sorbose utilisation (L-sorbose-phosphate-reductase, transcriptional regulator, sorbitol-6-phosphate dehydrogenase, four components of sorbose specific PTS system and fructose-bisphosphate aldolase (Yebra et al., 2000)) were missing in DPC2071, but were present in DPC4206 and DPC4536. The presence of *fos* operon, which is involved in the utilisation of fructo-oligosaccharides (such as inulin), and the transport of free fructose (Goh et al., 2006), was confirmed in DPC4206 and DPC4536 (BWK52_0545 to BWK52_0551 in DPC4206 and B4Q23_187 to B4Q23_0193 in DPC4536), explaining the enhanced utilisation of this sugar by these two strains. The gene encoding the first enzyme in maltose degradation, maltose phosphorylase, is interrupted by a stop codon in DPC2071, resulting in an inability to use maltose.

The two strains that shared the same genomic structure fingerprints (DPC4206 and DPC4536) showed a broader range of potential carbohydrates as energy sources, as they were able to metabolise more sugars compared to strain DPC2071. However, the most interesting finding of this comparison was the absence of growth of DPC4536 in the presence of lactose. This was confirmed in the following

experiment, where the OD_{600nm} of this strain did not exceed 0.45, while two others reached level of 1.4 (Fig. 3). The presence of *lacG* gene, coding for 6-phospho-beta-galactosidase (EC 3.2.1.85), the first enzyme in lactose degradation in *Lactobacillus casei* (De Vos and Vaughan, 1994) in strain DPC4206 was confirmed by PCR (primers designed in this study, data not shown). This gene was located on the single plasmid present in DPC4206 (Contig 17), which complies with the findings of Siezen et al. (2005), who showed that lactose metabolism genes are often plasmid encoded. On the other hand, both genome analysis and PCR with *lacG* specific primers showed absence of this gene in strain DPC4536. Again, this gene could have been lost during plasmid depletion in strain DPC4536, thus affecting its ability to use lactose. The alternative way of lactose utilisation in some lactobacilli (*L. helveticus* and *L. acidophilus*) includes lactose transport *via* lactose permease (LacS) and further activity of beta-galactosidase, but this pathway has not been described in *L. paracasei* strains (Ganzle and Follador, 2012), and no permease was identified in the genome of DPC4536.

5.4.3.2 Higher numbers of BglG transcriptional regulators were present in the genomes of DPC4206 and DPC4536

A beta-glucoside operon (*bgl* operon) was firstly described in *E. coli*, where it regulated metabolism of beta-glucosides, such as salicin and arbutin (Mahadevan et al., 1987). Afterwards, similarly organised regulation systems involved in sugar metabolism in other bacteria were described, including lactose metabolism in *L. casei* (Alpert and Siebers, 1997). The *bgl*-type operons are induced by sugars, and they are regulated by two operon products: BglG- a transcriptional regulator (antiterminator), and BglF- a phosphotranferase that regulates phosphorylation of BglG and enables formation of dimers, the only active form of BglG (Nussbaum-

Shochat and Amster-Choder, 1999). BglG essentially binds to *bgl* transcript and prevents the formation of terminator structures (Nussbaum-Shochat and Amster-Choder, 1999). BglG homologs control synthesis of a specific EII component of carbohydrate transporters of the bacterial phosphotransferase system (PTS) for utilization of a particular sugar *via* a transcription antitermination mechanism reminiscent of the bgl system (Rothe et al., 2012). Five genes designated as coding for BglG transcriptional regulators (antiterminators) have been detected in the genomes of DPC4206 and DPC4536, and were not found in DPC2071. They differed in length of the protein, and they probably regulate transport of different sugars into the cell. The higher number of BglG transcriptional regulators could be connected with the higher span of sugar utilisation genes and wider number of sugars used as energy sources by these two strains compared to DPC2071, but only deeper analysis of substrate specificities of these antiterminators could reveal their actual significance in observed phenotypes.

5.4.3.3 Specific genes for the fermentation of plant derived carbohydrates

Myo-inositol (MI) is a sugar alcohol present in soil, and it is part of phytic acid, a phosphate storage molecule in plants. It can also be metabolised by bacteria that live in soil, but it is not frequently used as an energy source in LAB (Yebra et al., 2007). So far, strains of *L. casei* are the only members of LAB that are capable of weakly metabolising MI, but the presence of a MI metabolism cluster of genes is not a common feature of *L. casei* strains, and it does not necessarily mean that the strain carrying the cluster will use it as an energy source (Zhang et al., 2010). Previously, the presence of the complete MI utilisation operon was confirmed in the probiotic strain *L. casei* BL23 (Yebra et al., 2007). Here, strains DPC4206 and DPC4536 (BWK52_0229c to BWK52_0239 in DPC4206 and B4Q23_0140c to B4Q23_0150

in DPC4536), that possess the whole cluster of gene needed for utilisation of MI did not exceed in growth (OD_{600nm}=0.55) when compared in the same media without MI added (OD_{600nm}=0.45) (Fig. 3), analogous to API assay, where only a slight change of colour was observed. Similar findings were shown for strain *L. casei* 12A (Vinay-Lara et al., 2014).

Pullulan is one of the polysaccharides produced from starch, present in plant material or fermented products of plant origin. In the *Lactobacillus* genus, species that are connected with plant niches, such as *L. amylovorus*, *L. acidophilus*, *L. amylophilus*, *L. cellobiosus* have the ability to metabolise starch (Petrova et al., 2013). Strains able to degrade starch and its derivatives possess specific enzymes, such as neopullulanase, pullulanase and amylopullulanase that differ in the specificity of the link they break in a polysaccharide chain (Doman-Pytka and Bardowski, 2004, Ganzle and Follador, 2012). However, in dairy isolates, starch metabolism genes are not expressed but rather exist as pseudogenes due to the mutation in promotor, amylase catalytic domain or signal peptide (Petrova et al., 2013).

Interestingly, the genomes of all three strains analysed in this study possessed genes encoding starch degradation enzymes. Apart from neopullulanase (BLL69_0750, BWK52_1091, B4Q23_0861), and amylopullulanase (BLL69_2007c, BWK52_2351c, B4Q23_1259) encoding genes detected in all three genomes, strain DPC2071 possessed also a pullulanase encoding gene (BLL69_ 0389) located on plasmid. However, none of the strains examined in this study could use pullulan or starch as an energy source (Fig. 3). An alignment of the amylopullulanase protein sequence from the three strains matched with the protein previously reported in *L. paracasei* B41 (Petrova and Petrov, 2012) (Fig. 6) but the substitution of three amino acids in the catalytic domain could be the reason for the lack of the starch degrading

phenotype. The gene encoding neopullulanase in *L. mucosae* LM1 was recently characterised (Balolong et al., 2016). When LM1 neopullulanse protein sequence was aligned with neopullulanase from the three sequenced genomes (by Clustal W), all four catalytic modules and a conserved MPKL motif were detected (not shown). The possible explanations for the lack of pullulan degradation could be the absence of transporters facilitating transportation of oligosaccharides such as panose, since these are scarce in the dairy niche. Additionally, transportation of maltodextrines, partial starch degradation products, requires specific ABC transporters (Sauvageot et al., 2017), which were not identified in any of the three strains sequenced in this study.

5.4.4 Genomic content as an indicator of the flavour development potential of the cheese isolates

Flavour development in bacterial ripened cheeses originates mainly from the metabolic activities of bacteria present during ripening (Marilley and Casey, 2004). Although glycolysis and lipolysis contribute to the development of flavour compounds, proteolysis and amino acid metabolism particularly are seen as major contributors (Smit et al., 2005). In the previous work of our group, it was shown that the three strains analysed in this study possess different activities of enzymes of the proteolytic cascade (cell envelope proteinase, aminopeptidases, aminotransferases) and they had variable volatile profiles in two cheese model systems (Stefanovic et al., 2017b, Stefanovic et al., 2017c). However, the genomic comparison of the three strains did not reveal any genetic content differences in regard to the components of proteolytic cascade, except for methionine metabolic pathway described below. This means that varying abilities of these strains for the development of flavour compounds most probably come as the consequence of different activities of the key

enzymes or their regulation, such as the impact of coenzymes, and not due to the different number of key enzyme encoding homologs.

5.4.4.1 Higher number of cystathionine lyases encoding genes explains higher potential for volatile sulfur compounds production in DPC4206

Volatile sulfur compounds (VSC) that arise during the microbial metabolism of sulfur compounds (methionine, cysteine) are essential for the aroma of many food products including cheese (Landaud et al., 2008). Compounds such as methanethiol, dimethyl-disulfide, dimethyl-trisulfide, dimethyl-tetrasulfide, and methional contribute to notes of onion, garlic, and cabbage in some types of cheese, such as Cheddar (Singh et al., 2003). In bacterial amino acid metabolism, transamination represents the main pathway of amino acid degradation. The aminotransferase converts methionine to 4-methylthio-2-oxobutanoic acid, which is further converted to various VSC (Landaud et al., 2008). Besides the aminotransferase pathway, the possible involvement of cystathionine lyases in VSC production has been recently reported (Fernandez and Zuniga, 2006, Lee et al., 2007, Bustos et al., 2011, Bogicevic et al., 2013), although these enzymes are primarily involved in methionine biosynthesis (Yvon and Rijnen, 2001). Cystathionine lyases (cystathionine beta lyase (CBL), EC 4.4.1.8; and cystathionine gamma lyase (CGL), EC 4.4.1.1)) can use various sulfur containing substrates, including methionine, to produce methanethiol (Fernandez and Zuniga, 2006). In the study of Bustos et al. (2011) it was shown that VSC producing abilities of LAB (*Lactococcus lactis, Lactobacillus spp., Streptococcus thermophilus* and *Brevibacterium linens*) correlated with the cystathionine lyase activities. Similarly, strains possessing cystathionine lyase genes used in cheese manufacture contributed to significantly higher levels of VSC at the end of ripening (Bogicevic et al., 2013). Lee et al. (2007) showed that

overexpression of CBL in *L. helveticus* resulted in higher production of VCS from methionine and cystathionine.

The three genomes analysed in this study differed in content of CBL and CGL. Strain DPC2071 had one gene encoding CBL (BLL69_0664), and two genes encoding CGL (BLL69_0264, BLL69_0493c). In strain DPC4206, two CBL genes (BWK52_1002, BWK52_3061c) and two CGL genes (BWK52_0733c, BWK52_3092) were identified, while in strain DPC4536 two CBL genes (B4Q23_0772, B4Q23_2254c), and only one CGL gene (B4Q23_0463c) were present. Additionally, all three genomes possess genes encoding cystathionine beta synthase (CBS, EC 4.2.1.22), that catalyses reaction of conversion homocysteine to cystathionine, indirectly involved in sulfur compounds metabolism. Strains DPC2071 and DPC4536 have one homolog of CBS (BLL69_0263 and B4Q23_0715, respectively), while DPC4206 has two homologs (BWK52_0941, BWK52_3091). Closer investigation showed that BWK52_3092 and BWK52_3091 in DPC4206 are located on plasmid-associated contigs, and appear to have been lost from strain DPC4536. The presence of the higher number of homologs for both CBL and CGL in strain DPC4206 could be the reason for more efficient methionine degradation observed when strains were grown in the media with an increased concentration of methionine (Stefanovic et al., 2017b). This feature is seen as a very important in cheese manufacture, and strains with optimal VCS production are potential candidates for adjunct selection.

5.4.4.2 DPC4206 and DPC4536 have additional genes encoding pyridoxine-5'-phosphate oxidases

The other possibility for the observed diversity of flavour compounds produced could be the effect of the coenzymes, such as vitamin B_6. Pyridoxine-5'-phosphate oxidase (EC 1.4.3.5) catalyses the conversions of pyridoxamine-phosphate and pyridoxine-phosphate to pyridoxal-phosphate, the biologically active form of vitamin B_6. Pyridoxal-phosphate presents the coenzyme for amino acid converting enzymes, such as amino acid transaminases and decarboxylases, which have an important role in flavour compound development. While the same gene for pyridoxine-5'-phosphate oxidase is present in all three strains (BLL69_1931c in DPC2071, BWK52_0201 in DPC4206 and B4Q23_0112 in DPC4536), strains DPC4206 and DPC4536 possess additional homologs of this enzyme, three of which were common for these two strains: BWK52_1254, BWK52_1493c, and BWK52_2866c in DPC4206 and B4Q23_1069c, B4Q23_2263 and B4Q23_2899 in DPC4536. However, DPC4206 had an additional gene, located on plasmid designated Contig 41 (BWK52_3104c), which was not present in DPC4536. Higher number of homologs in strains DPC4206 and DPC4536 could lead to higher activities of amino acid converting enzymes. However, in our previous study, we did not see that effect, although transaminase activity towards phenylalanine was the only one determined (Stefanovic et al., 2017b). Nevertheless, in a more complex surrounding with the higher number of both substrates available and active metabolic pathways, increased level of pyridoxal-phosphate could be the determining point in more efficient amino acid metabolism of the strains.

5.4.5 DPC2071 shows resistance to the toxic effect of cadmium and arsine salts

Bacteria possess numerous mechanisms that protect them from the increased levels of heavy metal ions they potentially encounter in certain environments. The presence of these ions may result in the formation of reactive oxygen species (ROS), which disrupt the normal physiology of the cell. Some of the protective measures include heavy metal efflux pumps and oxido-reductive reactions (Solioz et al., 2011). The growth of the strains was examined in the presence of three metal salts. All three strains were able to grow in MRS containing up to 6 mM of $CoCl_2$ with only slight decrease in growth at the highest concentration of $CoCl_2$ (Fig. 4). Cobalt is an essential component of coenzyme B_{12} (Rodionov et al., 2006), involved in numerous metabolic reactions, and cells grew well in the presence of cobalt and were able to efficiently use the ion as a cofactor, but suffered no toxic effects in the analysed concentration of cobalt salt. On the other hand, cadmium and arsine are not identified to be involved in normal metabolic processes in the cell and express toxic effects (Trevors et al., 1986, Cervantes et al., 1994). Cells of the analysed strains were sensitive to $CdCl_2$ at concentrations higher than 1 mM. However, strain DPC2071 was the most adaptable to the presence of cadmium ion, as $OD600_{nm}$ of this strain was significantly higher in all concentrations of $CdCl_2$ above 0.25 mM (Fig. 4). Similarly, this strain was the only one able to grow in 0.5 mM of Na_2HAsO_4, while the other two strains could not grow in this concentration of arsenic salt (Fig. 4). The exclusive presence of the arsenical pump ATPase (BLL69_0465c) and arsenical resistance operon repressor (BLL69_0466c) in DPC2071 could explain the growth of this strain in presence of up to 0.5 mM of arsenic ion. Besides that, this strain possesses additional specific genes that could help in resisting oxidative stress caused by the elevated concentrations of heavy

metals (Birben et al., 2012), such as specific glutathione reductase and thiol disulfide isomerases. Glutathione reductase (EC 1.8.1.7) (BLL69_0554) catalyses reduction of glutathione disulfide to glutathione, a critical molecule in resisting oxidative stress. Thiol disulfide isomerases (EC 5.3.4.1) (BLL69_0399, BLL69_0417), which appear to be plasmid-encoded, catalyse the proper formation and breakage of disulfide bonds between cysteine residues within proteins as they fold and correct wrongly folded proteins as well. They are also involved in oxido-reductive stability of proteins and protein isomerisation (Ali Khan and Mutus, 2014).

5.4.6 DPC2071 possess unusual arginine metabolism gene

Apart from the development of flavour compounds, microbial metabolism of amino acids in food products can lead to biogenic amines (BA) production. In microorganisms, BA contribute to numerous physiological functions, such as supply of metabolic energy, resistance to acidic pH and regulation of osmotic and oxidative stresses (Benkerroum, 2016). However, excessive BA production is undesirable in dairy products, since their toxic effects on humans have been shown (Ladero et al., 2010). Putrescine is a BA that originates from arginine metabolism. It is one of the most common BAs in food produced by microorganisms used in food manufacture, such as starter cultures, but also by food contaminants, such as *Pseudomonas spp*. or *Enterobacteriaceae* (Wunderlichová et al., 2014).

Generally, in Gram-positive bacteria, there are two metabolic pathways of putrescine biosynthesis: the ornithine decarboxylase pathway (ODC) and the agmatine deiminase (AgDI) pathway (Fig. 5) (Wunderlichová et al., 2014). Additionally, a biosynthetic route where agmatine is directly converted into putrescine by the action of the agmatinase (EC 3.5.3.11) has been described mainly in *Enterobacteriaceae*,

but also in some dairy-borne contaminants such as *Bacillus* spp. and *Pseudomonas* spp. (Benkerroum, 2016) (Fig. 5). In the publicly available genomes of *Lactobacillus paracasei*, the agmatinase-encoding gene was reported in only three strains (Lpl7, Lpl14 and CNCM I-4270), all isolated from cereals (Smokvina et. al, 2013). Interestingly, in the genome of DPC2071 the same gene, (BLL69_2612) was detected. The presence of agmatinase in strain DPC2071 adds up to the set of unusual genes present in DPC2071 genome. However, this route cannot contribute to putrescine production in DPC2071, since the gene encoding arginine decarboxylase, which transforms arginine to agmatine, was not identified.

In regard to putrescine production, although some components of putrescine synthesis pathways were detected in the genomes of the strains, no putrescine production was confirmed in the assay. All three strains have gene encoding for biodegradable ornithine decarboxylase that transforms L-ornithine to putrescine and possess genes for transport of putrescine/spermidine. Nevertheless, arginase, which converts arginine to L-ornithine, was not detected in any of the three strains, thus confirming incomplete putrescine synthetic route in the strains.

5.4.7 CRISPR array content provides evidence of the independent evolution of DPC4206 and DPC4536

Clustered regularly interspaced short palindromic repeats (CRISPR) systems coupled with CRISPR associated proteins, are the most recently described phage resistance system. They are composed of a *cas* operon and a CRISPR array that contains a string of DNA repeats and spacers. Spacers correspond to foreign DNA inserted between two repeats and confirm previous encounters of the strain with different phages. Several types of CRISPR systems have been reported so far (Types I, II and

III), which differ in mechanism of action and the target molecule (Rath et al., 2015, Hille and Charpentier, 2016). Novel systems (Types IV, V and VI) have been recently described (Wright et al., 2016).

In DPC2071, a Type II CRISPR system was detected. Upon analysis of spacers in DPC2071 in two separate CRISPR arrays, 30 and 18 spacers were identified, 17 of which were common for both of the arrays. The genome analysis showed that the *cas9* gene, a signature gene of Type II systems was broken by a transposase. It means that, at least in the past, this CRISPR system was efficient in conferring phage resistance, as confirmed by the presence of spacers, and the transposase probably had been inserted in the *cas9* recently, thus impairing its activity.

Both DPC4206 and DPC4536 possessed Type I CRISPR systems. The CRISPR arrays of DPC4206 and DPC4536 contained 34 and 24 spacers were identified, respectively, 21 of which were present in the genomes of both strains. Although the genomes of these two strains are highly similar, their CRISPR systems differ in numbers and specificity of spacers, confirming their recent divergence and independent evolutions during which they encountered different phages.

5.4.8 Cell surface molecules and secreted components

5.4.8.1 Exopolysaccharide production was detected in anaerobic conditions

Many LAB produce exopolysaccharide (EPS) that are excreted as slime (ropy form) or remain attached to the bacterial cell wall forming capsular EPS (Vuyst and Degeest, 1999, Peant et al., 2005). However, compared to strains isolated from the plant environment or gut isolates, dairy isolates usually carry the smallest number of EPS biosynthesis genes (Smokvina et al., 2013). EPS production is considered a valuable feature, as EPS improves the rheology and texture of dairy products, such

as yoghurt (Welman and Maddox, 2003). However, the sole presence of these enzymes and molecules is not a guarantee of EPS synthesis, as these molecules are part of numerous metabolic pathways in the cell, and should be referred to as "housekeeping enzymes" (Welman and Maddox, 2003).

A number of genes required for EPS biosynthesis were observed in all three genomes. In addition to various EPS synthesis genes (Contig 26, reported above), strain DPC2071 possesses specific genes components of *rfb* operon (dTDP-glucose pyrophosphorylase, dTDP-4-dehydrorhamnose 3,5-epimerase, dTDP-glucose 4,6-dehydratase and dTDP-4-dehydrorhamnose reductase (BLL669_2024c to BLL669_2027c)) that enable rhamnosyl-units to be incorporated into the repeat unit of EPS (Trefzer et al., 1999, Boels et al., 2004). Strains DPC4206 and DPC4536 also possess genes for EPS backbone production (BWK52_0503 to BWK52_0515 in DPC4206 and B4Q23_1965c to B4Q23_1977c in DPC4536), different to the ones encoded in DPC2071. However, although the necessary genes are present in all three strains, no ropy phenotype on MRS plates with increased concentration of glucose or saccharose was detected for any of the tested strains. On the other hand, when grown in anaerobic conditions (anaerobic jar) on ruthenium red milk agar plates, white colonies (considered as positive for EPS production) were observed for each of the three strains, as well as for the positive control. However, no ropy phenotype was observed for any of the three analysed strains in ruthenium red plates. It is possible that anaerobic conditions contribute to development of some type of sugar cell envelope. On the other hand, in aerobic conditions, no EPS production was observed on ruthenium red plates under aerobic conditions of incubation. This confirms that despite the extensive knowledge of EPS gene organisation, definite mechanisms of regulation of EPS biosynthesis remain unclear (Peant et al., 2005).

5.4.8.2 Specific collage-adhesion encoding genes in DPC2071 suggest the potential for gut colonisation

Adhesion proteins play an important role in adherence of bacteria to epithelia of tissue and colonise, at least transiently, mucosal surfaces of gastrointestinal tract (Cai et al., 2009). Successful adherence of cells is an important feature of strains with potential probiotic application (Sanchez et al., 2008). In the three sequenced strains, common fibronectin binding proteins (BLL69_1428, BWK52_1733 and B4Q23_1531) were detected, as well as large adhesion proteins (BLL69_2806c, B4Q23_2822c, and in strain DPC4206 two parts of adhesin were in separate contigs and after designing the specific primers, the presence of the whole gene protein was confirmed). On the other hand, strain DPC2071 possessed two additional collagen-adhesion proteins (BLL69_0529c and BLL69_2882), which could contribute to more efficient adherence for gut epithelial cells of this strain compared to two other strains. Besides that, all three strains possess the same pili synthesis genes of spaCBA type, where SpaA is a backbone-forming major pilin, SpaB is a minor pilin and SpaC is essential for the mucus adherence (Reunanen et al., 2012). However, none of the three strains possessed any genes encoding mucin-binding proteins, unlike typical gut isolates that have high number of mucin-binding proteins (Cai et al., 2009), which could be the limiting factor for their successful adherence to the gut surface.

