

**UCC Library and UCC researchers have made this item openly available.  
Please [let us know](#) how this has helped you. Thanks!**

<b>Title</b>	Advances in the genomics and metabolomics of dairy lactobacilli: A review
<b>Author(s)</b>	Stefanovic, Ewelina; Fitzgerald, Gerald F.; McAuliffe, Olivia
<b>Publication date</b>	2016-08-31
<b>Original citation</b>	Stefanovic, E., Fitzgerald, G. and McAuliffe, O. (2016) 'Advances in the genomics and metabolomics of dairy lactobacilli: A review', Food Microbiology, 61, pp. 33-49. doi:10.1016/j.fm.2016.08.009
<b>Type of publication</b>	Review
<b>Link to publisher's version</b>	<a href="http://dx.doi.org/10.1016/j.fm.2016.08.009">http://dx.doi.org/10.1016/j.fm.2016.08.009</a> Access to the full text of the published version may require a subscription.
<b>Rights</b>	© 2016, Elsevier Ltd. All rights reserved. This manuscript version is made available under the CC-BY-NC-ND 4.0 license. <a href="http://creativecommons.org/licenses/by-nc-nd/4.0/">http://creativecommons.org/licenses/by-nc-nd/4.0/</a>
<b>Item downloaded from</b>	<a href="http://hdl.handle.net/10468/5004">http://hdl.handle.net/10468/5004</a>

Downloaded on 2021-04-19T15:03:50Z

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

2

3 Ewelina Stefanovic<sup>a,b</sup>, Gerald Fitzgerald<sup>b</sup> and Olivia McAuliffe<sup>a,\*</sup>

4

5 <sup>a</sup>Teagasc Food Research Centre, Moorepark, Fermoy, Ireland

6 <sup>b</sup>School of Microbiology, University College Cork, Cork, Ireland

7

8 \*Corresponding author: Olivia McAuliffe, Teagasc Food Research Centre, Moorepark

9 Fermoy, Ireland. Tel: +3532542609, Fax +3532542340, Email: olivia.mcauliffe@teagasc.ie

10 **Abstract**

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

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

## 29 **1. Introduction**

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

53 developments in genomics and metabolomics provide the tools for a more ‘knowledge-based’  
54 approach to selection of desirable cultures for application in industry (McAuliffe, 2017).  
55 LAB are phylogenetically closely related, but the number of predicted protein-coding genes  
56 in the LAB varies between 1,700 and 2,800 (Makarova et al., 2006). Genomic studies of  
57 members of the LAB have confirmed the overall trend of minimisation of genomes, which is  
58 in close agreement with the transition to nutritionally rich environments. Nevertheless, some  
59 gene families were expanded by gene duplication or acquisition of paralogous genes via  
60 horizontal gene transfer (HGT) (Makarova et al., 2006). Based on the analysis of the  
61 genomes of 12 LAB species it was concluded that the core LAB genome, comprising  
62 orthologous genes conserved in all analysed genomes (Collins and Higgs, 2012), consists of  
63 567 genes, mostly encoding translation, transcription and replication processes, but 41 of the  
64 genes were uncharacterized and 50 had only general functions predicted. This study also  
65 identified two core genes exclusive for LAB, the products of which are LysM  
66 (peptidoglycan-binding) domain and the highly conserved LaCOG01237 with no known  
67 domains, but based on its localisation, it is probably involved in modification of tRNA  
68 (Makarova et al., 2006).

69 The genus *Lactobacillus* comprises a diverse group of bacteria currently consisting of more  
70 than 200 species and subspecies (Sun et al., 2015a) that share the common features of other  
71 LAB, including low GC content, acid tolerance and conversion of sugars to lactic acid as one  
72 of the main end products of metabolism. Species of lactobacilli are present in various  
73 environments such as plants, fermented food products (dairy, meat, wine), and both the  
74 human and animal gastrointestinal tracts. Their ability to ferment milk, meat and plant  
75 material presents the basis for their artisanal and industrial usage (Sun et al., 2015a). Apart  
76 from this, strains of *Lactobacillus* are well known for their probiotic properties (Lebeer et al.,  
77 2008).