5.5 Conclusion

This study demonstrated the variability that exists between genomes of cheese isolates of *L. paracasei*. The specific genes and specific homologs of genes detected in three strains facilitated the differences in their metabolic potential, production of flavour contributing compounds and ability to survive in presence of growth inhibitors, such as heavy metals. Strain DPC2071 was characterised by high number of plasmids, unusual for *Lactobacillus* strains. The genetic content of DPC2071 revealed its interesting past and potential habitats, as well as numerous interactions with other strains of lactobacilli, not usually connected with the dairy niche. Two strains with the same PFGE pattern and with highly similar genomes (DPC4206 and DPC4536) shared genetic content, but some differences were evident. The plasmid was not present in DPC4536, and probably parts of it were integrated in chromosome, while rest of it was lost. One of the most important differences between the two strains is the loss of ability of DPC4536 to use lactose as an energy source. Apart from that, distinct constitution of parts of their CRISPR arrays confirms recent independent evolution of these two strains and independent encounter with phages. From all of these, it could be proposed that DPC4536 evolved from DPC4206, and this event was recent, as still 99 % of genomes were identical. This study shows far reaching conclusions based on the genome comparison of strains isolated from the same ecological niche, and additionally confirms previously reported high level of genetic diversity of *L. paracasei* strains.

5.6 Acknowledgments

Many thanks to Francisco Avila Cobos and Dr Raul Cabrera Rubio for their valuable advice in genome comparison.

5.7 References

Ali Khan, H. & Mutus, B. 2014. Protein disulfide isomerase a multifunctional protein with multiple physiological roles. *Front Chem,* 2**,** 70.

Alikhan, N.-F., Petty, N. K., Zakour, N. L. B. & Beatson, S. A. 2011. BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. *BMC Genomics,* 12**,** 402.

Alpert, C.-A. & Siebers, U. 1997. The lac operon of *Lactobacillus casei* contains *lacT,* a gene coding for a protein of the BglG family of transcriptional antiterminators. *J Bacteriol,* 179**,** 1555-1562.

Amina, Z., Noureddine, S., Venkatesan, A., Perumal, V., Hichem, B., Asma, Z., Yamina, M., Miloud, H. & Mebrouk, K. 2014. Characterization and potential probiotic attributes of *Lactobacillus plantarum* DU10 isolated from Algerian raw camel milk. *Biotechnol,* 13**,** 282.

Arndt, D., Grant, J. R., Marcu, A., Sajed, T., Pon, A., Liang, Y. & Wishart, D. S. 2016. PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Res,* 44**,** W16-W21.

Balolong, M. P., Chae, J. P. & Kang, D. K. 2016. Expression and characterisation of neopullulanase from *Lactobacillus mucosae*. *Biotechnol Lett,* 38**,** 1753-1760.

Benkerroum, N. 2016. Biogenic Amines in Dairy Products: Origin, Incidence, and Control Means. *Compr Rev Food Sci Food Saf,* 15**,** 801-826.

Birben, E., Sahiner, U. M., Sackesen, C., Erzurum, S. & Kalayci, O. 2012. Oxidative stress and antioxidant defense. *World Allergy Organ J,* 5**,** 9-19.

Boels, I. C., Beerthuyzen, M. M., Kosters, M. H. W., Van Kaauwen, M. P. W., Kleerebezem, M. & de Vos, W. M. 2004. Identification and Functional Characterization of the *Lactococcus lactis rfb* Operon, Required for dTDP-Rhamnose Biosynthesis. *J Bacteriol,* 186**,** 1239-1248.

Bogicevic, B., Fuchsmann, P., Breme, K., Portmann, R., Guggenbuhl, B. & Irmler, S. 2013. A preliminary study on the effect of *Lactobacillus casei* expressing cystathionine lyase1/cystathionine lyase2 on Cheddar cheese and the formation of sulphur-containing compounds. *Int Dairy J,* 33**,** 97-103.

Broadbent, J. R., Neeno-Eckwall, E. C., Stahl, B., Tandee, K., Cai, H., Morovic, W., Horvath, P., Heidenreich, J., Perna, N. T., Barrangou, R. & Steele, J. L. 2012. Analysis of the *Lactobacillus casei* supragenome and its influence in species evolution and lifestyle adaptation. *BMC Genomics,* 13**,** 533.

Bustos, I., Martinez-Bartolome, M. A., Achemchem, F., Pelaez, C., Requena, T. & Martinez-Cuesta, M. C. 2011. Volatile sulphur compounds-forming abilities of lactic acid bacteria: C-S lyase activities. *Int J Food Microbiol,* 148**,** 121-127.

Cai, H., Thompson, R., Budinich, M. F., Broadbent, J. R. & 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.

Ceapa, C., Lambert, J., van Limpt, K., Wels, M., Smokvina, T., Knol, J. & Kleerebezem, M. 2015. Correlation of *Lactobacillus rhamnosus* Genotypes and Carbohydrate Utilization Signatures Determined by Phenotype Profiling. *Appl Environ Microbiol,* 81**,** 5458-5470.

Cervantes, C., Ji, G., Ramirez, J. & Silver, S. 1994. Resistance to arsenic compounds in microorganisms. *FEMS Microbiol Rev,* 15**,** 355-367.

Darling, A. C., Mau, B., Blattner, F. R. & Perna, N. T. 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res,* 14**,** 1394-1403.

De Vos, W. M. & Vaughan, E. E. 1994. Genetics of lactose utilization in lactic acid bacteria. *FEMS Microbiol Rev,* 15**,** 217-237.

Desmond, C., Ross, R. P., Fitzgerald, G. & Stanton, C. 2005. Sequence analysis of the plasmid genome of the probiotic strain *Lactobacillus paracasei* NFBC338 which includes the plasmids pCD01 and pCD02. *Plasmid,* 54**,** 160-175.

Doman-Pytka, M. & Bardowski, J. 2004. Pullulan degrading enzymes of bacterial origin. *Crit Rev Microbiol,* 30**,** 107-121.

Fernandez, M. & Zuniga, M. 2006. Amino acid catabolic pathways of lactic acid bacteria. *Crit Rev Microbiol,* 32**,** 155-183.

Ganzle, M. G. & Follador, R. 2012. Metabolism of oligosaccharides and starch in lactobacilli: a review. *Front Microbiol,* 3**,** 340.

Geissler, A. J., Behr, J. & Vogel, R. F. 2016. Multiple Genome Sequences of the Important Beer-Spoiling Species *Lactobacillus backii*. *Genome Announc,* 4**,** e00826-16.

Goh, Y. J., Zhang, C., Benson, A. K., Schlegel, V., Lee, J. H. & Hutkins, R. W. 2006. Identification of a putative operon involved in fructooligosaccharide utilization by *Lactobacillus paracasei*. *Appl Environ Microbiol*, 72**,** 7518-30.

Grissa, I., Vergnaud, G. & Pourcel, C. 2007. CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Res,* 35**,** W52-W57.

Hille, F. & Charpentier, E. 2016. CRISPR-Cas: biology, mechanisms and relevance. *Philos Trans R Soc Lond B Biol Sci,* 371.

Ibarburu, I., Puertas, A. I., Berregi, I., Rodriguez-Carvajal, M. A., Prieto, A. & Duenas, M. T. 2015. Production and partial characterization of exopolysaccharides produced by two *Lactobacillus suebicus* strains isolated from cider. *Int J Food Microbiol,* 214**,** 54-62.

Krooneman, J., Faber, F., Alderkamp, A., Elferink, S. O., Driehuis, F., Cleenwerck, I., Swings, J., Gottschal, J. & Vancanneyt, M. 2002. *Lactobacillus diolivorans* sp. nov., a 1,2-propanediol-degrading bacterium isolated from aerobically stable maize silage. *Int J Syst Evol Microbiol,* 52**,** 639-646.

Ladero, V., Calles-Enríquez, M., Fernández, M. & A Alvarez, M. 2010. Toxicological effects of dietary biogenic amines. *Curr Nut Food Sci,* 6**,** 145-156.

Landaud, S., Helinck, S. & Bonnarme, P. 2008. Formation of volatile sulfur compounds and metabolism of methionine and other sulfur compounds in fermented food. *Appl Microbiol Biotechnol,* 77**,** 1191-1205.

Lee, W. J., Banavara, D. S., Hughes, J. E., Christiansen, J. K., Steele, J. L., Broadbent, J. R. & Rankin, S. A. 2007. Role of cystathionine beta-lyase in catabolism of amino acids to sulfur volatiles by genetic variants of *Lactobacillus helveticus* CNRZ 32. *Appl Environ Microbiol,* 73**,** 3034-3039.

Mahadevan, S., Reynolds, A. & Wright, A. 1987. Positive and negative regulation of the *bgl* operon in *Escherichia coli*. *J Bacteriol,* 169**,** 2570-2578.

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

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

Martinez Viedma, P., Abriouel, H., Sobrino Lopez, A., Ben Omar, N., Lucas Lopez, R., Valdivia, E., Martin Belloso, O. & Galvez, A. 2009. Effect of enterocin AS-48 in combination with high-intensity pulsed-electric field treatment against the spoilage bacterium *Lactobacillus diolivorans* in apple juice. *Food Microbiol,* 26**,** 491-496.

McKay, L. L. & Baldwin, K. A. 1984. Conjugative 40-megadalton plasmid in *Streptococcus lactis* subsp. *diacetylactis* DRC3 is associated with resistance to nisin and bacteriophage. *Appl Environl Microbiol,* 47**,** 68-74.

Medini, D., Donati, C., Tettelin, H., Masignani, V. & Rappuoli, R. 2005. The microbial pan-genome. *Curr Opin Genet Dev,* 15**,** 589-594.

Moller, V. 1954. Distribution of amino acid decarboxylases in *Enterobacteriaceae*. *Acta Pathol Microbiol Scand,* 35**,** 259-277.

Mora, D., Fortina, M., Parini, C., Ricci, G., Gatti, M., Giraffa, G. & Manachini, P. 2002. Genetic diversity and technological properties of *Streptococcus thermophilus* strains isolated from dairy products. *J Appl Microbiol,* 93**,** 278-287.

Nussbaum-Shochat, A. & Amster-Choder, O. 1999. BglG, the transcriptional antiterminator of the bgl system, interacts with the β′ subunit of the *Escherichia coli* RNA polymerase. *Proc Natl Acad Sci U S A,* 96**,** 4336-4341.

O'Sullivan, D. J. & Klaenhammer, T. R. 1993. Rapid mini-prep isolation of high-quality plasmid DNA from *Lactococcus* and *Lactobacillus* spp. *Appl Environ Microbiol,* 59**,** 2730-2733.

O'Sullivan, O., O'Callaghan, J., Sangrador-Vegas, A., McAuliffe, O., Slattery, L., Kaleta, P., Callanan, M., Fitzgerald, G. F., Ross, R. P. & Beresford, T. 2009. Comparative genomics of lactic acid bacteria reveals a niche-specific gene set. *BMC Microbiol,* 9**,** 50.

Peant, B., LaPointe, G., Gilbert, C., Atlan, D., Ward, P. & Roy, D. 2005. Comparative analysis of the exopolysaccharide biosynthesis gene clusters from four strains of *Lactobacillus rhamnosus*. *Microbiology,* 151**,** 1839-1851.

Petrova, P. & Petrov, K. 2012. Direct starch conversion into L-(+)-lactic acid by a novel amylolytic strain of *Lactobacillus paracasei* B41. *Starch - Stärke,* 64**,** 10-17.

Petrova, P., Petrov, K. & Stoyancheva, G. 2013. Starch-modifying enzymes of lactic acid bacteria - structures, properties, and applications. *Starch - Stärke,* 65**,** 34-47.

Pfeiler, E. A. & Klaenhammer, T. R. 2007. The genomics of lactic acid bacteria. *Trends Microbiol,* 15**,** 546-553.

Rath, D., Amlinger, L., Rath, A. & Lundgren, M. 2015. The CRISPR-Cas immune system: biology, mechanisms and applications. *Biochimie,* 117**,** 119-128.

Reunanen, J., von Ossowski, I., Hendrickx, A. P., Palva, A. & de Vos, W. M. 2012. Characterization of the SpaCBA pilus fibers in the probiotic *Lactobacillus rhamnosus* GG. *Appl Environ Microbiol,* 78**,** 2337-2344.

Rodionov, D. A., Hebbeln, P., Gelfand, M. S. & 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.

Rothe, F. M., Bahr, T., Stülke, J., Rak, B. & Görke, B. 2012. Activation of *Escherichia coli* antiterminator BglG requires its phosphorylation. *Proc Natl Acad Sci U S A,* 109**,** 15906-15911.

Sanchez, B., Bressollier, P. & Urdaci, M. C. 2008. Exported proteins in probiotic bacteria: adhesion to intestinal surfaces, host immunomodulation and molecular cross-talking with the host. *FEMS Immunol Med Microbiol,* 54**,** 1-17.

Sauvageot, N., Mokhtari, A., Joyet, P., Budin-Verneuil, A., Blancato, V. S., Repizo, G. D., Henry, C., Pikis, A., Thompson, J., Magni, C., Hartke, A. & Deutscher, J. 2017. *Enterococcus faecalis* Uses a Phosphotransferase System Permease and a Host Colonization-Related ABC Transporter for Maltodextrin Uptake. *J Bacteriol,* 199**,** e00878-16.

Schroeter, J. & Klaenhammer, T. 2009. Genomics of lactic acid bacteria. *FEMS Microbiol Lett,* 292**,** 1-6.

Siezen, R. J., Renckens, B., van Swam, I., Peters, S., van Kranenburg, R., Kleerebezem, M. & 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.

Singh, T., Drake, M. & Cadwallader, K. 2003. Flavor of Cheddar cheese: a chemical and sensory perspective. *Compr Rev Food Sci Food Saf,* 2**,** 166-189.

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

Smokvina, T., Wels, M., Polka, J., Chervaux, C., Brisse, S., Boekhorst, J., van Hylckama Vlieg, J. E. & Siezen, R. J. 2013. *Lactobacillus paracasei* comparative genomics: towards species pan-genome definition and exploitation of diversity. *PLoS One,* 8**,** e68731.

Solioz, M., Mermod, M., Abicht, H. K. & Mancini, S. 2011. Responses of lactic acid bacteria to heavy metal stress. *In*: Tsakalidou, E. & Papadimitriou, K. (eds.) *Stress Responses of Lactic Acid Bacteria.* Springer.

Stefanovic, E., Fitzgerald, G. & McAuliffe, O. 2017a. Draft Genome Sequences of Three *Lactobacillus paracasei* Strains, Members of the Nonstarter Microbiota of Mature Cheddar Cheese. *Genome Announc,* 5**,** e00655-17.

Stefanovic, E., Kilcawley, K. N., Rea, M. C., Fitzgerald, G. F. & McAuliffe, O. 2017b. Genetic, enzymatic and metabolite profiling of the *Lactobacillus casei* group reveals strain biodiversity and potential applications for flavour diversification. *J Appl Microbiol,* 122**,** 1245-1261.

Stefanovic, E., Thierry, A., Maillard, M. B., Bertuzzi, A., Rea, M. C., Fitzgerald, G. F., McAuliffe, O. & Kilcawley, K. N. 2017c. Strains of the *Lactobacillus casei* group show diverse abilities for the production of flavor compounds in 2 model systems. *J Dairy Sci,* 100**,** 6918-6929.

Sun, Z., Harris, H. M., McCann, A., Guo, C., Argimon, S., Zhang, W., Yang, X., Jeffery, I. B., Cooney, J. C., Kagawa, T. F., Liu, W., Song, Y., Salvetti, E., Wrobel, A., Rasinkangas, P., Parkhill, J., Rea, M. C., O'Sullivan, O., Ritari, J., Douillard, F. P., Paul Ross, R., Yang, R., Briner, A. E., Felis, G. E., de Vos, W. M., Barrangou, R., Klaenhammer, T. R., Caufield, P. W., Cui, Y., Zhang, H. & O'Toole, P. W. 2015. Expanding the biotechnology potential of lactobacilli through comparative genomics of 213 strains and associated genera. *Nat Commun,* 6**,** 8322.

Tanizawa, Y., Tohno, M., Kaminuma, E., Nakamura, Y. & Arita, M. 2015. Complete genome sequence and analysis of *Lactobacillus hokkaidonensis* LOOC260(T), a psychrotrophic lactic acid bacterium isolated from silage. *BMC Genomics,* 16**,** 240.

Tindall, B. 2008. The type strain of *Lactobacillus casei* is ATCC 393, ATCC 334 cannot serve as the type because it represents a different taxon, the name *Lactobacillus paracasei* and its subspecies names are not rejected and the revival of the name *'Lactobacillus zeae'*contravenes Rules 51b (1) and (2) of the International Code of Nomenclature of Bacteria. Opinion 82. *Int J Syst Evol Microbiol,* 58**,** 1764-1765.

Toh, H., Oshima, K., Nakano, A., Takahata, M., Murakami, M., Takaki, T., Nishiyama, H., Igimi, S., Hattori, M. & Morita, H. 2013. Genomic adaptation of the *Lactobacillus casei* group. *PLoS One,* 8**,** e75073.

Tohno, M., Kitahara, M., Irisawa, T., Masuda, T., Uegaki, R., Ohkuma, M. & Tajima, K. 2013. Description of *Lactobacillus iwatensis* sp. nov., isolated from orchardgrass (*Dactylis glomerata* L.) silage, and *Lactobacillus backii* sp. nov. *Int J Syst Evol Microbiol,* 63**,** 3854-3860.

Trefzer, A., Bechthold, A. & Salas, J. A. 1999. Genes and enzymes involved in deoxysugar biosynthesis in bacteria. *Nat Prod Rep,* 16**,** 283-299.

Trevors, J. T., Stratton, G. W. & Gadd, G. M. 1986. Cadmium transport, resistance, and toxicity in bacteria, algae, and fungi. *Can J Microbiol,* 32**,** 447-464.

Vinay-Lara, E., Hamilton, J. J., Stahl, B., Broadbent, J. R., Reed, J. L. & Steele, J. L. 2014. Genome-scale reconstruction of metabolic networks of *Lactobacillus casei* ATCC 334 and 12A. *PLoS One,* 9**,** e110785.

Vuyst, L. & Degeest, B. 1999. Heteropolysaccharides from lactic acid bacteria. *FEMS Microbiol Rev,* 23**,** 153-177.

Wang, T. T. & Lee, B. H. 1997. Plasmids in *Lactobacillus*. *Crit Rev Biotechnol,* 17**,** 227-272.

Welman, A. D. & Maddox, I. S. 2003. Exopolysaccharides from lactic acid bacteria: perspectives and challenges. *Trends Biotechnol,* 21**,** 269-274.

Wright, A. V., Nunez, J. K. & Doudna, J. A. 2016. Biology and Applications of CRISPR Systems: Harnessing Nature's Toolbox for Genome Engineering. *Cell,* 164**,** 29-44.

Wunderlichová, L., Buňková, L., Koutný, M., Jančová, P. & Buňka, F. 2014. Formation, Degradation, and Detoxification of Putrescine by Foodborne Bacteria: A Review. *Compr Rev Food Sci Food Saf,* 13**,** 1012-1030.

Yebra, M. J., Veyrat, A., Santos, M. A. & Pérez-Martínez, G. 2000. Genetics of L-sorbose transport and metabolism in *Lactobacillus casei*. *J Bacteriol,* 182**,** 155-163.

Yebra, M. J., Zuniga, M., Beaufils, S., Perez-Martinez, G., Deutscher, J. & Monedero, V. 2007. Identification of a gene cluster enabling *Lactobacillus casei* BL23 to utilize myo-inositol. *Appl Environ Microbiol,* 73**,** 3850-3858.

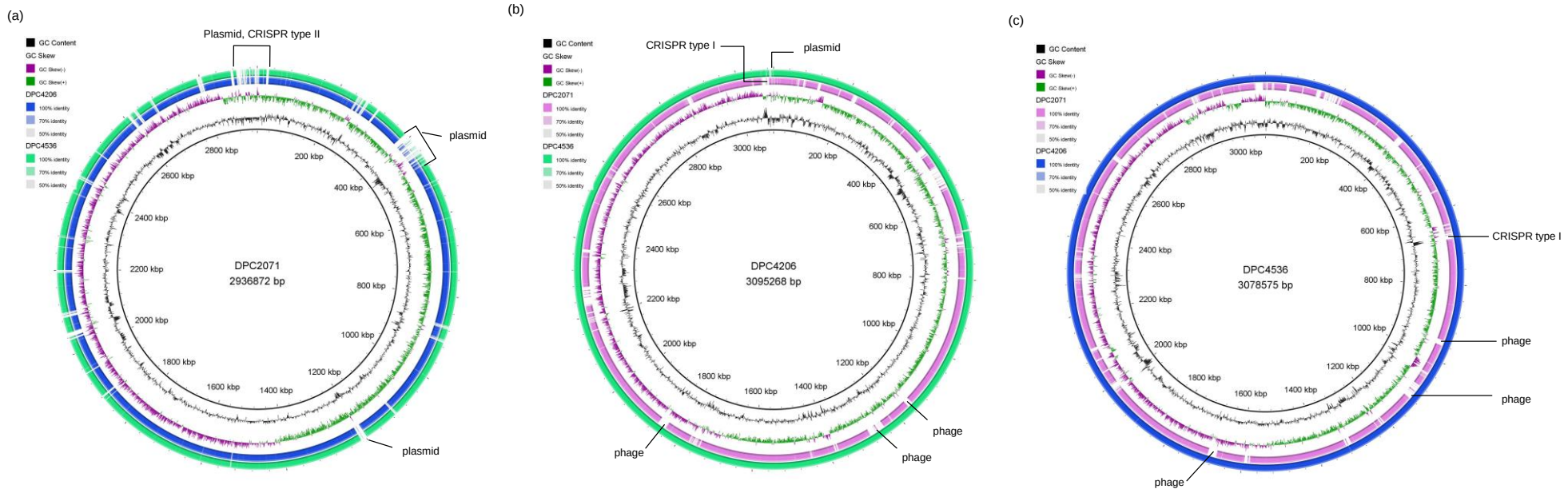
Yvon, M. & Rijnen, L. 2001. Cheese flavour formation by amino acid catabolism. *Int Dairy J,* 11**,** 185-201.

Zhang, W. Y., Sun, Z. H., Yu, D. L., Airideng, C., Chen, W., Meng, H. & Zhang, H. P. 2010. Comparative analysis of iol clusters *in Lactobacillus casei* strains. *World J Microbiol Biotechnol,* 26**,** 1949-1955.

Table 1: General characteristics of genomes of three strains of *Lactobacillus paracasei* (Stefanovic et al., 2017a).

	DPC2071	DPC4206	DPC4536
Genome length (bp)	2.936.872	3.095.268	3.078.575
Contigs	41	49	35
GC content	46.3 %	46.3 %	46.3%
No of CDS	2827	2951	2931
No of plasmids	11	1	0
Locus tag	BLL69	BWK52	B4Q23
Nucleotide sequence blast (BLASTn) between strains (% of query coverage, % of identity, E-value)			
	DPC2071 subject	DPC4206 subject	DPC4536 subject
DPC2071 query	-	89%, 99%, 0.0	89%, 99%, 0.0
DPC4206 query	86%, 99%, 0.0	-	99%, 99%, 0.0
DPC4536 query	86%, 99%, 0.0	99%, 99%, 0.0	-

Figure 1: Circular maps of *Lactobacillus paracasei* strains using (a) strain DPC2071, (b) strain DPC4206 and (c) strain DPC4536 as reference genomes, denoted as the inner rings. The outer rings denote genomes DPC2071 (pink) (b,c), DPC4206 (blue) (a,c) , and DPC4536 (green) (a,b).



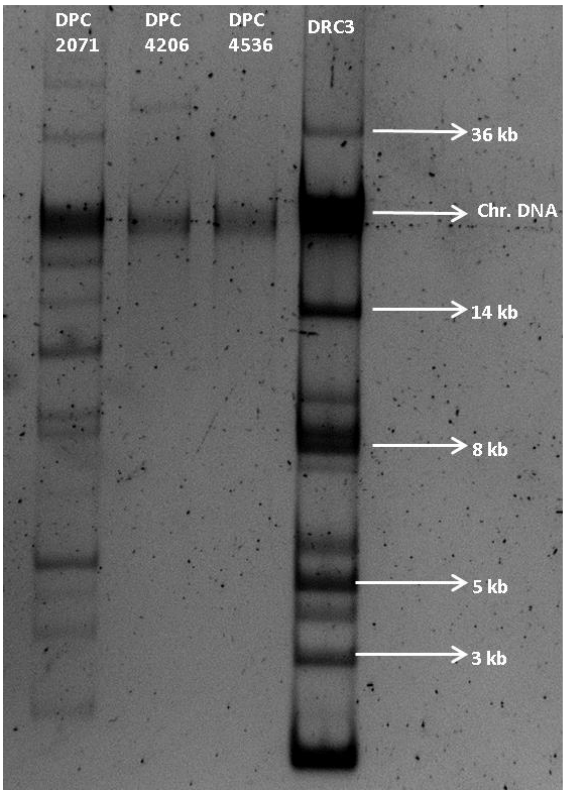


Figure 2: Plasmid profile of three *Lactobacillus paracasei* strains: DPC2071, DPC4206 and DPC4536. The plasmid profile of *Lactococcus lactis* ssp. *lactis* biovar. *diacetylactis* DRC3 provides reference for estimation of plasmids sizes (McKay and Baldwin, 1984).

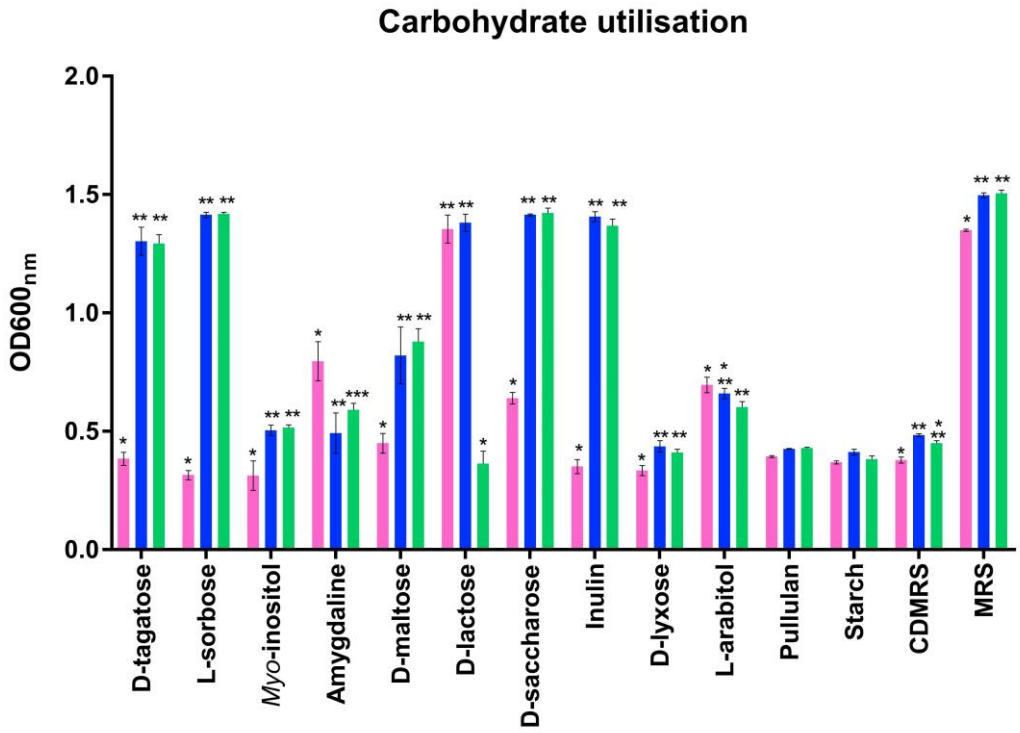
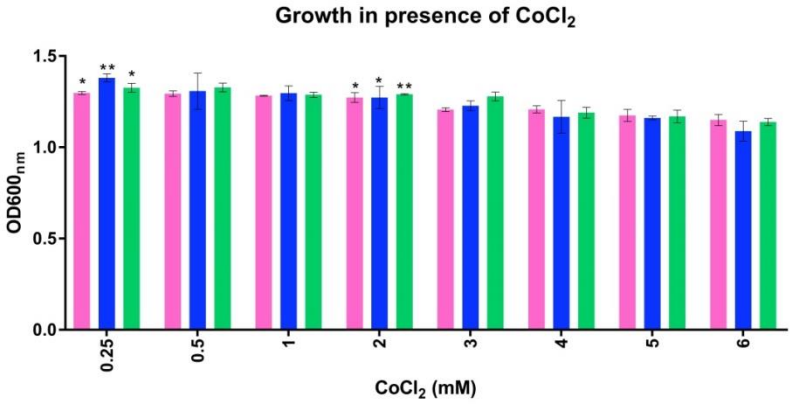


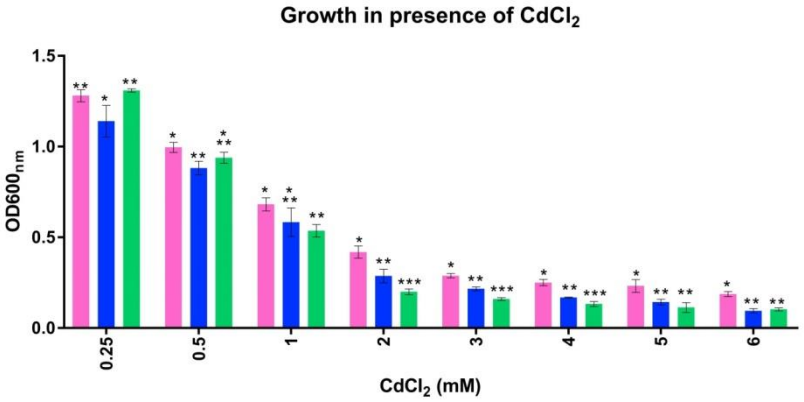
Figure 3: Growth of three strains of *Lactobacillus paracasei* in reconstituted MRS (CDMRS) supplemented with a single sugar in concentration 1 % (w/v) and incubated over 48 h at 30°C. Bars represent OD_{600nm} at the end of 48 h. Bars for the same sugar sharing the same asterix symbol show no statistical difference in growth ($p>0.05$), after mean comparison by performing One-way Analysis of Variance (ANOVA) followed by Least Significant Test (LSD). Error bars present standard deviation.

Legend: DPC2071 DPC4206 DPC4536

(a)



(b)



(c)

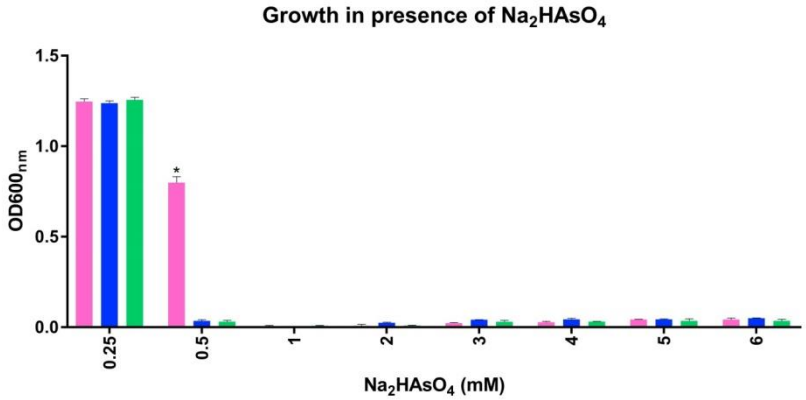


Figure 4: Growth of three strains of *Lactobacillus paracasei* in MRS supplemented with corresponding heavy metal salt (a) $CoCl_2$, (b) $CdCl_2$ and (c) Na_2AsO_4, inoculated in concentration 1 % (w/v) and incubated over 48 h at 30°C. Bars represent OD_{600nm}. Bars for the same heavy metal salt concentration sharing the same asterix symbol show no statistical difference in growth ($p>0.05$), after mean comparison by performing One-way Analysis of Variance (ANOVA) followed by Least Significant Test (LSD). Error bars present standard deviation.

Legend: DPC2071 DPC4206 DPC4536

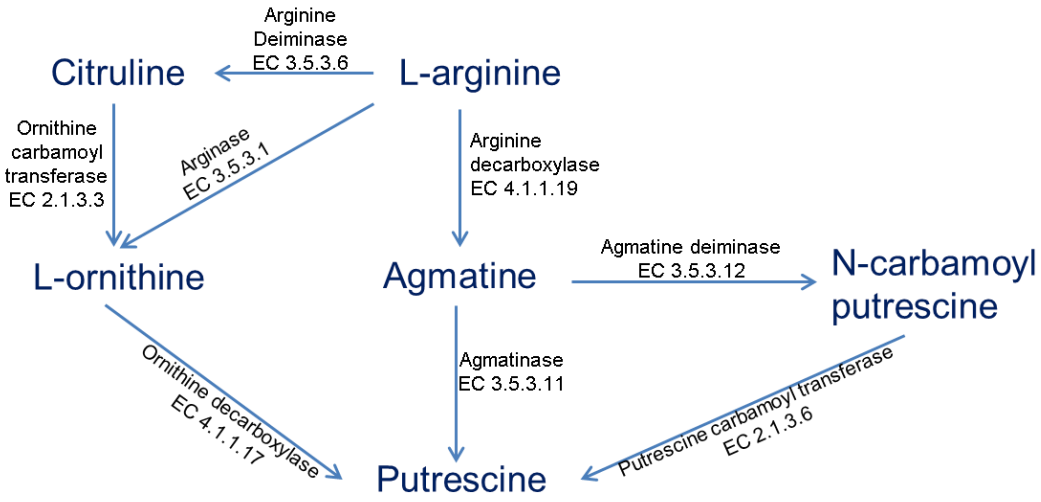


Figure 5: Arginine metabolic pathways in bacteria. The biogenic amines can be produced in food by microorganisms *via* presented metabolic pathways. Modified from Benkerroum, 2016 and Wunderlichová et al., 2014.



Figure 6: The partial representation of the alignment of amylopullulanase protein in *Lactobacillus paracasei* strains DPC2071, DPC4206, DPC4536 and B41, by ClustalW. The conserved regions, reported by Doman-Pytka and Bardowski, 2004 and Petrova and Petrov, 2012 are boxed. The amino acids that differ from the consensus sequence are marked with the different colour.

Chapter 6

Evaluation of the potential of *Lactobacillus paracasei* adjuncts for cheese flavour development and diversification

6.1 Abstract

The non-starter microbiota of Cheddar cheese mostly comprises mesophilic lactobacilli, such as *Lactobacillus casei, L. paracasei, L. rhamnosus* and *L. plantarum*. These bacteria are recognised for their potential to improve Cheddar cheese flavour when used as adjunct cultures. In this study, three strains of *L. paracasei* (DPC2071, DPC4206 and DPC4536) were evaluated for their contribution to enhancement and diversification of Cheddar flavour. The strains were selected based on their genomic diversity, variability in proteolytic enzyme activities and volatile profiles generated in cheese model systems. The addition of adjunct cultures did not affect the gross composition of the cheeses. The levels of total free amino acids (FAA) in cheeses showed a significant increase after 28 days of ripening. However, the concentrations of individual amino acids did not significantly differ between cheeses except for several amino acids (aspartic acid, threonine, serine, and tryptophan) at Day 14. Volatile profile analysis revealed that the main compounds that differentiated the cheeses were of lipid origin, such as long-chain aldehydes, acids, ketones and lactones, and the differences were more pronounced in earlier stages of ripening. Sensorial analysis showed that cheeses were perceived as similar, and cheese with the DPC2071 adjunct having a slightly better acceptance. This study showed that the three *Lactobacillus paracasei* strains used as adjuncts had a minimal impact on Cheddar flavour diversity under the conditions used, and are more suitable for flavour diversification in short-ripened cheeses.

6.2 Introduction

The production of cheese worldwide shows a global increase year-on-year with an annual production of over 22 million tonnes (www.dairy.ahdb.org, report from Feb 2017). With such a high market demand, the dairy industry is challenged by increasing consumer requirements for products of novel flavour. Thus, the industry is seeking a means of enhancement and diversification of cheese flavour. One of the factors influencing flavour development is the general chemical composition of cheese (Lynch et al., 1999). However, the metabolic activities driven by the cheese microbiota during ripening represent the major force of flavour development (El Soda et al., 2000, Yvon, 2006). Besides innovation in the technology applied in cheese manufacturing, the alteration of the microbial populations in the cheese represents a potential tool for flavour diversification (Van Hoorde et al., 2010).

The microbiota of Cheddar cheese comprises the starter lactic acid bacteria (SLAB) and non-starter lactic acid bacteria (NSLAB). SLAB acidify milk during fermentation (El Soda et al., 2000), but they also contribute to the flavour development due to their metabolic activity (Wouters et al., 2002, Kieronczyk et al., 2003). NSLAB represent the endogenous secondary flora (Wouters et al., 2002), and these organisms dominate the later stages of Cheddar cheese ripening (Burns et al., 2012). The NSLAB population of Cheddar cheese includes homo- and heterofermentative mesophilic lactobacilli: *Lactobacillus casei, L. paracasei, L. rhamnosus, L. plantarum* and *L. brevis* (Fitzsimons et al., 1999, Gobbetti et al., 2015). These bacteria show adaptability to environments with limited nutrient amounts which occur in later stages of Cheddar ripening (Hussain et al., 2009). They use mainly peptides and amino acids as nitrogen and energy sources, since residual

lactose is present at low levels (Settanni and Moschetti, 2010). Additionally, the potential substrates for NSLAB include nucleic acids and sugars from glycoproteins and glycolipids of the lysed starters (Steele et al., 2006).

NSLAB have a prominent role in cheese flavour development (Crow et al., 2002), as cheeses made in aseptic conditions with starter bacteria developed poor flavour profiles (Wijesundera et al., 1997). Additionally, cheeses made with raw milk that have higher NSLAB levels than pasteurised milk, develop stronger flavour (Fox et al., 1998). NSLAB contribute to the intensification of flavour and increased overall acceptability mainly through impact on secondary proteolysis in cheese and metabolism of free amino acids (FAA) (McSweeney and Fox, 1997, Lynch et al., 1999, Di Cagno et al., 2006, Milesi et al., 2009). However, in some instances, non-starter flora can contribute to the formation of off-flavours, especially in the later phases of ripening (Crow et al., 2001, Gobbetti et al., 2015).

Because of their generally positive effect on cheese flavour development, mesophilic lactobacilli are often added deliberately to cheese milk as adjunct cultures during industrial production. Apart from the direct impact on flavour development, they expedite the ripening and control the adventitious microflora (Milesi et al., 2010, Singh and Singh, 2014). In terms of their ability to improve cheese flavour, strains of the *L. casei* group, especially the species *L. casei* and *L. paracasei*, are one of the most extensively explored NSLAB. In Cheddar, the application of *L. paracasei* strains as adjuncts improved both flavour intensity and bitterness (Lynch et al., 1999, Ong et al., 2007a). Milesi et al. (2010) showed that Cremoso cheeses with a *L. paracasei* adjunct strain had a similar composition to the control cheese with no adjunct, but the overall acceptability was higher. In Manchego cheeses, the addition of *L. paracasei* strains as adjuncts improved the flavour of the cheeses (Poveda et al.,

2014). In Gouda cheeses, *L. paracasei* adjuncts contributed to cheese flavour diversification (Van Hoorde et al., 2010). Strain-specific effects of *L. paracasei* adjuncts originally isolated from Danbo cheese were observed when these strains were examined in cheese model system, as some strains contributed to flavour improvement, while others led to the development of off-flavours (Antonsson et al., 2003).

Proteolytic reactions that occur during cheese manufacture and ripening are seen as major contributors to texture and flavour development (McSweeney and Sousa, 2000). The main protein of milk is casein, and its degradation by rennet and intrinsic milk proteinases releases large peptides (primary proteolysis), that are further metabolised by proteinases of starter and non-starter (or adjunct) bacteria to release small peptides (Sousa et al., 2001). Subsequently, bacterial peptidases release free amino acids (FAA) (secondary proteolysis), which contribute directly to the cheese flavour, but also indirectly through their metabolism by microbial amino acid converting enzymes, which is considered to be one of the main pathways for flavour development (Ardo, 2006, Yvon, 2006). In addition to proteolysis, other pathways such as lipolysis and glycolysis also contribute to cheese flavour. In the lactose/citrate pathway, pyruvate represents the central metabolite, and it is further degraded into acetaldehyde, ethanol, diacetyl, acetoin, all of which are important cheese flavour contributors (Marilley and Casey, 2004). The lipolytic pathways include a complex network of reactions, in which numerous long-chain alcohols, acids, methyl-ketones and lactones that have various aroma notes arise (Collins et al., 2003).

The aim of the present work was to evaluate the flavouring capacity of three *L. paracasei* strains when used as adjunct cultures in Cheddar cheese manufacture. The

three strains used in this study were selected according to their proteolytic characterisation in enzymatic assays, and their production of flavour compounds in cheese model systems (Stefanovic et al., 2017a, Stefanovic et al., 2017b).

6.3 Materials and methods

6.3.1 Bacterial strains used in cheese manufacture

The starter culture used in cheese production was *Lactococcus lactis* ssp. *lactis* 303. In addition to the starter, three adjunct cultures, all belonging to the *L. casei* group and designated as *Lactobacillus paracasei* (DPC2071, DPC4206 and DPC4536) were used. The strains were kept frozen at -80°C in the appropriate medium (LM17 broth (Merck, Darmstadt, Germany) for the starter culture, and MRS broth (Oxoid, Basingstoke, UK) for adjunct cultures, supplemented with glycerol (20 % (v/v)). Prior to the cheese making, strains were grown on LM17 or MRS agar plates, for starter and adjuncts, respectively, at 30°C.

6.3.2 Cheese manufacture and ripening

The Control cheese contained only starter culture, while each of the three test cheeses contained, apart from the starter, one of the three adjunct cultures. Test cheeses were named according to the adjunct used (i.e. cheese DPC2071, cheese DPC4206 and cheese DPC4536). Cheeses were manufactured in a pilot plant. For cheese making, bulk starter cultures (1 % v/v) were inoculated in 7 L of 10 % (w/v) heat treated (90°C for 30 min) reconstituted skim milk (RSM) and incubated at 30°C for 18 h. Afterwards, cultures were cooled, and kept at 4°C for 18 h until the cheese manufacture the following day. Adjunct strains DPC2071 and DPC4206 were grown in 500 mL of 10 % RSM (w/v) (autoclaved at 121°C for 5 min) with addition of 1 % (v/v) of yeast extract. Strain DPC4536 was grown 500 mL of MRS, as previous tests showed its poor growth in milk (data not shown). For inoculation into the vat, the 500 mL culture of DPC4536 was centrifuged (4000 g, 10 min 4 °C) and resuspended

in 500 mL of sterile 10 % (w/v) RSM. Three cheese making trials were performed on different dates.

Raw milk was standardised to a protein-to-fat ratio of approximately 0.96:1. Milk was pasteurised at 72°C for 15 s and pumped into cylindrical, jacketed 500 L vats. Milk (454 kg/vat) was inoculated with starter and appropriate adjunct, as described above. Coagulation was achieved over 30 min by adding chymosin (18 mL/100 kg, Chr. Hansen, Cork, Ireland) before a 5 min cut program. After curd cutting, the curd and whey mixture was cooked at a rate of 1°C/5 min to a maximum scald of 38°C. Subsequently, the curds and whey were drained at pH 6.20 and cheddared until pH of 5.3. Curds were milled and salted (2.75 % of NaCl (w/w)), and left to mellow for 20 min. Salted curds were moulded (2 × 22 kg) and pressed for 18 h. Cheeses were vacuum-packed and transferred to 8°C ripening room. Cheeses were ripened for 9 months.

6.3.3 Enumeration of starter and adjunct bacteria in cheeses

For bacteriological analysis, cheeses were aseptically sampled at Day 1, 14, 28 and Month 3, 6 and 9 of ripening. The samples were placed in a sterile stomacher bag, diluted 1:10 with sterile 2 % trisodium citrate and homogenised using a stomacher (Iul Instruments, Barcelona, Spain) for 5 min. Independent duplicate samples were taken at each time point and serial dilutions were prepared as required. Starter cells were enumerated on LM17 agar after incubation at 30°C for 3 days. Total NSLAB (lactobacilli) were enumerated on *Lactobacillus* selective (LBS) agar (BD, Oxford, UK) after 5 days incubation at 30°C. Coliforms were enumerated on violet red bile agar (BD) after incubation at 30°C for 1 day. Enterococci were enumerated on Kanamycin aesculin azide agar (Merck) after incubation at 30°C for 1 day.