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

## 85 **2. Genomics of the *Lactobacillus* genus**

86 Due to their importance in various biotechnological and health-related applications, there has  
87 been a growing interest in exploring the genomic features of the genus *Lactobacillus*, which  
88 is the largest and most diverse genus of LAB (Broadbent et al., 2012). *Lactobacillus* genomes  
89 range in size from 1.23 Mbp (*Lb. sanfranciscensis*) to 4.91 Mbp (*Lb. parakefiri*) (Sun et al.,  
90 2015a). Species of this genus are present in dairy products (*Lb. delbrueckii* ssp. *bulgaricus*,  
91 *Lb. helveticus*), human and animal gastrointestinal tracts (*Lb. acidophilus* and *Lb. gasseri*) or  
92 in a variety of niches (*Lb. plantarum*, *Lb. pentosus*, *Lb. brevis*, and *Lb. paracasei*) (Smokvina  
93 et al., 2013). The first genome of the *Lactobacillus* genus sequenced was *Lb. plantarum*  
94 WCFS1 (Kleerebezem et al., 2003) followed by *Lb. johnsonii* NC533 (Pridmore et al., 2004)  
95 and *Lb. acidophilus* NCFM (Altermann et al., 2005). These studies revealed some interesting  
96 genomic features of the *Lactobacillus* genus, such as lifestyle adaptation islands in *Lb.*  
97 *plantarum* WCFS1, lack of general biosynthetic pathways in the probiotic strain *Lb. johnsonii*  
98 NC553 and unique structures called potential autonomous units (PAU) in *Lb. acidophilus*  
99 NCFM, all of which triggered further investigation and comparison with newly sequenced  
100 strains of the same species. Currently (July 2016), there are 214 *Lactobacillus* genome  
101 sequencing projects available in public databases (<http://www.ncbi.nlm.nih.gov>).  
102 The pangenome (or supragenome) is considered as the full set of all genes within a selected  
103 genome set (species, genera or higher taxonomic groups) (Medini et al., 2005; Collins and  
104 Higgs, 2012). The size of the pangenome generated for *Lactobacillus* and associated genera  
105 of LAB reaches almost 45000 gene families, while 73 genes mainly responsible for cell  
106 growth and replication make up the core genome (Sun et al., 2015a). In a study based on the  
107 features of 20 complete *Lactobacillus* genomes representing 14 species whose genomes  
108 ranged from 1.8 to 3.3 Mbp, the number of proteins within these genomes was between 1721-  
109 3100 (Kant et al., 2011). The estimated size of the pangenome of the *Lactobacillus* genus

110 consists of almost 14000 proteins, while the core genome consists of 383 orthologs (Kant et  
111 al., 2011). This number is higher than the 141 core genes reported in the study of Claesson et  
112 al. (2008), who used more strict criteria and took into account only 12 completely sequenced  
113 *Lactobacillus* genomes. Over 100 out of 383 genes of the *Lactobacillus* core genome were  
114 organised in operon-like clusters that are conserved in other related Gram-positive bacteria  
115 (Kant et al., 2011). Among 41 genes specific for *Lactobacillus*, 13 were predicted to code for  
116 ribosomal proteins, and 13 were annotated as hypothetical (Kant et al., 2011). Taken together,  
117 comparative genomic studies of lactobacilli confirmed the overall trend observed in other  
118 LAB, which is loss of ancestral genes and minimisation of genomes, as well as acquisition of  
119 genes by HGT as a response to adaptation to the primary habitat of these bacteria (Makarova  
120 et al., 2006).

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

### 130 **2.1 *Lactobacillus delbrueckii***

131 From the perspective of the dairy industry, *Lactobacillus delbrueckii* contains two  
132 industrially important subspecies: subspecies *bulgaricus* and subspecies *lactis*. Of the 22  
133 genome sequences available for these two subspecies, five are complete sequences. While *Lb.*  
134 *delbrueckii* ssp. *bulgaricus* is widely used in the production of yoghurt, subspecies *lactis* is



135 used primarily as a starter in the manufacture of cheeses like Emmental, Grana Padano and  
136 Parmigiano Reggiano (El Kafsi et al., 2014). The core genome of the three *Lb. delbrueckii*  
137 *ssp. bulgaricus* strains (2038, ATCC 11842 and ATCC BAA-365) consists of 1276 genes,  
138 with the genomes of strains 2038, ATCC 11842 and ATCC BAA-365 consists of 211, 150  
139 and 166 unique genes, respectively (Hao et al., 2011). An alignment of the three genomes  
140 revealed two duplicated segments flanking the predicted replication terminus, but strain 2038  
141 has a unique 8.5 kbp region between the duplication regions, which could be the reason for  
142 the bigger genome size (1.87 Mbp compared to 1.86 Mbp ATCC 11842 and ATCC BAA-  
143 365). This region is most likely inherited from an ancestor, but lost in the other two strains,  
144 probably due to their independent evolution from strain 2038 (Hao et al., 2011).

145 A genome analysis of sequenced *Lb. delbrueckii* strains showed that the average GC and  
146 GC3 content (GC at codon position 3) in coding sequences (CDSs) is approximately 52% and  
147 65%, respectively (El Kafsi et al., 2014), which is in agreement with a previously reported  
148 higher GC content in *Lb. delbrueckii ssp. bulgaricus* compared to other lactobacilli (van de  
149 Guchte et al., 2006). Higher GC content is a sign of rapid ongoing evolution in these species  
150 (O'Sullivan et al., 2009). In both subspecies, decay and inactivation of superfluous genes was  
151 evident, indicating an evolutionary trend towards adaptation to the dairy environment. A  
152 deeper insight into the genomics of these subspecies revealed some interesting genetic  
153 differences. Firstly, it was shown that the size of the *ssp. bulgaricus* genomes is smaller  
154 compared to *ssp. lactis* (1.8 Mbp and 2 Mbp, respectively). However, the number of CDS did  
155 not differ considerably between the two subspecies, as it varied in range from 1333-1783 for  
156 subspecies *bulgaricus* to 1593-1721 for subspecies *lactis*. Comparison of the core proteomes  
157 of five *ssp. lactis* and five *ssp. bulgaricus* strains surprisingly revealed quite similar sizes of  
158 core proteomes and significant overlapping of these. The overall core proteome consists of  
159 989 proteins, with 65 proteins specific for *ssp. lactis* and 25 proteins specific for *ssp.*

160 *bulgaricus*. The majority of the 65 specific *ssp. bulgaricus* proteins have unknown functions,  
161 while those of known function are mainly membrane transporter-associated proteins. The 25  
162 specific *ssp. lactis* proteins have mainly known functions, involved in carbohydrate and  
163 amino acid metabolism. For both subspecies, fragments of other subspecies-specific genes  
164 could be found as pseudogenes, implying that differential loss of genes caused subspecies  
165 divergence. Another important finding of the extensive genomic analysis is re-classification  
166 of strain ND02, which was designated as *ssp. bulgaricus* but confirmed to be *ssp. lactis*, not  
167 only due to the larger genome but also due to the higher number of insertion sequences (IS).  
168 Besides that, it was previously shown that *Lb. delbrueckii* subspecies can be distinguished  
169 based on the number of *EcoRI* sites in their 16S rDNA sequences, where *ssp. lactis* possesses  
170 one, and *ssp. bulgaricus* has two restriction sites (Giraffa et al., 1998). The detailed analysis  
171 of 16S rRNA of strain ND02 showed it did not contain two specific restrictions sites, adding  
172 an argument to its re-classification as *ssp. lactis* (El Kafsi et al., 2014).