To confirm that the majority of NSLAB belonged to the inoculated adjuncts, pulsed-field gel electrophoresis (PFGE) was performed as described previously (Stefanovic et al., 2017a). Isolates from two time points (Month 3 and Month 9) were analysed, and the PFGE patterns were compared with the patterns of the three adjuncts.

6.3.4 Cheese compositional and biochemical analysis

At Day 14 post manufacture, cheeses were sampled and grated for salt, protein, moisture and fat content determination. Salt and moisture were determined according to the IDF methods (IDF (1979) and IDF (1982), respectively). Fat content was determined by CEM Smart Turbo Moisture/Solids analyser (CEM Corporation, Matthews, NC, USA). Additionally, primary and secondary proteolysis was monitored from Day 14 until Month 9 of ripening. Primary proteolysis was determined using the macro-Kjeldahl method (IDF, 1993). Secondary proteolysis was determined by measuring the free amino acid (FAA) content of the pH 4.6 soluble nitrogen extracts (pH4.6SN) according to the method described by McDermott et al. (2016).

6.3.5 Free fatty acids (FFA) analysis of cheese lipid extracts by Gas Chromatography with Flame Ionisation Detector (GC-FID)

FFA content of cheeses was determined at two time points of ripening: Month 3 and Month 9. Lipid extraction was performed according to the procedure outlined by De Jong and Badings (1990) with the following modifications: 4 g of sample was mixed with 10 g anhydrous Na_2SO_4 by grinding with a mortar and pestle. 0.3 mL of 2.5 M H_2SO_4 and 1 mL of an internal standard (ISTD) (C5, C11, C17 at 1000 ppm in heptane) were added to each sample. The samples were extracted 3 times with 15 mL of diethyl ether/heptane (1:1) and each time the solution was clarified by

centrifugation at 3000 g for 5 min. The collected extracts were pooled for solid phase extraction.

The 500 mg aminopropyl columns were pre-conditioned with 10 mL of heptane. The lipid extract was applied to the column and the neutral lipids removed using 10 mL of 20 % diethyl ether in hexane. At no point were the columns left to dry. The FFA were collected using 5 mL of 2 % formic acid/diethyl ether (2 % FA/DE) in glass test tubes. The entire extract was immediately separated and stored in 2 mL amber vials which were capped with PTFE/white silicone septa (Agilent Technologies, Cork, Ireland). Amber vials were used to prevent ultraviolet light degradation of any polyunsaturated fatty acids that may be present in the extract.

Gas chromatography was performed on Varian CP3800 GC with a CP FFAP CB capillary column (30 m × 0.25 mm × 0.32 μm, Agilent Technologies). The injector was held at 25°C using cryogenics (liquid carbon dioxide) for 6 sec, this was raised to 250°C at 30°/min. The injector and auto-sampler were operated in on-column mode. The injection volume of the extracts obtained above was 0.5 μL. The inlet liner used was a SPI direct liner (Agilent Technologies). The carrier gas was helium and was held at a constant flow of 1.2 mL/min. The column oven was held at 40°C for 2 min and raised to 240°C at 7.5°C/min, and this was held for 23.33 min. The total run time was 52 min. The Flame Ionisation Detector (FID) was operated at 300°C. The identification of FFA in the samples was performed based on retention times of FFA in the standard mix (GLC Reference STANDARD 74 “Free acid”, Nu-Chek-prep, Inc., Waterville, MN, USA) used for the instrument calibration.

6.3.6 Cheese volatile analysis by Gas Chromatography-Mass Spectrometry (GC-MS)

The volatile compounds present in the cheeses were determined at two time points, Month 3 and Month 9 of ripening. For each cheese sample, 4 g of grated cheese was placed in an amber 20 mL screw capped HS-SPME vial with a silicone/PTFE septum (Apex Scientific, Maynooth, Ireland). Initially, the vials were equilibrated to 40°C for 10 min with pulsed agitation of 5 seconds at 500 rpm (Shimadzu AOC 5000 plus autosampler). Solid phase microextraction (SPME) was performed with a 50/30 µm Carboxen®/ divinylbenzene/ polydimethylsiloxane (CAR/DVB/PDMS, Agilent Technologies) fibre, which was exposed to the headspace above the samples for 20 min at 40°C. After extraction, the fibre was injected into the GC inlet and desorbed for 2 min at 250 °C into a SPL injector with a SPME liner. Gas chromatography was performed on a Shimadzu 2010 Plus GC with a DB-5 (60 m × 0.25 mm × 0.25 µm, Agilent Technologies) column using a split/splitless injector in split mode 1:10. The carrier gas (helium) was maintained at 23 psi. The temperature of the column oven was set at 35°C, held for 5 min, increased at 6.5°C/min to 230°C then increased at 15°C/min to 320°C. The mass spectrometer detector Shimadzu TQ8030 was run in single quad mode. The ionisation was done by electronic impact (-70 eV) and the mass range m/z scanned between 35 and 250 amu. All samples were analysed in triplicate. To ensure there was no carry over between the samples, the SPME fibre was cleaned using a bake-out station at 270°C for 3 min. During the run, vials with external standards (dimethyl-sulfide, benzaldehyde, cyclohexanone, butyl acetate, acetone, and ethanol at concentrations of 10 ppm) were analysed to ensure that analysis was done within specification. Blanks (empty vials) were injected regularly to monitor possible carry over.

The volatile compounds were identified using mass spectra comparisons to the NIST 2014 mass spectral library (Scientific Instrument Services, Ringoes, NJ, USA), Flavors and Fragrances of Natural and Synthetic Compounds Library and an in-house library created in GC-MS Solutions software (Mason Technology, Dublin, Ireland) with target and qualifier ions and linear retention indices for each compound. Spectral deconvolution was also performed to confirm identification of compounds using AMDIS. An auto-tune of the GC-MS was carried out prior to the analysis to ensure optimal GC-MS performance. The compounds of interest were selected according to previously published review of compounds considered as main flavour contributors in cheese (Curioni and Bosset, 2002, Singh et al., 2003).

6.3.7 Sensory Affective Evaluation and Ranking Descriptive Analysis of cheeses

Cheeses were assessed by a sensory panel at Month 9 of ripening. Forty-two assessors aged between 19 and 25 were recruited in University College Cork, Ireland. Selection criteria were availability and motivation to participate on all days of the experiment and consumption of Cheddar cheese in everyday nutrition. Sensory Affective Evaluation (SAE) was performed according to Stone and Sidel (2004) and Stone et al. (2012a), and Ranking Descriptive Analysis (RDA) according to Dairou and Sieffermann (2002) and Richter et al. (2010). For RDA, panellists were briefly trained. Assessors used the sensory hedonic descriptors for SAE and intensity descriptors for RDA listed in Tables 3 and 4, respectively.

Sensory analysis was carried out in panel booths conforming to international standards (ISO 8589: 2007). All samples were stored at 4˚C until required. The cheeses were presented to the assessor panel at ambient temperature (21°C) and

coded with a randomly selected 3 digit code. Each assessor was provided with deionised water and instructed to cleanse their palates between tastings. For SAE, each assessor was asked to indicate their degree of liking on a 10 cm line scale ranging from 0 (extremely dislike) at the left to 10 (extremely like) at the right and rating subsequently scored in cm from left. For RDA, each assessor was asked to assess the intensity of the attributes, according a 10 cm line scale ranging from 0 (none) at the left to 10 (extreme) at the right and rating subsequently scored in cm from left. The order of the presentation of all test samples was randomised to prevent first order and carryover effects. For RDS, samples were presented in duplicate. Samples were presented under white light (1000 LUX).

6.3.8 Statistical analysis

To determine if significant differences exist in cheese composition, fat, salt, moisture, protein, FAA and FFA content among cheese samples were analysed by Analysis of Variance (ANOVA) carried out using R statistical software (www.r-project.org). Sensory analysis attributes were analysed by ANOVA in SPSS v24 (IBM Corporation, Armonk, NY, USA). Means were compared using the least significant difference (LSD) test, and the level of significance was determined at $p<0.05$. For volatile analysis, the principal component analysis (PCA) was performed on selected signals resulting from chromatogram processing using the package FactomineR of the R software.

6.4 Results

6.4.1 Starter and adjunct enumeration showed typical evolution of cheese microbiota

In all four cheeses, similar trends in starter numbers were observed (Fig. 1a). The starter culture was inoculated at approximately 10^9 CFU/g cheese, and remained at this level in the first 14-28 days of ripening, after which the numbers decreased by about 2 $\log_{10}$ units by Month 9, when the lowest numbers of starter culture cells were present in cheese DPC2071. The inoculum of the three adjuncts used was between 10^8 and 10^9 CFU/g. Regarding the NSLAB counts, the numbers gradually increased over the ripening time in the Control cheese from 10^1 CFU/g at production to 10^8 CFU/g at Month 9 of ripening. In the three test cheeses, numbers of NSLAB increased after Day 28, reached a peak at Month 6, and slightly decreased at Month 9 (Fig. 1b). Using unique PFGE profiles as an indicator of the presence of individual strains revealed that in each test cheese, the patterns of NSLAB at the highest dilution corresponded to the patterns of inoculated adjunct in each of the vats. At Month 3 in the Control cheese, different PFGE profiles were observed (I1, I6, Fig. 2a), but the profiles of the inoculated adjuncts used in the experimental vats were still dominating (Fig. 2a). At Month 9, the profiles of the isolates from the Control cheese fully corresponded to the adjuncts used in the other three vats (Fig. 2b).

6.4.2 The presence or choice of adjunct cultures did not influence gross cheese composition

The determination of cheese composition was performed on Day 14 (Table 1). No significant differences ($p<0.05$) were observed between the cheeses for levels of fat,

moisture, dry matter, salt, pH, salt in moisture (S/M), fat in dry matter (FDM) and moisture in non-fat solids (MNFS).

6.4.3 Free amino acid content in cheeses significantly differed in first weeks of ripening

The level of pH 4.6 soluble nitrogen expressed as a percentage of total nitrogen (pH4.6SN/TN (%)) was used to measure the level of primary proteolysis. It increased significantly over the ripening period ($p<0.05$) in all four cheeses, reaching approx. 19 % at Month 9 (Fig. 3a). No effect specific to the inoculated adjuncts was observed.

Secondary proteolysis was determined as the level of FAA liberated from peptides. A significant increase in total free amino acids for each of the cheeses was observed after Day 28 of ripening ($p<0.05$, Fig. 3b). When total FAA content was compared among the cheeses at each time point, no significant difference was observed except at Month 6, where levels of total FAA were significantly higher in DPC4536 than in the Control cheese (Fig. 3b).

The concentrations of individual amino acids at each time point did not significantly differ among the four cheeses, except for aspartic acid, threonine, serine, and tryptophan in the samples at Day 14 (Fig. 4a). No significant differences were observed at Day 28 among the samples, while at Months 3 and 6, a significant difference was observed only in the case of aspartic acid, which was present at a significantly higher concentration in cheese DPC4536 (data not shown). At Month 9, no significant differences in the concentration of any individual amino acid in all four cheeses were observed (Fig. 4b).

6.4.4 The content of FFA in cheese lipid extracts did not significantly differ

There were no significant differences in the concentrations of any of 11 FFA between cheeses at Month 3 and Month 9. At both time points the concentration of C16 (palmitic) FFA was significantly higher in all cheeses compared to all other FFA. In addition, C18:1 (vaccenic), C18 (stearic) and C14 (myristic) were present in high concentration in all cheeses, but no significant differences in concentrations were observed.

6.4.5 Month 3 cheeses show higher differentiation in volatile profiles

In Month 3 cheeses, 48 volatiles that are considered as cheese flavour contributors were identified, 17 of which were present in significantly different (SD) abundances (Table 2). The ratio between the highest and the lowest value of abundances for a single compound among the four cheeses ranged between 1.85 for 3-methyl-3-butan-1-ol to >5000 for benzoic acid. The SD compounds were present in higher abundances in test cheeses compared to the Control. The exceptions are butan-2-one, and 3-methyl-2-buten-1-ol, the latter being in significantly higher abundances in the Control compared to the test cheeses. Cheese DPC4206 had the highest abundances of 2-decenal and pentadecan-2-one and significantly higher abundance 3-methyl-3-buten-1-ol compared to all other cheeses. Cheese DPC4536 had significantly higher abundances of ethyl octanoate, ethyl decanoate, octan-1-ol, benzoic acid and 2-undecenal compared to all other cheeses. Cheese DPC2071 was characterised by significantly lower abundances of octanal and 3-methyl-2-buten-1-ol compared to all other cheeses (LSD test, data not shown).

In Month 9 cheeses, 40 volatiles that are considered as cheese flavour contributors were identified, 8 of which were present in SD abundances in cheeses (Table 2). The

ratio between the highest and the lowest value of abundances for a single compound among the four cheeses ranged between 2.31 for ethyl hexanoate to >36000 for propanoic acid. The results of the Least significant test (LSD) showed that the majority of compounds present in SD abundances were present at the highest concentrations in cheese DPC4536, and for ethyl hexanoate, ethyl decanoate, and octan-1-ol they were present in significantly higher concentrations compared to all other cheese (LSD test, data not shown).

The PCA plot for Month 3 cheeses based on the abundances of all identified flavour contributors is presented in Figure 5. The first two axes described 82 % of the total variability among cheeses, with dimension 1 (PC1), describing 54 % of variability and dimension 2 (PC2) described 28 % of variability. Cheeses were discriminated mainly in PC1, while Control cheese and cheese DPC2071 were discriminated between themselves in PC2. The position of the Control cheese was determined by butan-2-one, carbon-disulfide (CDS), 3-hydroxy-butan-2-one (acetoin), pentan-1-ol, dimethyl-disulfide (DMDS), dimethyl-trisulfide (DMTS), octanal, dimethyl-sulfone, decanal, propan-1-ol, 3-methyl-2-buten-1-ol, heptanal, acetic acid and D-limonene. The position of cheese DPC2071 was determined by butanoic acid, ethyl butanoate, dimethyl-sulfone and dimethyl-sulfide (DMS). Cheese DPC4206 was positioned according to the abundances of 2,3-pentanedione, DMS, hexanoic acid, ethyl acetate, nonan-2-one, heptan-2-one, 3-methyl-3-buten-1-ol, while the position of DPC4536 was determined by the abundance of benzoic acid, 3-methyl-butanal, octanoic acid, 2,3-butanedione, ethyl hexanoate, decanoic acid, γ- and δ- dodecanolactone, nonanal, undecane, dodecanal and pentadecan-2-one (Fig. 5).

The PCA plot for Month 9 cheeses is presented in Figure 6. The first two axes described 75 % of the total variability among cheeses, with dimension 1 (PC1),

describing 45 % of variability and dimension 2 (PC2) describing 30 % of variability. Cheeses were discriminated mainly in PC1, while cheese DPC2071 and DPC4206 were discriminated in PC2. The position of the Control cheese was defined by 2,3-butanedione, benzeneacetaldehyde, CDS, butan-2-one, pentan-2-one, octanal, nonan-2-one, nonanal, δ-octalactone and δ-decalactone. Cheese DPC2071 was positioned according to the abundances of propan-1-ol, propanoic acid, butan-2-one, 3-methyl-butanal, 2,3-pentanedione, butanoic acid, heptan-2-one, undecane and decanal. The position of cheese DPC4206 cheese was determined by the abundances of 3-methyl-3-buten-1-ol, DMS, acetic acid, 3-hydroxy-butan-2-one, while the position of cheese DPC4536 was correlated with the abundances of ethyl butanoate, ethyl hexanoate, ethyl octanoate, ethyl decanoate, octan-1-ol, octanoic acid, ethyl acetate, DMDS, 3-methyl-2-buten-1-ol, dimethyl-sulfone, hexanoic acid and DMTS (Fig. 6).

6.4.6 Sensory analysis showed minimal differences in the sensory attributes of cheeses

In the SAE, cheese DPC2071 showed significantly higher score for liking of flavour and overall acceptability, compared to cheese DPC4536 (Fig. 7). All other treatments and variables were not significantly different.

Figure 8 displays the data for the RDA. Statistical analysis showed that cheese DPC4536 had a significantly higher score for pasty texture compared to cheese DPC2071, and significantly higher scores for off-flavour compared to cheeses DPC2071 and DPC4206. Cheese DPC2071 had significantly higher score for Cheddar flavour compared to cheese DPC4536. All other treatments and variables were not significantly different.

6.5 Discussion

In bacterial ripened cheeses, such as Cheddar, the dynamic evolution of both the starter and non-starter microbial populations depends on the environmental conditions and available nutrients during the manufacture and ripening thus affecting acidification, biochemical transformation of substrates and flavour development (Bautista-Gallego et al., 2014). Unlike the starter bacteria, whose viability is rapidly reduced in the initial phases of ripening, non-starter bacteria slowly increase in numbers using various available molecules to become the dominant microbiota in cheese (Gatti et al., 2014). In this study, cell enumerations of the starter follow the general trends observed during Cheddar cheese ripening (Fox et al., 1998, Sousa et al., 2001, Settanni and Moschetti, 2010). The starter culture was inoculated into the vats at a high level which was maintained until Day 28, after which starter numbers decreased, probably due to the utilisation of most of the lactose (Crow et al., 2002). Conversely, the numbers of adjuncts inoculated in the three test cheeses reached peaks at Months 3 and 6. In the Control cheese, with no adjunct added, NSLAB started thriving after Day 28, similarly as reported previously (Fox et al., 1998, Steele et al., 2006).

To evaluate the persistence and the dominance of the added adjuncts in relation to naturally-present or contaminating flora in the cheese, PFGE profiles were obtained over the ripening period. The PFGE patterns showed that the dominating profiles corresponded to the inoculated *L. paracasei* strains in each vat (Fig. 2). At Month 3, PFGE patterns not representing the adjunct strains were detected in the Control cheese. They corresponded to other contaminating microbiota originating from the environment, personnel or the pasteurised milk. However, this microbiota was not detected in subsequent sampling and the three adjunct strains became dominant in

the Control cheese, probably due to the cross-contamination between the vats during cheese manufacture. This means that the Control cheese was not a real control of the experiment, since non-starter flora that developed by the end of ripening predominantly corresponded to the adjuncts that were used in the test cheeses, thus influencing differentiation between the Control cheese and three test cheeses.

The gross composition of the manufactured cheeses was determined at Day 14. No significant difference in any of the parameters among the four cheeses was observed, indicating that inoculated adjuncts had no effect on the cheese composition. Gilles and Lawrence (1973) proposed a grading system of cheese quality according to the values of cheese composition indices, such as salt in moisture (S/M), fat in dry matter (FDM), moisture in non-fat solids (MNFS) and pH. In terms of the overall quality of the cheeses produced in this study, the only parameter deviating from 'premium grade' is FDM, which in our case is approx. 48 %. In the case of premium grade cheeses FDM values are typically between 52 and 55 %, while values of 50-56 % are typical in graded cheeses. Since lipolysis in Cheddar cheese is not extensive, the fat content plays a minor role in determining cheese quality, and FDM values from a relatively wide range are acceptable. However, if FDM value is lower than 48 %, it is highly possible that cheeses will be more firm and have less acceptable flavour at the end of ripening (Fox et al., 2004). Taking into account all of the parameters, cheeses produced in this study were of satisfactory quality.

Primary proteolysis in cheese refers to the degradation of casein into large polypeptides (Gobbetti et al., 2007). The level of primary proteolysis can be expressed through various calculations (McSweeney and Fox, 1997) and the most common way is to calculate the soluble nitrogen in cheese extracts at a pH of 4.6 expressed as a percentage of total nitrogen (pH4.6SN/TN (%)). In this study, the

levels of pH4.6SN/TN (%) significantly increased during ripening, and the results obtained correspond to the usual trends observed in Cheddar (Lane and Fox, 1996). However, no significant differences among the cheeses were observed (Fig. 3a). These results confirm that adjunct (or NSLAB) bacteria have a minimal effect on primary proteolysis in cheese, as shown by Lane and Fox (1996). In cheeses where *L. paracasei* strains are added primarily for their probiotic effect, they showed only a minor impact on proteolysis (Bergamini et al., 2006). Similarly, Bielecka and Cichosz (2017) showed that addition of *L. paracasei* LPC-37 during ripening of a Dutch-type cheese did not significantly affect proteolysis and peptidolysis. In contrast, Ong et al. (2007b) showed that after 20 weeks of ripening, Cheddar cheeses with added *L. casei* and *L. paracasei* had significantly higher values of water-soluble nitrogen (WSN) compared to the control cheese and cheeses with other probiotic adjuncts. This confirms that contribution of added *L. paracasei* strains to proteolysis, whether they were used as flavour adjuncts or probiotics, is strain-specific. In the present study, where three adjuncts belonging to the same species were compared, no differences were observed; however, if different starter culture, cheese technology and ripening conditions were used, they could potentially influence the individual performances of the three adjuncts.

The degradation of large polypeptides to shorter peptides and free amino acids by proteinases and peptidases is considered as secondary proteolysis in cheese (Gobbetti et al., 2007). As expected, the level of total FAA in the cheeses increased over the ripening time, as the proteolysis advanced. However, statistical differences in the concentrations of individual amino acids among the cheeses were observed only at the early stage of ripening (Day 14). Similar results for control and adjunct added cheeses were reported by Lane and Fox (1996), where it was shown that the

peptidase systems of added lactobacilli contributed much less to the release of free amino acids than starter peptidases, that are capable of degrading a wide range of medium and small peptides to amino acids. Although we have selected the adjunct strains for this study according to the variable activities of the enzymes of the proteolytic cascade and their contribution to proteolysis (Stefanovic et al., 2017a), we did not observe any direct or synergistic effect of the inoculated adjuncts towards secondary proteolysis. This is in contrast to findings of Lynch et al. (1999), who showed that the concentration of total amino acids were similar in control and test cheeses up to 3 months of ripening, but then higher levels of total amino acids developed in cheeses with added adjuncts.