## 173 **2.2 *Lactobacillus helveticus***

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

### 185 **2.3. The *Lactobacillus casei/paracasei* group**

186 The taxonomic status of *Lb. casei* is still a matter of much debate (Smokvina et al., 2013) as  
187 molecular studies have implied that the majority of *Lb. casei* strains are more related to *Lb.*  
188 *casei* ATCC 334 (also named *Lb. paracasei*) than to the official type strain *Lb. casei* ATCC  
189 393 (Dellaglio et al., 2002). Because of this uncertainty, the information available for both  
190 *Lb. casei* and *Lb. paracasei* will be reviewed together here. The members of this group have  
191 been isolated from dairy and plant materials (cheese, wine, pickle, silage) (Toh et al., 2013)  
192 and reproductive and gastrointestinal tracts of humans and animals (Cai et al., 2009). In the  
193 cheese industry, they are used as adjunct cultures for development of desired flavour (Milesi  
194 et al., 2010; Van Hoorde et al., 2010). Besides application in fermented food production,  
195 members of this group are well known for their probiotic characteristics (Herias et al., 2005;  
196 Ya et al., 2008). Such a diverse range of sources and broad ranging possible applications  
197 makes this group one of the best explored species within the *Lactobacillus* genus with eight  
198 and seven genome sequences completed for *Lb. casei* and *Lb. paracasei*, respectively, and 27  
199 and 46 draft genome sequences available for *Lb. casei* and *Lb. paracasei*, respectively.  
200 Genome sizes range from 2.38 Mbp for *Lb. paracasei* ssp. *tolerans* DMS20258 and 3.27 Mbp  
201 for *Lb. casei* Lbs2, with an average GC content of 46.5%. Analysis of the draft sequences of  
202 12 strains of *Lb. casei* of different origins (dairy, plant and human) along with five fully  
203 sequenced genomes have determined that the size of the *Lb. casei* pangenome is 3.2 X the  
204 average genome size, consisting of 1715 core and 4220 accessory genes (Broadbent et al.,  
205 2012). Another comparative study (Yu et al., 2015) performed on 12 draft *Lb. casei* genomes  
206 revealed 806 novel regions larger than 500 kbp harbouring both hypothetical proteins and  
207 mobile genetic elements in these strains compared to the seven complete genomes. This  
208 suggested that the *Lb. casei* pangenome expands with every new sequenced genome and  
209 potential for environmental adaptation within the species increases (Yu et al., 2015).

210 Similarly, when 37 genomes of *Lb. paracasei* were analysed, 1800 core and 4200 accessory  
211 genes were detected (Smokvina et al., 2013). A common feature of all 37 analysed genomes  
212 of *Lb. paracasei* is a cluster involved in the conversion of branched chain alpha-keto acids  
213 into branched chain fatty acids important for maintenance of the colonic epithelium. This  
214 gene cluster is unique for *Lb. paracasei*, implying its acquisition through HGT (Smokvina et  
215 al., 2013). Pangenome analysis revealed the ability of *Lb. paracasei* to utilise a broad range  
216 of carbohydrates. In total, 74 sugar utilisation cassettes were detected 15 of which belonged  
217 to the core genome. These cassettes were localised on two genomic islands (Smokvina et al.,  
218 2013), structures usually connected with the environmental adaptation (described in details  
219 below).

#### 220 **2.4 *Lactobacillus acidophilus***

221 Taxonomically, *Lactobacillus acidophilus* is part of a larger complex comprising several  
222 species: *Lb. acidophilus*, *Lb. amylovorus*, *Lb. crispatus*, *Lb. gallinarum*, *Lb. gasseri*, and *Lb.*  
223 *johnsonii* (Berger et al., 2007; Ramachandran et al., 2013). Strains of *Lb. acidophilus* are  
224 often used in dairy products as probiotics and as flavour contributing strain in certain dairy  
225 products, such as yoghurt, sweet acidophilus milk and cheese (Buriti et al., 2005; Ong et al.,  
226 2007; Ejtahed et al., 2011). The genome of *Lactobacillus acidophilus* NCFM was the first *Lb.*  
227 *acidophilus* to be sequenced (Altermann et al., 2005). Presently, 16 strains of this species  
228 have been sequenced, with three complete genomes available. Genomes range in size from  
229 1.25 - 2.05 Mbp, with GC content of 34.7%. Although phenotypic and biochemical  
230 characterisation of strains show a certain level of diversity, genotypic analysis indicates less  
231 variation within genomes of this species (Ramachandran et al., 2013; Stahl and Barrangou,  
232 2013; Bull et al., 2014). In a recent study reporting the genome sequences of *Lb. acidophilus*  
233 strains isolated from yoghurt (Iartchouk et al., 2015), the alignment of the three sequenced  
234 genomes (FSI4, NCFM, and La-14) confirmed a high level of genome similarity for these

235 strains at the DNA level. Similarly, alignment of La14 and NCFM showed extremely high  
236 similarity between these two strains and synteny with ATCC 4769 (Stahl and Barrangou,  
237 2013). Strain 30SC was initially designated as *Lb. acidophilus*, but unlike other strains of this  
238 species, it possesses 2 plasmids and has higher GC content (38%) (Stahl and Barrangou,  
239 2013). After detailed phylogenetic analysis of its genome, it was re-classified as *Lb.*  
240 *amylovorus* (Bull et al., 2014)

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

## 260 **2.5 *Lactobacillus rhamnosus***

261 *Lactobacillus rhamnosus* is present in various dairy products, such as cheese and yoghurt, but  
262 also in human cavities and gastrointestinal tract (GIT) (Douillard et al., 2013; Kant et al.,  
263 2014). In dairy products, it is mainly present as part of the non-starter LAB (NSLAB) in  
264 Italian cheese varieties (Gobbetti et al., 2015), and there is evidence of its positive effect on  
265 flavour development in these products (Sgarbi et al., 2013; Innocente et al., 2016). However,  
266 its main application is as probiotic cultures (Tuo et al., 2013), often administered through  
267 fermented dairy products. To date, 102 genome sequences have been elucidated, with the  
268 completed sequences of six strains available. The size of the genomes range from 2.52 Mbp  
269 for strain MTCC 5462 up to 3.41 Mbp for strain CRL1505, and the average GC content is  
270 46.7%. General genomic features of this species were determined based on 100 sequenced  
271 strains of various origin (cheese, yoghurt, vaginal cavity, oral cavity, intestinal tract, abscess,  
272 blood, clinical isolates) mapped according to the reference strain *Lb. rhamnosus* GG. The  
273 number of shared genes between these 100 strains and strain GG ranged from 87-100%  
274 (Douillard et al., 2013). The pangenome analysis based on the complete or draft genomes of  
275 13 strains, originating from various environments (milk, human airways, faeces, dairy starter,  
276 infected dental pulp, Cheddar cheese and gut biopsy), estimates a total of 4893 genes, 1.6 X  
277 the average size of a *Lb. rhamnosus* genome (Kant et al., 2014). Pangenome studies show  
278 that, in general, the rate of increase of the size of the pangenome slows down with every  
279 additional genome being sequenced (Kant et al., 2014). As the pangenome curve of *Lb.*  
280 *rhamnosus* reaches a plateau at about 5000 genes, it is predicted that with only a few more  
281 additional genomes of strains from different origins would be sufficient to reach total genome  
282 variability of the species (Kant et al., 2014). The core genome of *Lb. rhamnosus* is estimated  
283 to encode 2095 genes, or approximately 43% of the pangenome. There are at least 75 genes  
284 present only in *Lb. rhamnosus* species, and the majority of these are hypothetical proteins

285 followed by membrane transporters, transcriptional regulators and glycosyl-transferases. The  
286 dispensable genome, which contains genes present in two or more strains (Medini et al.,  
287 2005), of *Lb. rhamnosus* is estimated to contain 2798 genes, and the number of unique  
288 (strain-specific) genes is 855, which is approximately 30% of the dispensable genome. Most  
289 of the dispensable genes in the *Lb. rhamnosus* pangenome are annotated as hypothetical and  
290 it remains unknown what proportion of these would actually encode functional proteins (Kant  
291 et al., 2014).