A direct correlation between the concentration of FAA and cheese flavour cannot be made, since different types of cheeses have similar relative proportions of amino acids, but have distinctly different flavour (Sousa et al., 2001). In addition, as metabolism of amino acids seems to be strain-specific, a similar pool of FAA will be converted to different volatiles by different strains (Peralta et al., 2016). In this study, a very limited differentiation of cheeses according to the FAA levels was observed, mainly in the early stage of ripening. This was also reflected in GC volatile profiles of cheeses, where significant differences in concentrations of only several FAA-derived volatiles were observed in Month 3 samples. The examples are benzoic acid, originating from phenylalanine, that has rosy, honey-like aroma, and branched-chain alcohols, 3-methyl-3-buten-1-ol and 3-methyl-2-buten-1-ol, that most probably originated from leucine metabolism (Urbach, 1995, Bintsis and Robinson, 2004) or from grass used to feed the cows and were present in milk (Mariaca et al., 2001, Di Cagno et al., 2003), and are known for their cheese, fruity notes (Curioni and Bosset, 2002).

Cheddar belongs to the group of cheeses with moderate levels of lipolysis (Collins et al., 2003). In Cheddar, starter cultures are observed as the main FFA producers, while the contribution of non-starter appears to be minimal (Hickey et al., 2006). Free fatty acids (C4-C12), lactones and methyl-ketones are important flavour contributors with low threshold points (Collins et al., 2003, O'Mahony et al., 2005). The lipolytic activity of both starter and adjunct cultures used in this study were confirmed in a quantitative assay with 4-nitrophenyl-dodecanoate (data not shown), although the network of lipolytic reactions occurring in cheese is much more complex and involves numerous specific and non-specific enzymes. The analysis of FFA content at Month 3 and Month 9 confirmed minimal contribution of adjunct cultures to lipolysis in Cheddar. As expected, palmitic acid was present in the highest levels of all FFA across all the cheeses, followed by stearic and vaccenic acids, since it is known that C16 and C18 acids dominate in bovine milk triglycerides (Collins et al., 2003).

The differences observed in the volatile profiles of the analysed cheeses occurred, apart from several FAA-derived compounds (reported above), mainly from lipolysis-driven compounds. Since no significant differences in FFA contents were observed, it can be implied that the main lipolytic reactions arose from starter activity in the vat during cheese manufacture (Hickey et al., 2006), but the further development of flavour contributing compounds came from the adjunct metabolism of some of the primarily developed metabolites. This metabolic activity was dominant in cheese DPC4536 and partially in cheese DPC4206. In Month 3 cheeses, numerous long-chain aldehydes, acids and lactones were present in the volatile profile of cheese DPC4536, and to a lesser extent in cheese DPC4206 and the Control cheese. Octanal and nonanal have green, fatty aroma (Curioni and Bosset, 2002) while 2-decenal and

2-undecenal are characterised by green grass-like and herbaceous aromas (Verzera et al., 2004, Ziino et al., 2005). Decanoic acid has stale butter flavour (Curioni and Bosset, 2002). Lactones, such as δ-octalactone, δ-decalactone and δ-dodecalactone are mainly contributing to the coconut, fruity notes, similarly as ethyl esters (octanoate, decanoate) (Curioni and Bosset, 2002). In Month 9 cheeses, only a few compounds contributed to the statistically significant differentiation of cheeses, and the majority of them were of fat origin (ethyl hexanoate, octanoate and decanoate, δ-decalactone, octan-1-ol, octanal). Apart from these, propanoic acid and propan-1-ol (both having pungent aroma (Singh et al., 2003)) were dominant in Month 9 DPC2071 cheese, and they most probably originated from the metabolism of lactate, FAA (e.g. threonine) or from the degradation of long-chain metabolites, as all lactose (which is considered to be the main source of propanoic acid) would be exploited in such late stage of ripening.

The differentiation of cheeses based on volatile profiles was confirmed in PCA plots and the cheeses were more differentiated in the earlier stage of ripening. Apart from benzoic acid, the main variables leading to the differentiation were the afore mentioned lipid metabolites. However, although the ratio between the highest and the lowest abundance detected in cheeses for some compounds was considerably high (>2000), this was a consequence of the complete absence of these volatiles in the profiles of some cheeses, and their presence in the profiles of other cheeses. For example, benzoic acid was detected in the Control cheese and cheeses DPC4206 and DPC4536 in Month 3 samples, but was completely absent in cheese DPC2071, which contributed to the ratio of >5000. Similarly, propanoic acid was detected in all three test cheeses, but was not present in the volatile profile of the Control cheese in Month 9 samples. These compounds were, in a statistical sense, a factor of

differentiation; however, their realistic contribution was much less important, since the peaks corresponding to these compounds would not be among the highest ones in the generated chromatograms. In addition, the majority of flavour compounds were detected in similar abundances ($p>0.05$) among the cheeses at both time points, and although a certain level of differentiation existed, cheeses were still highly similar in terms of total aroma profiles.

Sensorial analysis performed at Month 9 of ripening confirmed a high degree of similarity among the cheeses, and scores for only several attributes significantly differed, while for the majority of them, no significant differences were observed. In general, cheese DPC4536 had the least favourable organoleptic characteristics, while cheese DPC2071 showed the highest scores for Cheddar flavour, liking of flavour and overall acceptability. These findings can be linked to volatile analysis. In cheese DPC4536, lipid metabolites, such as lactones and long-chain acid esters were detected at the highest concentrations. These compounds are flavour contributors, but if they are present in high enough concentrations they can be perceived as off-flavours (Marsili, 2011). On the other hand, cheese DPC2071 was characterised mainly by moderate values of volatiles, especially ones occurring from lipolytic processes, and high values of some important flavour compounds, such as short- and medium-chain aldehydes, ketones and acids. Most probably, the optimal balance between the aroma compounds led to determination of cheese DPC2071 as the one with somewhat better organoleptic characteristics.

The previous characterisation of strains that were used as adjuncts showed considerable differences in enzyme activities and in volatile profiles obtained in cheese model systems. Based on those results, it was expected that the appropriate level of differentiation would be observed in cheese application. Nevertheless, the

differences were only minor, and cheeses were highly similar in terms of their flavour characteristics. The reasons for these observations are not clear. Potentially, the increase in the complexity of the environment caused minimisation of metabolic diversity of adjuncts observed in previous characterisation assays, or the metabolic activity of the starter was sufficient to mask the moderate differences that arose from the adjuncts activities. Additionally, the development of NSLAB flora in the Control cheese that corresponded mainly to the adjuncts used in tested cheeses could have diminished the differences in both volatile profiles and flavour differentiation between the Control and tested cheeses, especially in later stages of ripening. Only minor differences in cheese volatiles abundances among the three tested cheeses, mainly in earlier stages of ripening, could suggest that these adjuncts have better potential for flavour diversification in shorter ripened cheeses, but these assumptions should be confirmed in sensorial panel.

6.6 Conclusion

The influence of three adjunct *Lactobacillus paracasei* strains in flavour development of Cheddar cheese was assessed. The adjunct strains did not show an impact on gross composition, nor did they influence primary and secondary proteolysis or lipolysis. Volatile analysis at Month 3 showed that the differences in volatiles among the four cheeses were caused mainly by the variation in long-chain aldehydes, acids and esters that originated from the metabolism of FFA. On the other hand, flavour compounds originating from FAA metabolism showed only slight variation. In Month 9 cheeses, differentiation in volatile profiles was much less evident. Sensorial analysis confirmed a high degree of similarity among the cheeses and showed that in general cheese DPC2071 had slightly better organoleptic characteristics. Even though the strains showed considerable diversity based on genomic profiling, enzyme activities and metabolic capacities in cheese model systems, this was not entirely reflected in pilot scale production of Cheddar cheese.

6.7 Acknowledgments

Many thanks to Dr Diarmuid Sheehan for his input in cheese trial design, to Dr Clara Roces Rodriguez for participation in cheese manufacture and analysis and to Prof. Maurice O'Sullivan for organising and conducting sensorial analysis of cheeses.

6.8 References

Antonsson, M., Molin, G. & Ardo, Y. 2003. *Lactobacillus* strains isolated from Danbo cheese as adjunct cultures in a cheese model system. *Int J Food Microbiol,* 85**,** 159-169.

Ardo, Y. 2006. Flavour formation by amino acid catabolism. *Biotechnol Adv,* 24**,** 238-242.

Bautista-Gallego, J., Alessandria, V., Fontana, M., Bisotti, S., Taricco, S., Dolci, P., Cocolin, L. & Rantsiou, K. 2014. Diversity and functional characterization of *Lactobacillus* spp. isolated throughout the ripening of a hard cheese. *Int J Food Microbiol,* 181**,** 60-66.

Bergamini, C. V., Hynes, E. R. & Zalazar, C. A. 2006. Influence of probiotic bacteria on the proteolysis profile of a semi-hard cheese. *Int Dairy J,* 16**,** 856-866.

Bielecka, M. M. & Cichosz, G. 2017. The influence of an adjunct culture of *Lactobacillus paracasei* LPC-37 on the physicochemical properties of Dutch-type cheese during ripening. *LWT-Food Sci Technol,* 83**,** 95-100.

Bintsis, T. & Robinson, R. K. 2004. A study of the effects of adjunct cultures on the aroma compounds of Feta-type cheese. *Food Chem*, 88**,** 435-441.

Burns, P., Cuffia, F., Milesi, M., Vinderola, G., Meinardi, C., Sabbag, N. & Hynes, E. 2012. Technological and probiotic role of adjunct cultures of non-starter lactobacilli in soft cheeses. *Food Microbiol,* 30**,** 45-50.

Collins, Y. F., McSweeney, P. L. H. & Wilkinson, M. G. 2003. Lipolysis and free fatty acid catabolism in cheese: a review of current knowledge. *Int Dairy J,* 13**,** 841-866.

Crow, V., Curry, B., Christison, M. & Hellier, K. 2002. Raw milk flora and NSLAB as adjuncts. *Aust J Dairy Technol,* 57**,** 99.

Crow, V., Curry, B. & Hayes, M. 2001. The ecology of non-starter lactic acid bacteria (NSLAB) and their use as adjuncts in New Zealand Cheddar. *Int Dairy J,* 11**,** 275-283.

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.

Dairou, V. & Sieffermann, J. M. 2002. A comparison of 14 jams characterized by conventional profile and a quick original method, the flash profile. *J Food Sci,* 67**,** 826-834.

De Jong, C. & Badings, H. T. 1990. Determination of free fatty acids in milk and cheese procedures for extraction, clean up, and capillary gas chromatographic analysis. *J Sep Sci,* 13**,** 94-98.

Di Cagno, R., Banks, J., Sheehan, L., Fox, P. F., Brechany, E. Y., Corsetti, A. & Gobbetti, M. 2003. Comparison of the microbiological, compositional, biochemical, volatile profile and sensory characteristics of three Italian PDO ewes' milk cheeses. *Int Dairy J,* 13**,** 961-972.

Di Cagno, R., Quinto, M., Corsetti, A., Minervini, F. & Gobbetti, M. 2006. Assessing the proteolytic and lipolytic activities of single strains of mesophilic lactobacilli as adjunct cultures using a Caciotta cheese model system. *Int Dairy J,* 16**,** 119-130.

El Soda, M., Madkor, S. A. & Tong, P. S. 2000. Adjunct cultures: recent developments and potential significance to the cheese industry. *J Dairy Sci,* 83**,** 609-19.

Fitzsimons, N. A., Cogan, T. M., Condon, S. & Beresford, T. 1999. Phenotypic and genotypic characterization of non-starter lactic acid bacteria in mature cheddar cheese. *Appl Environ Microbiol,* 65**,** 3418-26.

Fox, P., McSweeney, P. & Lynch, C. 1998. Significance of non-starter lactic acid bacteria in chedder cheese. *Aust J Dairy Technol,* 53**,** 83.

Fox, P. F., McSweeney, P. L., Cogan, T. M. & Guinee, T. P. 2004. *Cheese: chemistry, physics and microbiology: general aspects*, Academic Press.

Gatti, M., Bottari, B., Lazzi, C., Neviani, E. & Mucchetti, G. 2014. Invited review: Microbial evolution in raw-milk, long-ripened cheeses produced using undefined natural whey starters. *J Dairy Sci,* 97**,** 573-591.

Gilles, J. & Lawrence, R. 1973. The assessment of cheddar cheese quality by compositional analysis. *New Zeal J Dairy Sci Technol,* 8**,** 148-151.

Gobbetti, M., De Angelis, M., Di Cagno, R., Mancini, L. & Fox, P. F. 2015. Pros and cons for using non-starter lactic acid bacteria (NSLAB) as secondary/adjunct starters for cheese ripening. *Trends Food Sci Tech,* 45**,** 167-178.

Gobbetti, M., De Angelis, M., Di Cagno, R. & Rizzello, C. 2007. The relative contributions of starter cultures and non-starter bacteria to the flavour of cheese. *In*: Weimer, B.C (ed.). *Improving the flavor of cheese.* Taylor&Francis.

Hickey, D. K., Kilcawley, K. N., Beresford, T. P. & Wilkinson, M. G. 2006. Starter bacteria are the prime agents of lipolysis in cheddar cheese. *J Agric Food Chem,* 54**,** 8229-8235.

Hussain, M. A., Rouch, D. A. & Britz, M. L. 2009. Biochemistry of non-starter lactic acid bacteria isolate *Lactobacillus casei* GCRL163: Production of metabolites by stationary-phase cultures. *Int Dairy J,* 19**,** 12-21.

IDF 1979. International IDF Standard 88:1979: Cheese and Processed Cheese Products: Determination of the chloride content: Potentiometric titration method, IDF.

IDF 1982. International IDF Standard 4A:1982: Cheese and Processed Cheese : Determination of the Total Solids Content Reference Method, IDF.

IDF 1993. International IDF Standard 20B:1993: Milk: Determination of Nitrogen Content: Inspection by Attributes, IDF.

ISO 1988. Sensory analysis. General guidance for the design of test rooms. Ref. no. International Organization for Standardization, Genève, Switzerland, ISO 8589:1988.

Kieronczyk, A., Skeie, S., Langsrud, T. & Yvon, M. 2003. Cooperation between *Lactococcus lactis* and nonstarter lactobacilli in the formation of cheese aroma from amino acids. *Appl Environ Microbiol,* 69**,** 734-739.

Lane, C. & Fox, P. 1996. Contribution of starter and adjunct lactobacilli to proteolysis in Cheddar cheese during ripening. *Int Dairy J,* 6**,** 715-728.

Lynch, C., Muir, D., Banks, J., McSweeney, P. & Fox, P. 1999. Influence of Adjunct Cultures of *Lactobacillus paracasei* ssp. *paracasei* or *Lactobacillus plantarum* on Cheddar Cheese Ripening. *J Dairy Sci,* 82**,** 1618-1628.

Mariaca, R. G., Fernandez-Garcia, E., Mohedano, A. F. & Nunez, M. 2001. Volatile fraction of ewe's milk semi-hard cheese manufactured with and without the addition of a cysteine proteinase. *Food Sci Technol Int,* 7**,** 131-139.

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

Marsili, R. 2011. Flavors and off-flavors in dairy foods. *In*: Fuquay, J. W. (ed.) *Encyclopedia of Dairy Sciences (Second Edition).* Academic Press.

McDermott, A., Visentin, G., De Marchi, M., Berry, D., Fenelon, M., O'Connor, P., Kenny, O. & McParland, S. 2016. Prediction of individual milk proteins including free amino acids in bovine milk using mid-infrared spectroscopy and their correlations with milk processing characteristics. *J Dairy Sci,* 99**,** 3171-3182.

McSweeney, P. & Fox, P. 1997. Chemical methods for the characterization of proteolysis in cheese during ripening. *Lait,* 77**,** 41-76.

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

Milesi, M. M., Vinderola, G., Sabbag, N., Meinardi, C. A. & Hynes, E. 2009. Influence on cheese proteolysis and sensory characteristics of non-starter lactobacilli strains with probiotic potential. *Food Res Int,* 42**,** 1186-1196.

Milesi, M. M., Wolf, I. V., Bergamini, C. V. & Hynes, E. R. 2010. Two strains of nonstarter lactobacilli increased the production of flavor compounds in soft cheeses. *J Dairy Sci,* 93**,** 5020-5031.

O'Mahony, J. A., Sheehan, E. M., Delahunty, C. M. & McSweeney, P. L. H. 2005. Lipolysis and sensory characteristics of Cheddar cheeses ripened using different temperature-time treatments. *Lait,* 86**,** 59-72.

Ong, L., Henriksson, A. & Shah, N. P. 2007a. Chemical analysis and sensory evaluation of Cheddar cheese produced with *Lactobacillus acidophilus, Lb. casei, Lb. paracasei* or *Bifidobacterium* sp. *Int Dairy J,* 17**,** 937-945.

Ong, L., Henriksson, A. & Shah, N. P. 2007b. Proteolytic pattern and organic acid profiles of probiotic Cheddar cheese as influenced by probiotic strains of *Lactobacillus acidophilus, Lb. paracasei, Lb. casei* or *Bifidobacterium* sp. *Int Dairy J,* 17**,** 67-78.

Peralta, G. H., Wolf, I. V., Perotti, M. C., Bergamini, C. V. & Hynes, E. R. 2016. Formation of volatile compounds, peptidolysis and carbohydrate fermentation by mesophilic lactobacilli and streptoccocci cultures in a cheese extract. *Dairy Sci Technol,* 96**,** 603-621.

Poveda, J. M., Nieto-Arribas, P., Seseña, S., Chicón, R., Castro, L., Palop, L. & Cabezas, L. 2014. Volatile composition and improvement of the aroma of industrial Manchego cheese by using *Lactobacillus paracasei* subsp. *paracasei* as adjunct and other autochthonous strains as starters. *Eur Food Res Technol,* 238**,** 485-494.

Richter, V. B., de Almeida, T. C. A., Prudencio, S. H. & de Toledo Benassi, M. 2010. Proposing a ranking descriptive sensory method. *Food Qual Prefer,* 21**,** 611-620.

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

Singh, S. & Singh, R. 2014. Phenotypic and genotypic characterization of non-starter *Lactobacillus* species diversity in Indian Cheddar cheese. *LWT-Food Sci Technol,* 55**,** 415-420.

Singh, T., Drake, M. & Cadwallader, K. 2003. Flavor of Cheddar cheese: a chemical and sensory perspective. *Compr Rev Food Sci Food Saf,* 2**,** 166-189.

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

Steele, J. L., Budinich, M. F., Cai, H., Curtis, S. C. & Broadbent, J. R. 2006. Diversity and metabolic activity of *Lactobacillus casei* in ripening cheddar cheese. *Aust J Dairy Technol,* 61**,** 53.

Stefanovic, E., Kilcawley, K. N., Rea, M. C., Fitzgerald, G. F. & McAuliffe, O. 2017a. Genetic, enzymatic and metabolite profiling of the *Lactobacillus casei* group reveals strain biodiversity and potential applications for flavour diversification. *J Appl Microbiol,* 122**,** 1245-1261.

Stefanovic, E., Thierry, A., Maillard, M. B., Bertuzzi, A., Rea, M. C., Fitzgerald, G. F., McAuliffe, O. & Kilcawley, K. N. 2017b. Strains of the *Lactobacillus casei* group show diverse abilities for the production of flavor compounds in 2 model systems. *J Dairy Sci,* In Press

Stone, H., Bleibaum, R. N., & Thomas, H. A. (2012). Affective testing. *In* H. Stone, B. R.N., & H. A Thomas (eds.), *Sensory evaluation practices*. Elsevier Academic Press.

Stone, H., & Sidel, J. L. (2004). Affective testing. *In*: H. Stone and J. L. Sidel (eds.), *Sensory Evaluation Practices.* Elsevier Academic Press.

Van Hoorde, K., Van Leuven, I., Dirinck, P., Heyndrickx, M., Coudijzer, K., Vandamme, P. & Huys, G. 2010. Selection, application and monitoring of *Lactobacillus paracasei* strains as adjunct cultures in the production of Gouda-type cheeses. *Int J Food Microbiol,* 144**,** 226-235.

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

Verzera, A., Ziino, M., Condurso, C., Romeo, V. & Zappala, M. 2004. Solid-phase microextraction and gas chromatography-mass spectrometry for rapid characterisation of semi-hard cheeses. *Anal Bioanal Chem,* 380**,** 930-936.

Wijesundera, C., Roberts, M. & Limsowtin, G. 1997. Flavour development in aseptic cheese curd slurries prepared with single-strain starter bacteria in the presence and absence of adjuncts. *Lait,* 77**,** 121-131.

Wouters, J. T. M., Ayad, E. H. E., Hugenholtz, J. & Smit, G. 2002. Microbes from raw milk for fermented dairy products. *Int Dairy J,* 12**,** 91-109.

Yvon, M. 2006. Key enzymes for flavour formation by lactic acid bacteria. *Aust J Dairy Technol,* 61**,** 88-96.

Ziino, M., Condurso, C., Romeo, V., Giuffrida, D. & Verzera, A. 2005. Characterization of "Provola dei Nebrodi", a typical Sicilian cheese, by volatiles analysis using SPME-GC/MS. *Int Dairy J,* 15**,** 585-593.

Table 1: The composition of manufactured cheeses, at Day 14 of ripening.

Compositional indices	Control cheese	DPC2071 cheese	DPC4206 cheese	DPC4536 cheese
Moisture (%)	38.06	37.23	36.97	37.36
Salt (%)	1.79	1.76	1.79	1.78
pH	5.10	5.06	5.04	5.08
Fat (%)	29.72	30.19	30.36	30.24
Salt in moisture (%)	4.70	4.73	4.85	4.77
Fat in dry matter (%)	47.97	48.09	48.16	48.27
Moisture in non-fat solids (%)	54.14	53.33	53.08	53.55

Table 2: Compounds identified in Cheddar cheeses in two time points of ripening: Month 3 and Month 9 along with LRI used for compounds identification. If the abundances of compound showed significant differences among cheeses ($p<0.05$) the ratio between the maximal and minimal abundance of a compound between the cheeses was calculated.