## 292 **2.6 *Lactobacillus plantarum***

293 *Lactobacillus plantarum* is present in many ecological niches ranging from vegetables, meat,  
294 dairy products and gastro-intestinal tract. Apart from a prominent role in fermentations such  
295 as sourdough (Corsetti and Settanni, 2007), strains of this species are present in dairy  
296 fermentations and non-starter flora (Settanni and Moschetti, 2010; Gobbetti et al., 2015).  
297 Besides that, they are well known for their probiotic characteristics (Siezen and van  
298 Hylckama Vlieg, 2011). To date, 114 genome sequences are publically available, with 18  
299 completely sequenced genomes. The genome of this species is one of the largest in the  
300 *Lactobacillus* group, with a size of approximately 3.4 Mbp, and a GC content of 44.4%. In an  
301 extensive study, 185 isolates from different environments were phenotypically characterized,  
302 and based on the observed phenotypic diversity, a set of 42 candidates were selected for  
303 genomic analysis (Siezen et al., 2010). The core genome of *Lb. plantarum* was found to  
304 comprise 2050 - 2200 genes. Approximately 120 fully conserved genes were unique to *Lb.*  
305 *plantarum*. Many of the unique genes encode hypothetical proteins, while some genes encode  
306 functions that could be used for phenotyping. The two candidates are a conserved cluster for  
307 tartarate and sulfur uptake and metabolism, which are associated with plant habitats (Siezen  
308 et al., 2010). The reference genome WCFS1 itself has over 50 genes not found in any of the  
309 other selected strains isolated from different environments. Most notable are three gene

310 clusters encoding exopolysaccharide, a putative macrolide and a non-ribosomal synthesized  
311 hybrid peptide-polyketide, all of which take part in the interaction with environment. They  
312 were most likely acquired in a recent evolutionary event due to their GC content, suggesting  
313 adaptations necessary for survival in a specific niche (Siezen et al., 2010). Apart from these  
314 50 genes, all other strains were estimated to lack between 9% and 20% of genes present in the  
315 reference genome, WCFS1. These genes are mainly organised in functional gene clusters, or  
316 cassettes as parts of operons and they encode prophages, restriction/modification systems,  
317 exopolysaccharide, bacteriocin and non-ribosomal peptide biosynthesis and carbohydrate  
318 utilisation components and are located on genomic islands (described in details in the next  
319 section) (Siezen et al., 2010; Siezen and van Hylckama Vlieg, 2011).



### 320 **3. Niche adaptability of lactobacilli**

321 The widespread dissemination of members of the lactobacilli in different environments  
322 testifies to their extraordinary niche adaptability. Lactobacilli are present in grass and on  
323 plant material, in dairy products, on human skin, in the mouth, intestine and in the female  
324 reproductive system (Claesson et al., 2007), habitats with many contrasting environmental  
325 conditions (temperature, pH value, available nutrients, and competing microorganisms).  
326 Comparative genomic analysis has revealed that adaptation to such highly variable  
327 environments is a result of genome evolution and the genetic basis for niche specialisation  
328 appears to be the result of eliminating anabolic systems that are not needed through  
329 adaptation to nutritionally rich habitats, such as milk. On the other hand, in all LAB,  
330 including lactobacilli, duplications of genes coding for transporters and metabolism of  
331 carbohydrates, amino acid transporters and peptidases occurred, further enhancing the ability  
332 of these species to live in nutrient-rich environments (Fig. 1a) (Douglas and Klaenhammer,  
333 2010; Makarova and Koonin, 2007; Mayo et al., 2008).