Chemical group		LRI	Month 3		Month 9	
alcohol						
	Propan-1-ol	548	+		+	4.1
	3-Methyl-3-buten-1-ol	728	+	1.85	+	
	Pentan-1-ol	766	+		+	
	3-Methyl-2-buten-1-ol	772	+	2.46	+	
	Octan-1-ol	1069	+	1028	+	17.6
aldehyde						
	3-Methyl-butanal	654	+		+	
	Heptanal	903	+			
	Octanal	1003	+	668	+	1400
	Benzeneacetaldehyde	1049	+		+	
	Nonanal	1106	+	2.54	+	
	Decanal	1207	+		+	
	2-Decenal	1263	+	3416		
	2-Undecenal	1365	+	3628		
	Dodecanal	1410	+			
ketone						
	2,3-Butanedione (Diacetyl)	591	+		+	
	Butan-2-one	598	+	6.49	+	
	Pentan-2-one	684	+		+	
	2,3-Pentanedione	696	+		+	
	3-Hydroxy-butan-2-one (Acetoin)	732	+		+	
	Heptan-2-one	889	+		+	
	Nonan-2-one	1091	+		+	
	Pentadecan-2-one	1695	+	4774		
acid						
	Acetic acid	638	+	3.27	+	
	Propanoic acid	718			+	>36000
	Butanoic acid	792	+		+	
	Hexanoic acid	972	+		+	
	Benzoic acid	1155	+	5444		
	Octanoic acid	1158	+		+	
	Decanoic acid	1353	+	5.39		
sulfur						
	Dimethyl-sulfide (DMS)	518	+		+	
	Carbon disulfide (CDS)	537	+		+	
	Dimethyl-disulfide (DMDS)	743	+		+	
	Dimethyl-sulfone	921	+		+	
	Dimethyl-trisulfide (DMTS)	979	+		+	
ester						
	Ethyl acetate	613	+		+	
	Ethyl butanoate	799	+		+	
	Ethyl hexanoate	996	+		+	2.31
	Ethyl octanoate	1191	+	2.83	+	2.47
	δ-Octalactone	1289	+	2.18	+	
	Ethyl decanoate	1387	+	5.52	+	2.93
	δ-Decalactone	1503	+	3.81	+	2.95
	γ-Dodecalactone	1685	+			
	δ-Dodecalactone	1716	+	3375		
other						

Trichloromethane	623	+	+
2,5-Dimethyl-furan	707	+	+
Toluene	769	+	+
m-Xylene	873	+	+
D-Limonene	1035	+	+
Undecane	1099	+	+

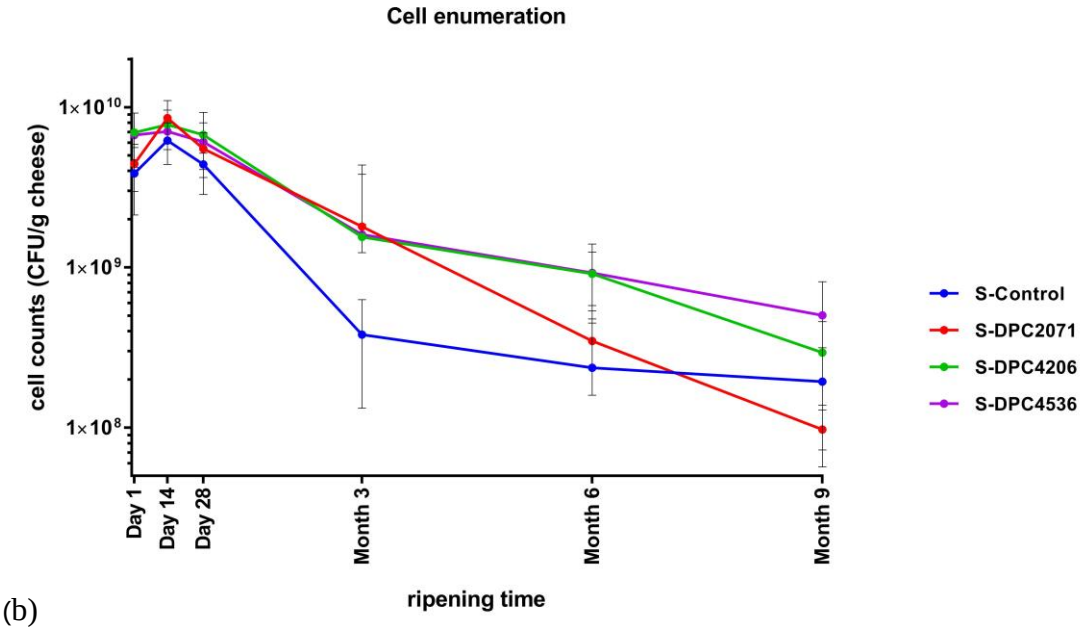
Table 3: Sensory terms used in Sensory Affective Evaluation of Cheddar cheeses.

Attribute	Definition	Scale
Appearance-Liking	The liking of appearance	0 = extremely dislike10 = extremely like
Flavour-Liking	The liking of flavour	0 = extremely dislike10 = extremely like
Aroma-Liking	The liking of aroma	0 = extremely dislike10 = extremely like
Texture-Liking	The liking of appearance	0 = extremely dislike10 = extremely like
Overall acceptability	The acceptability of the product	0 = extremely unacceptable10 = extremely acceptable

Table 4: Sensory terms used in Ranking Descriptive Analysis of Cheddar cheese

Attribute	Definition	Scale
Appearance-colour*	Appearance-Ivory to orange colour	0 = Ivory 10 = Orange
Creamy aroma	The smell associated with creamy/milky products	0 = none, 10 = extreme
Oxidised aroma	The smell associated with oxidised dairy products	0 = none, 10 = extreme
Barnyard aroma	The smell associated the farm, barnyard, ox tail	0 = none, 10 = extreme
Sweaty/sour aroma	The aromatics reminiscent of perspiration, foot odour. Sour, stale, slightly cheesy, moist, stained or odorous with sweat	0 = none, 10 = extreme
Firmness in the mouth	Firm texture in the mouth	0 = none, 10 = extreme
Crumbly	Crumbly texture in the mouth	0 = none, 10 = extreme
Pasty	Pasty texture in the mouth	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	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
Cheddar flavour	Intensity of cheddar cheese flavour	0 = none, 10 = extreme
Cream flavour	The flavour associated with creamy/milky products	0 = none, 10 = extreme
Dairy sweet flavour	The flavours associated with sweetened cultured dairy products such as fruit yoghurt	0 = none, 10 = extreme
Dairy fat flavour	Intensity of fat flavour	0 = none, 10 = extreme
Off-flavour	Off-flavour (Rancid)	0 = none, 10 = extreme
Oxidised flavour	The flavour associated with rancid or oxidised products	0 = none, 10 = extreme
Barnyard flavour	The flavour associated with the farm, barnyard, ox tail	0 = none, 10 = extreme
Fruity/Estery flavour	The flavours associated with fatty acid ethyl esters	0 = none, 10 = extreme

(a)



(b)

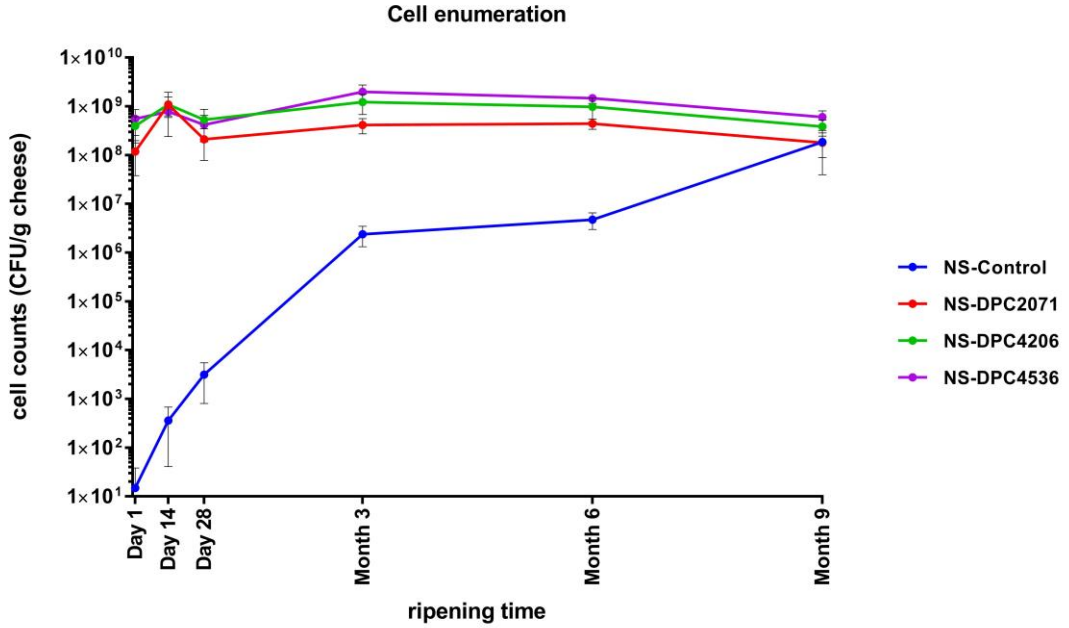
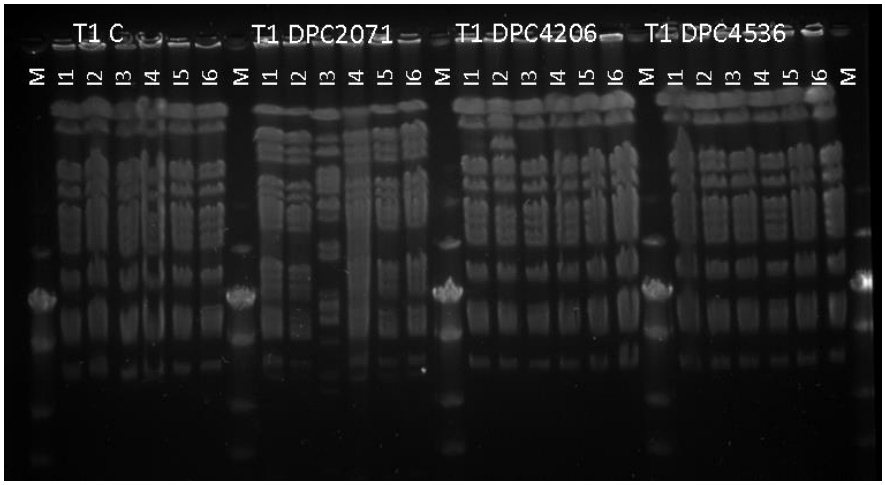


Figure 1: Enumeration of starter (a) and nonstarter (b) microbiota in cheeses during ripening. The values presented means obtained after enumeration of cells in cheeses of each of the three trials. Error bars present standard deviation.

(a)



(b)



(c)

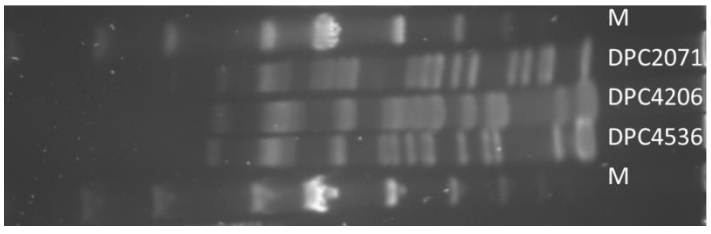
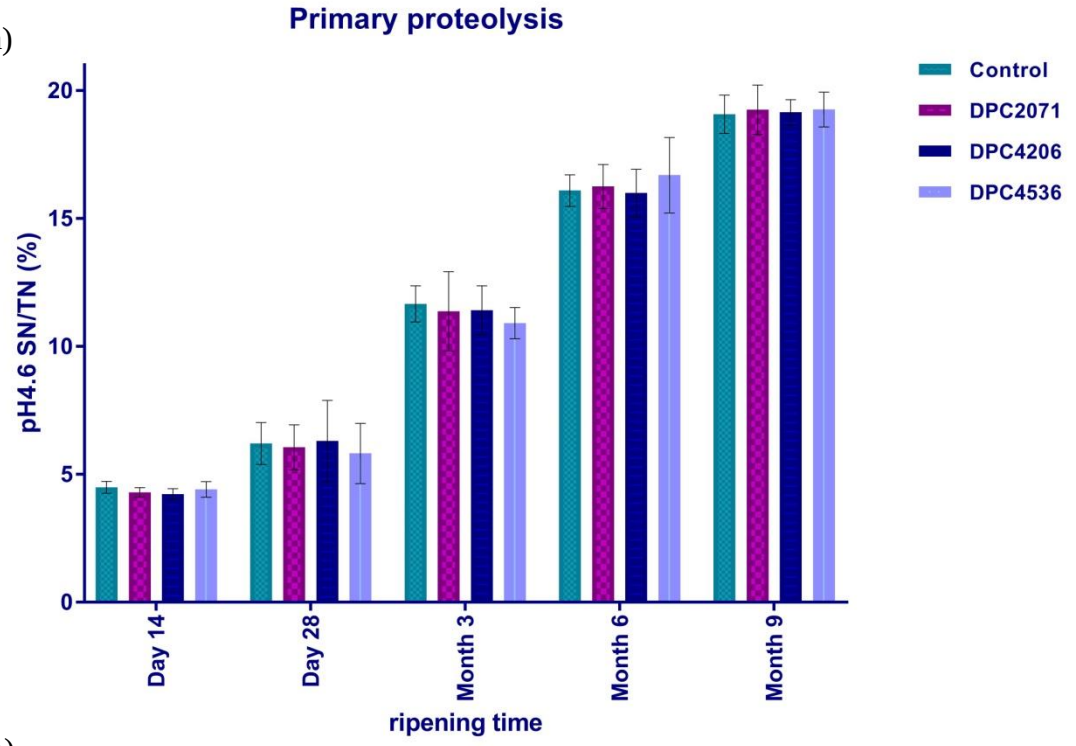


Figure 2: PFGE profiles of cheese (T1, Month 3 (a) and Month 9 (b)). Six isolates (I1-I6) from the highest dilution obtained in cell enumerations in each cheese were evaluated. Figure presents results for isolates obtained from cheeses manufactured in trial 1 (T1). For comparison, in (c), patterns of the three strains used as adjuncts are shown. M- Low range PFG marker, New England Biolabs.

(a)



(b)

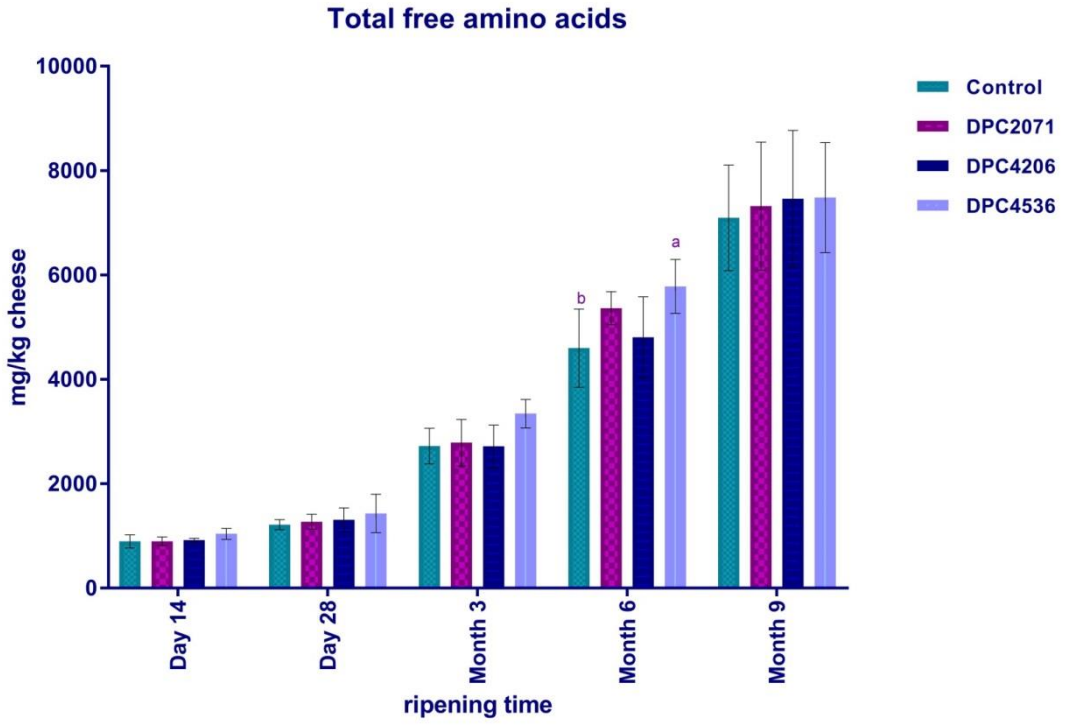
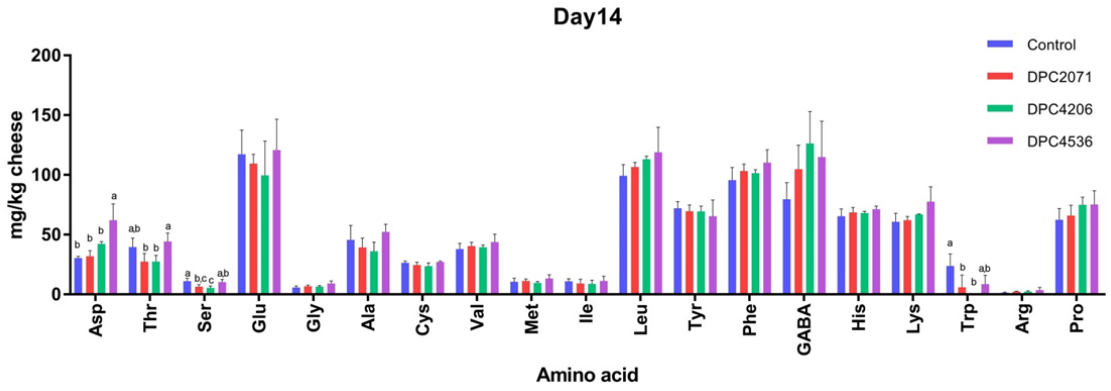


Figure 3: Primary proteolysis in cheese, calculated as soluble nitrogen at pH 4.6 expressed as a percentage of total nitrogen (pH4.6/TN (%)) (a), and secondary proteolysis, expressed as mg of total free amino acids per kg of cheese. Bars present mean of three values. Error bars present standard deviation. Letters (a, b) denote significant ($p<0.05$) differences observed among cheeses in a single time point.

(a)



(b)

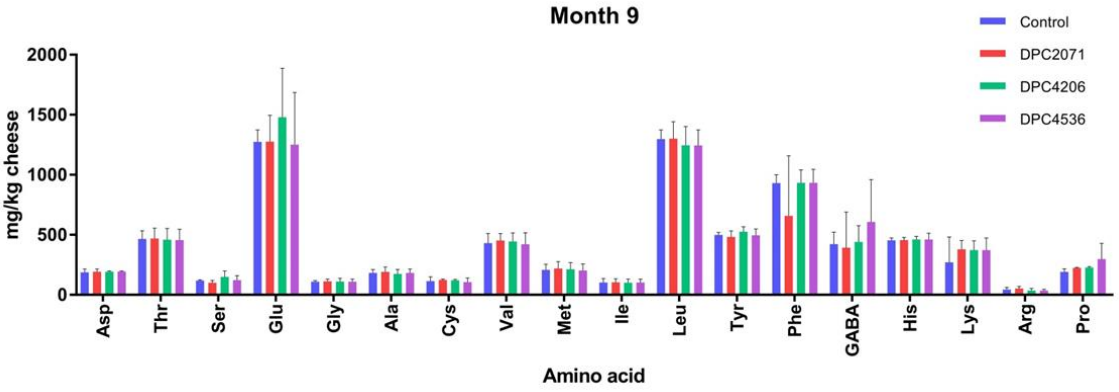
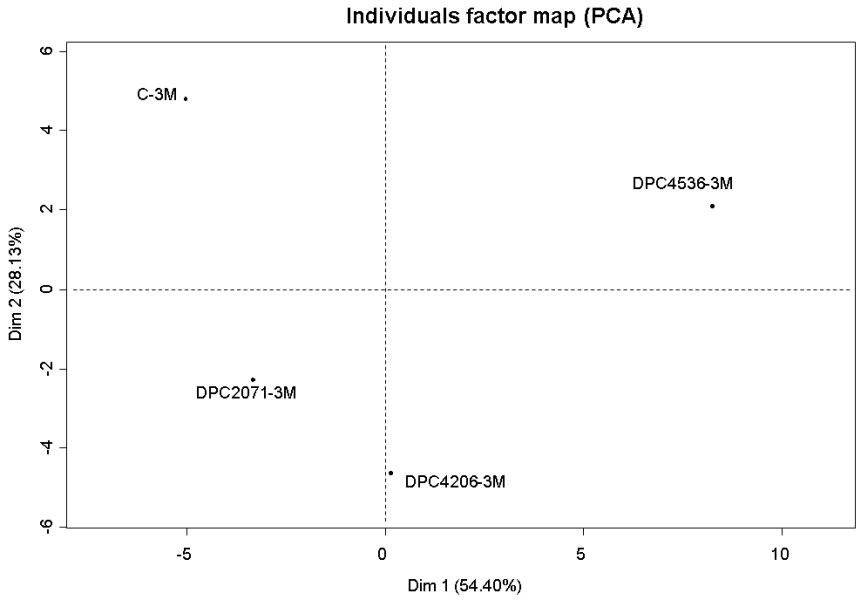


Figure 4: Free amino acids (mg/kg of cheese) determined in two time points (a) Day 14 of ripening and (b) Month 9 of ripening. Bars were labelled with different letters indicate significant differences ($p<0.05$). Error bars present standard deviation.

(a)



(b)

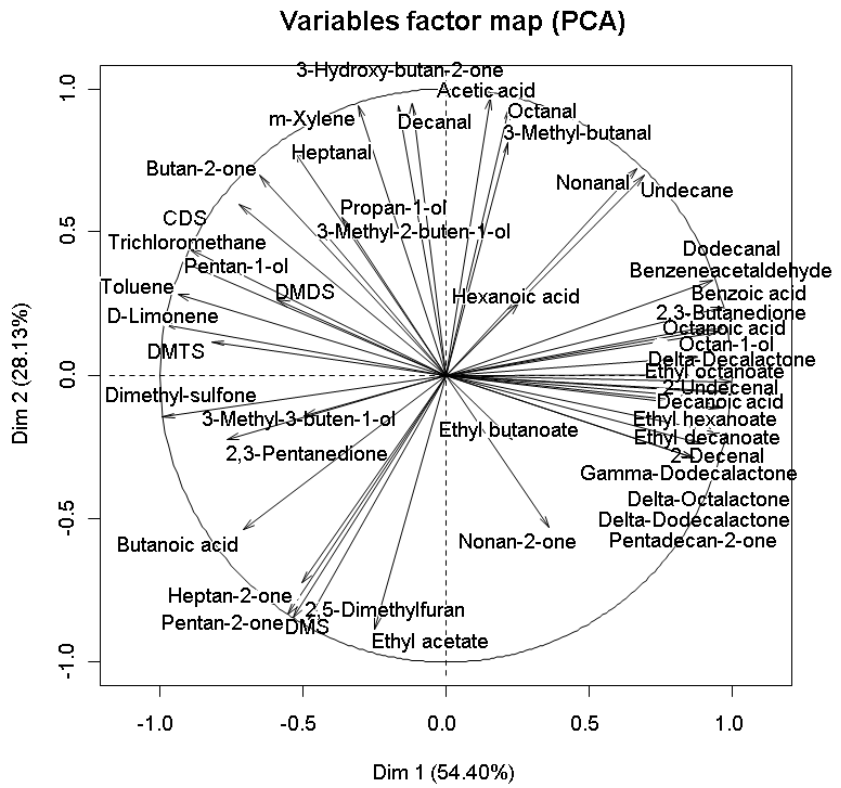
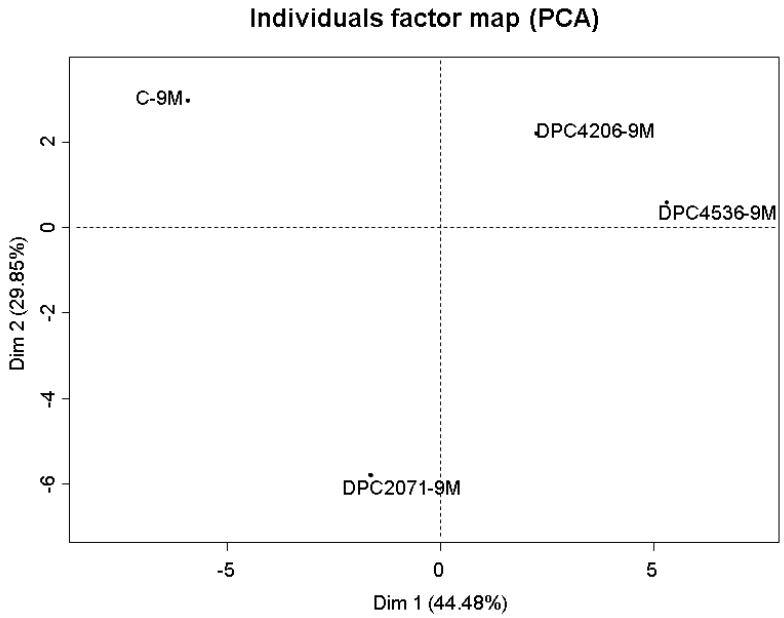


Figure 5: Individual factor map (a) and variable factor map (b) of principal component analysis (PCA) on 48 volatile compounds identified in Cheddar cheeses at Month 3 of ripening.

(a)



(b)

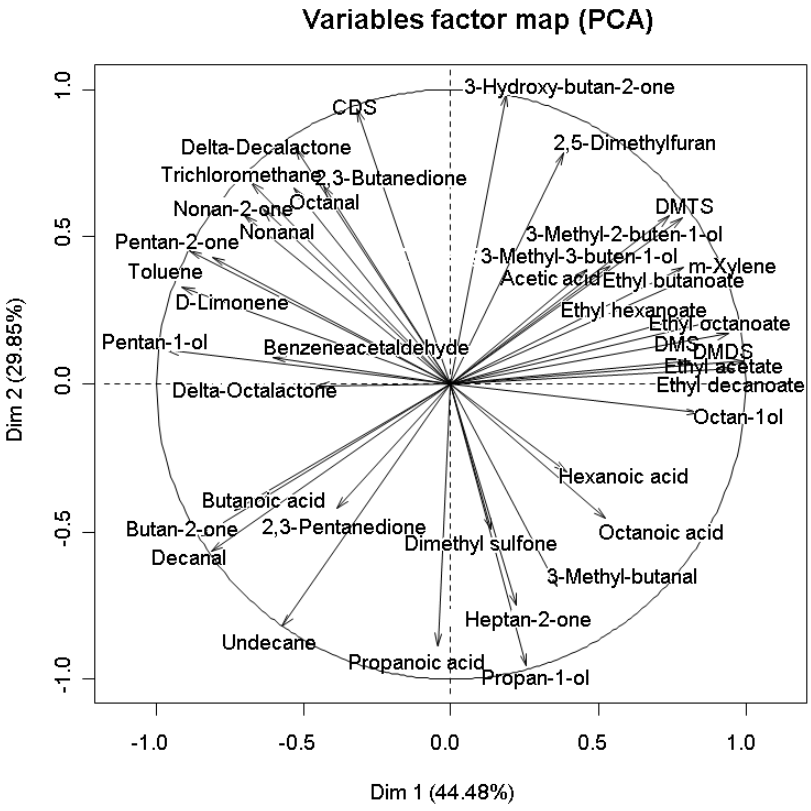


Figure 6: Individual factor map (a) and variable factor map (b) of principal component analysis (PCA) on 40 volatile compounds identified in Cheddar cheeses at Month 9 of ripening.

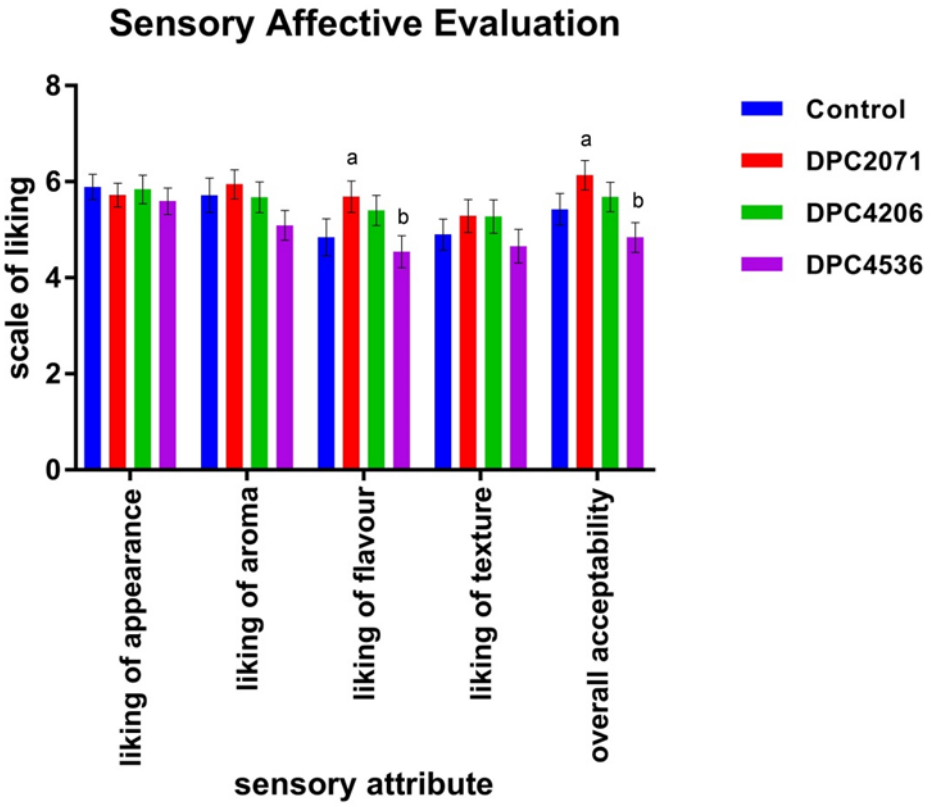
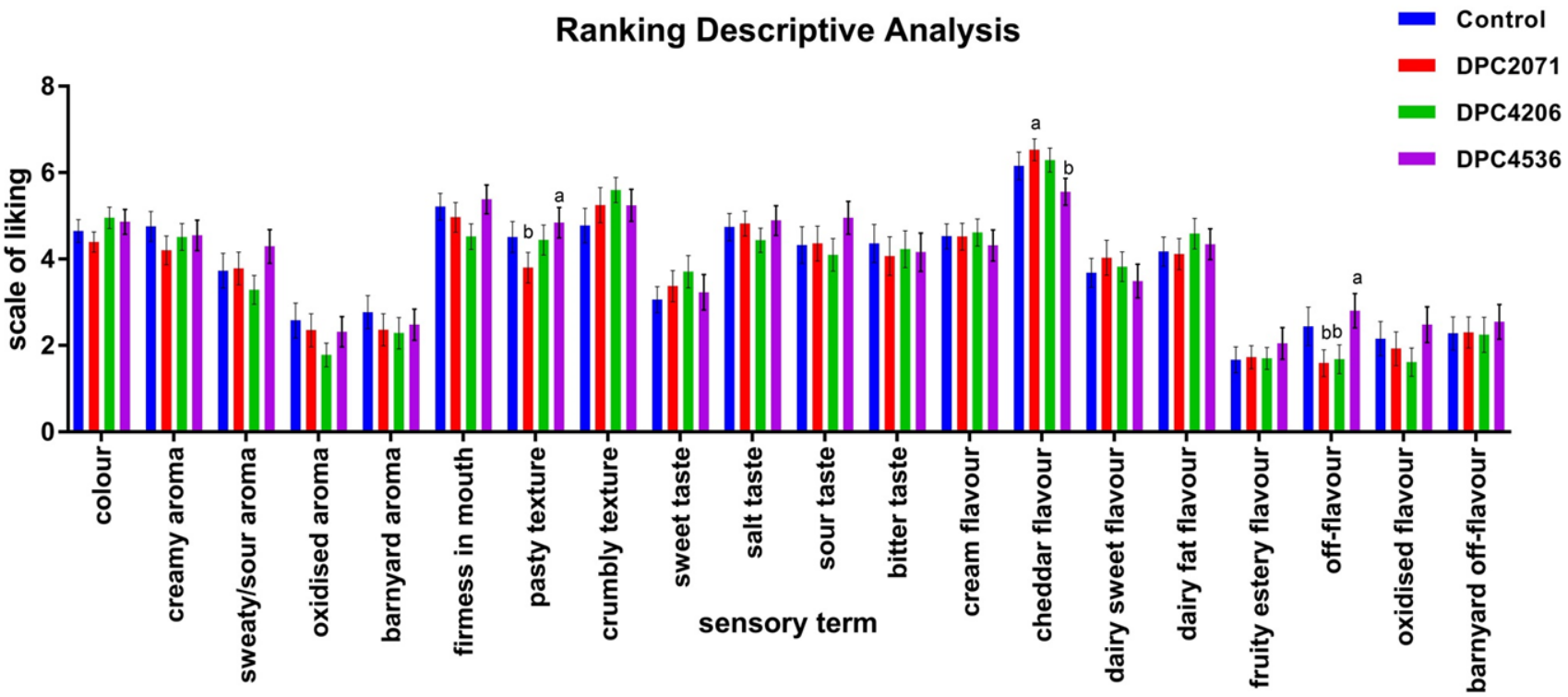


Figure 7: Figure 1. Sensory Affective Evaluation of Cheddar cheese. Bars present means of data from 42 assessors, and error bars present SEM (standard error of the mean). Letters (a, b) denote significant ($p<0.05$) differences. Scale of liking ranged from 0 (extremely dislike) to 10 (extremely like).

Figure 8: Ranking Descriptive Analysis of Cheddar Cheese. Bars present means of data from 42 assessors, and error bars present SEM (standard error of the mean). Letters (a, b and c) denote significant ($p<0.05$) differences. Scale of liking ranged from 0 (none) to 10 (extreme).



Chapter 7

General Discussion

7.1 General discussion

Lactobacillus paracasei is a lactic acid bacterium (LAB) successfully applied in numerous fields, such as the food industry, the biotechnological production of chemicals of interest and in health-related fields, where stains of this species have been extensively used as probiotics. Such a diverse spectrum of use confirms the extraordinary diversity and the ability of this species to successfully survive in niches with different environmental conditions (Cai et al., 2009). The available genome sequences present the basis for the analysis of the strains isolated from different habitats. The adaptation of *L. paracasei* to a broad range of habitats is facilitated by the evolution of their genomes, which tend to minimise through gene decay thus enabling niche specialisation, as has occurred in dairy strains. Conversely, isolates from the human and animal gut possess a diverse range of genes enabling these strains to survive and adapt to the constant changes in these habitats (Makarova et al., 2006).

The growing number of genome sequences of strains of the *L. casei* group has highlighted ongoing issues with the correct taxonomy of this group. Often, newly isolated and sequenced strains are designated as *L. casei*, although according to the current nomenclature rules (Tindall, 2008), strain ATCC 334 is the reference strain for *L. paracasei*, to which these isolates are more closely related than to ATCC 393, the reference strain for *L. casei.* In addition, ATCC 393 is more closely related to *L. rhamnosus* (Toh et al., 2013). Even though there is substantial evidence on discrepancies in the current nomenclature (Dellaglio et al., 2002), they were not acknowledged by the Judicial Commission of the International Committee on Systematics of Bacteria. However, new studies (Naser et al., 2007, Koirala et al.,

2015, Sardaro et al., 2016) continuously provide reports confirming that a consensus on nomenclature of the *L. casei* group is needed. Additionally, a thorough revision of the whole *Lactobacillus* genus taxonomy is suggested, possibly followed by its division into more homogenous genera (Salvetti and O'Toole, 2017).

One of the current aims of the fermented dairy industry is to develop products with novel flavour characteristics and to meet the consumers demand in the highly saturated market offer. One of the approaches is the use of diverse microorganisms in fermented food production, as flavour development mainly depends on microbial metabolism of the available substrates. Of the many different applications of *L. paracasei* strains, their potential to contribute to cheese flavour development has been recognised (Fox et al., 1998, Steele et al., 2006). This thesis focuses on the characterisation of strains of *L. paracasei* and their application in flavour development of Cheddar cheese. A bank of isolates was examined for their genetic diversity and phenotypic traits that contribute to the potential flavour diversification of dairy products. Furthermore, both practical application of the strains in cheese manufacture and the genomic background of their diversity were examined.

Chapter 2 describes the screening of isolates obtained from the Dairy Product Research Centre (DPC) Culture Collection. The strains were identified as belonging to the *L. casei* group based on 16S rRNA PCR, and as discussed above, they were designated as *L. paracasei* according to the current nomenclature rules (Tindall, 2008). The initial selection of strains was based on differences in the genomic structure profiles obtained by pulsed-filed gel electrophoresis (PFGE). This method is widely used for robust screening of strains, assuming that strains sharing the same pattern are identical. According to the PFGE profiles, a substantial level of diversity was observed, and 98 different patterns were detected among 252 strains designated

as *L. paracasei*. The phenotypic analysis included the determination of proteolytic enzymes activities since they are involved in the generation of metabolites that are seen as the most important cheese flavour contributors (Smit et al., 2005). The strains selected based on their differing PFGE profiles possessed variable activities, and it is envisaged that these could contribute to the variations in metabolite concentrations and subsequently to the flavour diversity. Nevertheless, when enzyme activities of two strains that showed the same PFGE pattern (DPC4206 and DPC4536) were compared, it was noticed that significant differences existed. This finding highlights that strains sharing the same PFGE pattern do not necessarily have the same phenotypic characteristics, and suggests the importance of gene expression and regulation and the limitation of a single technique, such as PFGE, to de-replicate collections of isolates.

The final stage of proteolysis, metabolism of free amino acids, has been recognised as the one having the highest contribution to flavour development in bacterial-ripened cheeses (Yvon and Rijnen, 2001, Ardo, 2006). In the continuation of strains screening, their ability to metabolise an amino acid present in higher concentration in the amino acid mix and to produce volatiles contributing to flavour was examined by GC-MS. An interesting finding was that even though some strains metabolised the amino acid present in abundance, the chromatograms of some of the strains (DPC2068) were more abundant in volatiles originating from specific amino acids (leucine) even in the environment with their lower concentration. The biotechnology of LAB presents a rapidly developing area and the examination of strains metabolic preferences in chemically-defined environment is the first step in guided microbial fermentations, where these organisms could be used in the production of desired molecules in the cost-effective yields.

Chapter 3 describes the potential of the strains selected according to the results obtained in Chapter 2 for flavour compound production. The estimation of a strains' ability to improve cheese flavour is best seen in an actual cheese production, which is often both time consuming and costly (Milesi et al., 2007). To overcome this, numerous model systems have been developed (Shakeel-Ur et al., 2001). Two model systems were used to examine the volatile production capacities of ten *L. paracasei* strains. In Model system 1 (MS1), which comprised a pancreatic digest of casein, the ability of strains to produce volatiles *via* amino acid catabolism was examined. The results illustrated the variation in volatiles produced by the strains, thus confirming the variability of *L. paracasei* metabolism reported in Chapter 2. On the other hand, Model system 2 (MS2), based on a processed curd, gave an indication of a strains behaviour in a cheese-like environment. In this model system, differentiation of strains was less obvious, and only one strain was recognised as slightly different. One of the reasons could be that the metabolism of additional substrates, such as lipids and lactose, contributed to a decreasing level of diversity. The strains' volatile profiles differed mainly in the amounts of produced compounds in each of the model systems. This was expected, as strains belonged to the same species, and the differences in volatiles abundances occurred because of the variations in the activities of the key flavour enzymes, as demonstrated in Chapter 2.

One of the most commonly used methods for determination of volatile compounds is gas chromatography coupled to mass spectrometry (GC-MS) (Soria et al., 2015). This method has found application in a plethora of analytical fields, including the food domain. Gas chromatography is preceded by the extraction of the volatiles, and numerous extraction techniques have been developed with the purpose of facilitating the isolation of the compounds of interest from the sample matrix. Chapter 4

illustrates the comparison of the results obtained when samples of MS1, described in Chapter 3, were analysed by two analytical methods based on different extraction techniques (HS-SPME and HS-Trap) and different types of GC columns (non-polar and polar).

The volatiles detected by the two methods were both qualitatively and quantitatively different. There are several factors that contributed to this observation, most important of which is the extraction step. Due to the different characteristics and affinities of sorbents, the surface available for the extraction and the conditions of extraction, preferable compounds were detected by both methods. In addition, the variation in the polarity of the columns used in the two methods contributed to differences in volatiles separation and detection. Some of the volatiles were exclusively detected by one method, while others were detected by both methods, but in variable abundances, mainly due to the differences in the affinity of the sorbent used for the extraction. In addition, two methods identified the same three strains as being the most distinct based on their volatile profiles, but their discrimination differed in the two methods, highlighting the impact of the analytical approach on the final results.

Since the conditions of analysis represent a significant factor in volatile detection, the reports on volatiles from different studies should be interpreted with caution. The development of guidelines based on the results obtained by different methods for different sample types could be helpful in standardising procedures for volatile analysis (Tait et al., 2014), and facilitate the comparison of the results obtained by different scientific teams.

Taking into the consideration all the results obtained in the previous chapters, three strains (DPC2071, DPC4206 and DPC4536) were selected and their genomes were sequenced and assembled. Further on, the specificities of their genome contents were assessed and presented in Chapter 5. As expected, a substantial level of diversity in genomic content was observed, especially when comparing DPC2071 on the one hand with DPC4206 and DPC4536 (that have identical PFGE profiles) on the other hand.

The genome of strain DPC2071 was characterised by the strikingly high number of plasmids, 11, compared to 0-4 usually present in lactobacilli (Douillard and De Vos, 2014). Additionally, some of the genes detected in this strain showed identity to species not often connected with the dairy niche, thus suggesting its previous habitats. This strain showed high potential for genetic content exchange and was involved in interaction with numerous species in its previous habitats. Based on the prevalent isolation sources of the interacting species, it could be suggested that in past, this strain probably inhabited a non-dairy niche, such as plant environment, at least for a short time during which it may have acquired genes from bacteria in its surroundings.

As expected, two strains that had the same PFGE pattern (DPC4206 and DPC4536) showed 99 % of genome identity, but still several important differences were noticed. While DPC4206 harboured a plasmid, DPC4536 had no plasmids. The other important difference was the loss of lactose utilisation ability in strain DPC4536, due to the lack of *lacG* gene encoding the first enzyme in lactose metabolic pathway. Based on the genomic analysis, it could be proposed that the two strains probably diverged from a common ancestor. This hypothesis could be confirmed by closer examination of their CRISPR arrays. The typical CRISPR array constitutes of string

of repeats that separate spacers, which are the "fingerprint" of previous phage attacks. Although the strains share the majority of the CRISPR spacers, they also possess specific ones, which testifies of their independent phage encounters, and thus independent existence. The genome comparison of these two strains illustrates that although PFGE is a good method for initial screening of strain diversity, it is not able to record smaller differences beyond the restriction patterns (Cai et al., 2007), confirming that identical PFGE patterns do not always imply identical strains.

In regard to the proteolytic pathway of flavour component development, the genomic comparison of the three strains did not reveal any genetic differences, except for the methionine metabolic pathway, where it was shown that strain DPC4206 possesses the highest number of homologs encoding for parts of sulfur-compounds generating pathways, which could be the determining point in flavour generation. However, in Chapter 2, differences in activities of all enzymes of the proteolytic cascade for the three strains were illustrated. This means that the different potential for flavour compound development of the three strains most probably comes as the consequence of different activities of the key enzymes and/or their regulation, such as the impact of coenzymes, and not due to the different number of key enzyme encoding homologs.

In all of the three analysed genomes, the sole presence of genes did not always secure the phenotype expression under the experimental conditions. Most likely, the regulation of these features led to non-observable phenotypes, as in case of EPS production or *myo*-inositol utilisation, although the complete pathways were detected in strains (for *myo*-inositol only in DPC4206 and DPC4536). In other cases, such as pullulan utilisation, the genes encoding the initial enzyme of the metabolic pathway was detected in each of the genomes, but no growth in presence of this sugar were

observed, probably since other components, such as specific oligosaccharide transporters were not encoded.

The genome comparison of strains isolated from the same niche confirmed a high level of diversity of *L. paracasei*. By looking at their genomic content assumptions on their evolution and interaction with other strains in the different environments they previously inhabited can be made. This type of analysis looks deeper in the importance of certain systems for survival in specific niches and enables better understanding of the nature of this striking *Lactobacillus* species.

Strains of the *Lactobacillus casei* group are among the most commonly isolated strains of the non-starter flora of bacterial ripened cheeses (Fox et al., 1998, Gobbetti et al., 2015). Recently, these strains have been examined for their application in cheese flavour improvement, and in numerous cases and different types of cheeses their positive or at least non-negative effect has been documented (Crow et al., 2001). Chapter 6 describes the application of the three selected *L. paracasei* strains as adjuncts in Cheddar production and comparison of their ability to contribute to and diversify the flavour. The starter was *Lactococcus lactis* ssp. *lactis* 303, previously shown to acidify curd at the satisfying rate. The application of adjuncts did not affect cheese making and had no impact on the cheese gross composition.

The enumeration of cells showed expected trends observed previously in Cheddar cheese, where a starter dominated the initial stages of ripening, while adjuncts were present in the higher numbers after 3 months of ripening. In regard to primary and secondary proteolysis, no significant difference was observed among cheeses, similar to other studies (Reale et al., 2016, Bielecka and Cichosz, 2017), which means that the pool of protein derived compounds was the same in all cheeses.

In the analysis of cheese volatiles, it was observed that the differences between cheeses were more evident in shorter ripened cheeses (3 months), while the level of differentiation decreased by the end of ripening. Only a few FAA metabolites were present in significantly different abundances in 3 months ripened cheeses. This result was expected, due to the low level of variation in FAA content, which was more prominent in early stages of ripening. Nevertheless, the metabolites of FFA, such as long-chain acids, aldehydes, ketones and aldehydes, contributed the most to the cheese differentiation. They were present in significantly higher abundances in cheese DPC4536 and to a lesser extent in DPC4206, and were the main differentiating compounds in both Month 3 and Month 9 cheeses. This finding showed that an unexpected metabolic route contributed to differentiation, since lipolysis is not the prominent pathway in Cheddar (Collins et al., 2003). However, the majority of other flavour compounds were detected in similar abundances among the cheeses in both time points. In addition, sensorial analysis of cheeses after 9 months of ripening showed that cheeses had highly similar organoleptic characteristics, and only cheese DPC2071 had slightly better attributes.

The results presented in Chapters 2, 3 and 6 lead to a conclusion on the diversity of strains and their flavour forming capacity based on the proteolytic reactions. In the determination of the enzyme activities, the focus was solely on a single substrate metabolism, and the observed level of differentiation was rather high. Further on, the differences in metabolic activities were highly observable in the media with a single amino acid present in the higher concentrations. When a mixture of amino acids was used to assess the metabolic differences (MS1), numerous enzymes were involved in metabolic processes and all of them contributed to the pool of volatiles. It is noticeable that the variation among the strains decreased, as products of other

metabolic pathways were included in the analysis. In an even more complex surrounding that included substrates belonging to the various groups (lipids, proteins, and sugars in MS2) the level of strain differentiation decreased even more. Finally, in the real cheese environment, the strains metabolic activities towards proteins and amino acids were highly similar.

This highlights both advantages and disadvantages of the screening approach used. Although it did show that great differences existed in a very simplified surrounding, they were less visible with the increase in the complexity of the environment. The differentiation of the strains based on the genomics and phenotypic assays provided useful information on the physiology of the strains, but the differences observed in *in vitro* analysis are not a guarantee of a different behaviour at the real application level, as they tend to be minimised by the numerous additional factors.

In conclusion, this thesis presented a comprehensive approach in determining genetic, phenotypic and ecological diversity of strains of *Lactobacillus paracasei*, and their potential to diversify cheese flavour. This species is characterised by the wide genome structure diversity, and even the strains sharing same genomic structure profiles are not entirely identical, in genetic, genomic and phenotypic sense. The metabolic activity of strains examined differed substantially, especially in simple systems, but the variation in flavour generation decreased with the increase of complexity of the environment. In cheese manufacture, strains showed greater ability for flavour differentiation in short ripened Cheddar. The genome analysis revealed numerous genes which provide potential information of strains evolutions and their previous habitats, as well as their specific metabolic characteristics or lack of them. Overall, this thesis demonstrated the genetic and metabolic diversity of *L. paracasei* strains and their potential application to cheese flavour diversification.

7.2 References

Ardo, Y. 2006. Flavour formation by amino acid catabolism. *Biotechnol Adv,* 24**,** 238-242.

Bielecka, M. M. & Cichosz, G. 2017. The influence of an adjunct culture of *Lactobacillus paracasei* LPC-37 on the physicochemical properties of Dutch-type cheese during ripening. *LWT-Food Sci Technol,* 83**,** 95-100.

Cai, H., Rodriguez, B. T., Zhang, W., Broadbent, J. R. & Steele, J. L. 2007. Genotypic and phenotypic characterization of *Lactobacillus casei* strains isolated from different ecological niches suggests frequent recombination and niche specificity. *Microbiology,* 153**,** 2655-2665.

Cai, H., Thompson, R., Budinich, M. F., Broadbent, J. R. & 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.

Collins, Y. F., McSweeney, P. L. H. & Wilkinson, M. G. 2003. Lipolysis and free fatty acid catabolism in cheese: a review of current knowledge. *Int Dairy J,* 13**,** 841-866.

Crow, V., Curry, B. & Hayes, M. 2001. The ecology of non-starter lactic acid bacteria (NSLAB) and their use as adjuncts in New Zealand Cheddar. *Int Dairy J,* 11**,** 275-283.

Dellaglio, F., Felis, G. E. & Torriani, S. 2002. The status of the species *Lactobacillus casei* (Orla-Jensen 1916) Hansen and Lessel 1971 and *Lactobacillus paracasei* Collins et al. 1989. Request for an opinion. *Int J Syst Evol Microbiol,* 52**,** 285-287.

Douillard, F. P. & De Vos, W. M. 2014. Functional genomics of lactic acid bacteria: from food to health. *Microb Cell Fact,* 13 Suppl 1**,** S8.

Fox, P., McSweeney, P. & Lynch, C. 1998. Significance of non-starter lactic acid bacteria in chedder cheese. *Aust J Dairy Technol,* 53**,** 83.

Gobbetti, M., De Angelis, M., Di Cagno, R., Mancini, L. & Fox, P. F. 2015. Pros and cons for using non-starter lactic acid bacteria (NSLAB) as secondary/adjunct starters for cheese ripening. *Trends Food Sci Tech,* 45**,** 167-178.

Koirala, R., Taverniti, V., Balzaretti, S., Ricci, G., Fortina, M. G. & Guglielmetti, S. 2015. Melting curve analysis of a *groEL* PCR fragment for the rapid genotyping of strains belonging to the *Lactobacillus casei* group of species. *Microbiol Res,* 173**,** 50-58.

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

Milesi, M. M., Candioti, M. & Hynes, E. 2007. Mini soft cheese as a simple model for biochemical studies on cheese-making and ripening. *LWT-Food Sci Technol,* 40**,** 1427-1433.

Naser, S. M., Dawyndt, P., Hoste, B., Gevers, D., Vandemeulebroecke, K., Cleenwerck, I., Vancanneyt, M. & Swings, J. 2007. Identification of lactobacilli by *pheS* and *rpoA* gene sequence analyses. *Int J Syst Evol Microbiol,* 57**,** 2777-2789.

Reale, A., Ianniello, R. G., Ciocia, F., Di Renzo, T., Boscaino, F., Ricciardi, A., Coppola, R., Parente, E., Zotta, T. & McSweeney, P. L. H. 2016. Effect of respirative and catalase-positive *Lactobacillus casei* adjuncts on the production and quality of Cheddar-type cheese. *Int Dairy J,* 63**,** 78-87.

Salvetti, E. & O'Toole, P. W. 2017. When regulation challenges innovation: the case of the genus *Lactobacillus*. *Trends Food Sci Technol.* 66**,** 187-194.

Sardaro, M. L. S., Levante, A., Bernini, V., Gatti, M., Neviani, E. & Lazzi, C. 2016. The *spxB* gene as a target to identify *Lactobacillus casei* group species in cheese. *Food Microbiol,* 59**,** 57-65.

Shakeel-Ur, R., Fox, P. F., McSweeney, P. L. H., Madkor, S. A. & Farkye, N. Y. 2001. Alternatives to pilot plant experiments in cheese-ripening studies. *Int J Dairy Technol,* 54**,** 121-126.

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

Soria, A. C., Garcia-Sarrio, M. J. & Sanz, M. L. 2015. Volatile sampling by headspace techniques. *Trends Anal Chem,* 71**,** 85-99.

Steele, J. L., Budinich, M. F., Cai, H., Curtis, S. C. & Broadbent, J. R. 2006. Diversity and metabolic activity of *Lactobacillus casei* in ripening cheddar cheese. *Aust J Dairy Technol,* 61**,** 53.

Tait, E., Perry, J. D., Stanforth, S. P. & Dean, J. R. 2014. Identification of volatile organic compounds produced by bacteria using HS-SPME-GC-MS. *J Chromatogr Sci,* 52**,** 363-373.

Tindall, B. 2008. The type strain of *Lactobacillus casei* is ATCC 393, ATCC 334 cannot serve as the type because it represents a different taxon, the name *Lactobacillus paracasei* and its subspecies names are not rejected and the revival of the name *'Lactobacillus zeae'*contravenes Rules 51b (1) and (2) of the International Code of Nomenclature of Bacteria. Opinion 82. *Int J Syst Evol Microbiol,* 58**,** 1764-1765.

Toh, H., Oshima, K., Nakano, A., Takahata, M., Murakami, M., Takaki, T., Nishiyama, H., Igimi, S., Hattori, M. & Morita, H. 2013. Genomic adaptation of the *Lactobacillus* casei group. *PLoS One,* 8**,** e75073.

Yvon, M. & Rijnen, L. 2001. Cheese flavour formation by amino acid catabolism. *Int Dairy J,* 11**,** 185-201.

Appendix

Appendix 1

Draft genome sequence of *Lactobacillus casei* DPC6800, an isolate with potential to diversify flavour in cheese

All of this appendix has been published in Genome Announcements (2016). 4(2). DOI:10.1128/genomeA.00063-16

8.1 Abstract

Lactobacillus casei is a non-starter lactic acid bacterium commonly present in various types of cheeses. It is believed that strains of this species have a significant impact on the development of cheese flavour. The draft genome sequence of *L. casei* DPC6800, isolated from a semi-hard Dutch cheese, is reported.

8.2 Genome Announcement

Lactobacillus casei is a member of the lactic acid bacteria, a group of Gram-positive, facultatively anaerobic and fastidious bacteria with many biotechnological and health-related applications (Broadbent et al., 2012). Strains of the *L. casei* species show extraordinary niche adaptability and have been found in various habitats, such as milk and dairy products, plant materials, and in the human and animal gastrointestinal tracts (Broadbent et al., 2012, Yu et al., 2015). In dairy products, this organism forms part of the non-starter microbial flora, which has a prominent role during cheese ripening in the development of specific flavour and aroma compounds (Settanni and Moschetti, 2010) through the breakdown of numerous substrates such as amino acids, fatty acids and carbohydrates, during cheese production and ripening (Marilley and Casey, 2004). The subject of this analysis, *Lactobacillus casei* DPC6800, was isolated from a semi-hard Dutch cheese.

Bacterial DNA from strain DPC6800 was extracted, and single end sequencing was performed on a Roche 454 FLX sequencer housed in the Teagasc Sequencing Centre (Moorepark, Fermoy, Cork, Ireland) using standard protocols from the manufacturer (Roche Diagnostics Ltd., West Sussex, UK). Quality filtering, adapter clipping, and trimming of the resulting reads as well as assembly were performed using the SeqMan NGen application of the DNAStar Lasergene Genomics Suite (DNASTAR Inc., Madison, WI, USA). Open reading frames (ORFs) were predicted using Glimmer v3.02 (Delcher et al., 2007) and RAST (Aziz et al., 2008). The genome was annotated using the RAST server, with subsequent annotations verified and manually curated using BLASTp (Altschul et al., 1990) and Artemis (Rutherford et al., 2000).

Sequence assembly yielded a 3053365 bp draft genome with 31 × average coverage, consisting of 58 non-overlapping contigs with a contig *N*50 of 98006 bp and a maximum contig size of 595092bp. Whole genome annotation determined that strain DPC6800 contained a total of 3300 protein-coding genes and 14 tRNAs. Genes that encode enzymes of crucial importance for flavour development were identified, including components of the proteolytic system such as proteinases, peptidases and aminotransferases. The cell wall-associated proteinase PrtP (AC564_0739c) was identified, along with numerous peptidases of broad or specific peptidolytic function, such as tripeptide aminopeptidase (AC564_0751c), methionine aminopeptidase (AC564_0890), aminopeptidase S (AC564_0896), aminopeptidase N (AC564_1879c), aminopeptidase V (AC564_3148c), aminopeptidase C (AC564_3291, AC564_3292), Xaa-Pro-dipeptidyl peptidase (AC564_2630, AC564_2631), Aminotransferases, responsible for the interconversion of amino acids in the later steps of the proteolytic process, are encoded by several genes, i.e. three aspartate aminotransferases (AC564_0742c, AC564_2175, AC564_2467c), two aromatic amino acid aminotransferases (AC564_1682c, AC564_3204) and one branched-chain amino acid aminotransferase (AC564_2001). A gene for glutamate dehydrogenase (AC564_0811c), an enzyme that supports aminotransferase activity through recycling of α-ketoglutarate, an intermediate molecule in aminotransferase reaction, was also identified. Also important for flavour development is the metabolism of citrate, and, the presence of a gene encoding a Mg^{2+}-citrate co-transporter CitMHS, necessary for the initial steps of citrate metabolism, was confirmed (AC564_1305). The findings of the genome analysis confirm the potential of *L. casei* DPC6800 for use as an adjunct culture in cheese production to direct or enhance cheese flavour.

Nucleotide sequence accession numbers. This Whole Genome Shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession LNQD00000000. The version described in this paper is version LNQD01000000.

8.3 Acknowledgments

E. Stefanovic is the recipient of the Teagasc Walsh Fellowship.

Funding information: This work was funded by Dairy Research Ireland (6156) and Teagasc (6224).

8.4 References

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

Aziz, R. K., Bartels, D., Best, A. A., DeJongh, M., Disz, T., Edwards, R. A., Formsma, K., Gerdes, S., Glass, E. M. & Kubal, M. 2008. The RAST Server: rapid annotations using subsystems technology. *BMC Genomics,* 9**,** 75.

Broadbent, J. R., Neeno-Eckwall, E. C., Stahl, B., Tandee, K., Cai, H., Morovic, W., Horvath, P., Heidenreich, J., Perna, N. T., Barrangou, R. & Steele, J. L. 2012. Analysis of the *Lactobacillus casei* supragenome and its influence in species evolution and lifestyle adaptation. *BMC Genomics,* 13**,** 533.

Delcher, A. L., Bratke, K. A., Powers, E. C. & Salzberg, S. L. 2007. Identifying bacterial genes and endosymbiont DNA with Glimmer. *Bioinformatics,* 23**,** 673-679.

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

Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M. A. & Barrell, B. 2000. Artemis: sequence visualization and annotation. *Bioinformatics,* 16**,** 944-945.

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

Yu, S., Peng, Y., Zheng, Y. & Chen, W. 2015. Comparative Genome Analysis of *Lactobacillus casei*: Insights into Genomic Diversification for Niche Expansion. *Indian J Microbiol,* 55**,** 102-107.

Appendix 2

Draft genome sequences of three *Lactobacillus paracasei* strains, members of the non-starter microbiota of mature Cheddar cheese

All of this appendix has been published in Genome Announcements (2017) 5(29)
DOI: doi.org/10.1128/genomeA.00655-17

9.1 Abstract

Lactobacillus paracasei strains are common members of the non-starter microbiota present in various types of cheeses. The draft genome sequences of three strains isolated from mature Cheddar cheeses are reported.

9.2 Genome Announcement

Lactic acid bacteria (LAB) are Gram positive aerotolerant bacteria with a wide spectrum of practical applications, including food production, biotechnology and medicine-related fields (Makarova et al., 2006). Strains of the genus *Lactobacillus* have been isolated from diverse habitats, such as fermented products, plant materials, and human and animal gastrointestinal tracts (Smokvina et al., 2013). In cheese, *Lactobacillus paracasei* form part of the non-starter microbiota, and are considered to have an important role in the ripening process and flavour development (Gobbetti et al., 2015). Three *Lactobacillus paracasei* strains (DPC2071, DPC4206 and DPC4536) analysed in this study were isolated from mature Cheddar cheeses as part of the non-starter LAB population.

Bacterial DNA was isolated from all three strains, and genomic libraries were prepared with the Nextera® XT DNA library preparation kit (Illumina, Inc. San Diego, CA, USA). The 2 × 250 bp paired end reads sequencing was performed on a Illumina MiSeq platform (MicrobesNG, University of Birmingham, UK) The assembly of each genome was performed using the SeqMan NGen application of the DNAStar Lasergene Genomics Suite (DNASTAR Inc., Madison, WI, USA). Glimmer v3.02 (Delcher et al., 2007) and RAST (Aziz et al., 2008) were used to predict open reading frames (ORFs). Initially, the RAST server was used to annotate each genome, and the annotations were verified and manually curated using BLASTp (Altschul et al., 1990) and Artemis (Rutherford et al., 2000).

Sequence assemblies for the three strains indicated coverage of 88 ×, 70 × and 101 × for DPC2071, DPC4206 and DPC4536 respectively. The length of the DPC2071 genome was 2936872 bp consisting of 41 non-overlapping contigs, with a contig

*N*50 of 300051 bp, a maximum contig size of 536232 bp. and a total of 2827 protein-coding genes. In the case of strain DPC4206, the assembly yielded a genome sequence of 3095268 bp, consisting of 49 contigs. The maximum contig size was 322047 bp and contig *N*50 was 142300 bp, while the total of 2951 protein-coding genes was identified. The draft genome sequence of strain DPC4536 was 3078575 bp long and it consisted of 35 contigs and 2,931 genes encoding proteins. The maximum contig size was 426277 bp and the contig *N*50 was 191,696 bp. The GC content of all three genomes was 46.3 %, which corresponds to the usual GC content of *L. paracasei* genomes.

This sequencing data will contribute to the pool of available *Lactobacillus paracasei* genomes and enable further comparative genome analysis of strains of this species.

Nucleotide sequence accession numbers. The Whole Genome Shotgun projects have been deposited at DDBJ/EMBL/GenBank under the accessions NCSN01000000, NCSO01000000 and NCSP01000000, while the versions described in this paper are versions NCSN00000000, NCSO00000000 and NCSP00000000 for strains DPC2071, DPC4206 and DPC4536, respectively.

9.3 Acknowledgments

E. Stefanovic is the recipient of the Teagasc Walsh Fellowship. This work was funded by Teagasc (Grant 6224).

9.4 References

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

Aziz, R. K., Bartels, D., Best, A. A., DeJongh, M., Disz, T., Edwards, R. A., Formsma, K., Gerdes, S., Glass, E. M. & Kubal, M. 2008. The RAST Server: rapid annotations using subsystems technology. *BMC Genomics,* 9**,** 75.

Delcher, A. L., Bratke, K. A., Powers, E. C. & Salzberg, S. L. 2007. Identifying bacterial genes and endosymbiont DNA with Glimmer. *Bioinformatics,* 23**,** 673-679.

Gobbetti, M., De Angelis, M., Di Cagno, R., Mancini, L. & Fox, P. F. 2015. Pros and cons for using non-starter lactic acid bacteria (NSLAB) as secondary/adjunct starters for cheese ripening. *Trends Food Sci Tech,* 45**,** 167-178.

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

Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M. A. & Barrell, B. 2000. Artemis: sequence visualization and annotation. *Bioinformatics,* 16**,** 944-945.

Smokvina, T., Wels, M., Polka, J., Chervaux, C., Brisse, S., Boekhorst, J., van Hylckama Vlieg, J. E. & Siezen, R. J. 2013. *Lactobacillus paracasei* comparative genomics: towards species pan-genome definition and exploitation of diversity. *PLoS One,* 8**,** e68731.

Acknowledgments

First and foremost I would like to thank to Dr Olivia McAuliffe for everything she did for me over the last four years. Thank you for your patience and words of comfort and for re-assuring me that I will succeed. Thank you for working so diligently on improving my practical but most importantly my writing skills. Thanks for the words of constructive critics, advice and all the challenges that you had put in front of me, to make me the best possible candidate. I am more than sure that all these will contribute in my future career.

Secondly, I would like to thank to Dr Kieran Kilcawley for his constant support over the course of the project. I highly appreciate the opportunity initiated by you to spend three months at INRA Laboratories, in Rennes, France. Also, I will be eternally grateful to you for having your doors always opened for me and my numerous questions.

Thirdly, I would like to sincerely thank to Dr Anne Thierry, the leader of Microbiology Team at INRA STLO, Rennes, France. Thank you for the lovely three months I spent in Rennes under your supervision. I learned so much with you. And thank you for endless patience and several hours' long Skype conversation that helped in successful writing of two chapters of this thesis.

I would also like to thank to prof. Gerald Fitzgerald for his input to my project.

A very special thanks to Teagasc Walsh Fellowship scheme for funding my studies.

I am also grateful to Dr Mary Rea for valuable advice she gave me over the four years I spent in Moorepark, and to Dr Brian Healy, whose words of support meant a lot.

Thanks to all people from Lab 96, Charlotte, Aidan and Dan who were there for me when I started, and Vincenzo, Clara, Leonardo, Shahneela, Alan, Conor, Ludivine, Carlos, Rebecca, Siobhan and many others who enriched serious lab work with jokes and so much laughter. And bowling!

Beatriz, you are my guardian angel, and sometimes just seeing you at the corridor helped me to overcome hurdles and difficult moments everyone experience during

their studies. Our long walks, talks and tremendous moments are forever encrypted in my heart, wherever our journeys bring us in future.

Heni, our dinners will be always remembered! And all the cartoons we watched in cinema! And my viva is on your birthday!

Gwynneth, you are one of the best students I have ever met! Thank you for the lovely collaboration we had over the four months you spent in Lab 96 and for the brilliant results we obtained together! For your words of comfort and support, and for believing in me when I didn't do that myself.

Sanja, my Cimi, an immense thanks for all the advice, support, travelling around Ireland, help in illness, shopping, going out to the Garden for 5 euro, small signs of attention that sometimes mean a lot more than anything else. Thank you, Zeko, for being with me in the good and the evil, for taking care of my Bubica while I was in Rennes, for your support when I changed houses. Thank you for being a true friend!

Lana, my Chocolate Banana, God Almighty sent you to me that day in Aula Maxima. Thank you for our unforgettable journeys across the Europe (and many more of them still to come!), for all your advice and criticism. For calling me when I least expect but I need the most, for putting up with all my outbursts. Your support and presence are an essential part of this doctorate.

Finally, the words of gratitude go to my parents. Dear mum and dad, I'd say the journey has finished, and I will never be a student again. I hope you are proud of me. Thank you for putting up with me for these several years, I know I can be impossible sometimes. Thank you for supporting me in all possible ways. For your immense devotion, and for your belief that I have chosen the right pathway in life. Mum, my Linda, thanks for all the help with Figures, Posters and Presentations. For every day Skypes, for your patience for my different moods. I dedicate this thesis to you, guys, and you are obliged to read it, from the beginning to the end! :)