#### 334 **3.1 Horizontal gene transfer (HGT) is the main pathway of niche adaptability in** 335 **lactobacilli**

336 Although gene loss and acquisition, which are the principal events resulting in niche  
337 adaptation, occur in different ways, HGT via bacteriophages, transposons and other mobile  
338 elements appears to be an especially dominant force of adaptation to novel environments in  
339 *Lactobacillus* species (Broadbent et al., 2012), and it is responsible for various genome  
340 rearrangements (Rossi et al., 2014). Such events have made the LAB amenable to adaption to  
341 different habitats, including milk and other food matrices, plant material, and GIT.  
342 Transposons and plasmids present the main mechanism of gene exchange that occurs  
343 amongst different taxonomic groups that do not possess strictly controlled  
344 restriction/modification systems (Rossi et al., 2014). Both niche specialists and generalists

345 have undergone multiple genetic changes which have led to restriction or broadening of the  
346 possible habitats in which these strains could survive.

347 Apart from the traditional classes of mobile genetic elements (plasmids and prophages),  
348 structures acquired by the host bacteria through HGT comprising mobile elements and genes  
349 contributing to the ability of the host to adapt to specific conditions of habitat, are known as  
350 genomic islands (GI) (Bellanger et al., 2014). The first record of “lifestyle adaptation” islands  
351 in *Lactobacillus* was in the genome of *Lb. plantarum* WCFS1, where numerous genes  
352 involved in sugar transportation and metabolism are grouped together in a region  
353 characterised by lower GC content (41.5%) than the rest of the genome (44.45%), suggesting  
354 recent acquisition by HGT (Kleerebezem et al., 2003). Apart from strong overrepresentation  
355 of genes involved in energy metabolism, regulatory proteins coordinating sugar metabolism  
356 are also present on GI (Molenaar et al., 2005). In strain *Lb. helveticus* DPC4571, a number of  
357 amino acid metabolism genes along with lipid biosynthesis genes were also identified in a  
358 region characterised with higher GC content (42% compared to 37% in the rest of the  
359 genome) and insertion sequences flanking this region suggest a recent transfer of this GI  
360 (Callanan et al., 2008). One of the GI of *Lb. casei* BL23 carries genes for catabolism of  
361 myoinositol, a cyclic polyol not commonly metabolised by LAB and potentially present in  
362 degrading plant material (Yebra et al., 2007, Cai et al., 2009). Genomic islands of *Lb. casei*  
363 ATCC 334 encode hypothetical proteins and transcriptional regulators, sugar transporters and  
364 metabolic enzymes and are characterised by high prevalence of insertion sequences,  
365 recombinases, integrases with higher GC content supporting their recent acquisition and a  
366 heterologous origin (Cai et al., 2009). The 26 genomic islands of *Lb. rhamnosus* ATCC  
367 53103, isolated from the human gut, include six carbohydrate utilisation gene clusters, which  
368 seem to have secured the survival of the strain in a less nutritionally rich environment, such

369 as the human intestine (Toh et al., 2013). These examples of different genes present in GI  
370 confirm their importance for adaptation and survival in specific environmental conditions.

### 371 **3.2 Niche adaptation studies reveal lactobacilli as niche specialists**

372 Niche specialists can be described as strains that are able to live in a limited number of  
373 habitats, while niche generalists have the capacity to populate various environments. Genome  
374 analysis of dairy specialists show that these strains have an abundance of sugar  
375 transportation, proteolysis and amino acid transportation encoding genes, some of which have  
376 undergone duplication as they enable the organism to uptake nutrients from the rich milk  
377 environment (Makarova et al., 2006). On the other hand, substantial gene decay has been  
378 confirmed in some lactobacilli, such as in the dairy *Lb. casei* strains, which have more than  
379 120 CDS absent. As a result, these strains have improved their ability to survive in the dairy  
380 niche but have a reduced capacity for survival in other niches (Cai et al., 2009). In the  
381 genomes of dairy LAB, more than 10% of coding genes are present only as pseudogenes  
382 (Zhu et al., 2009), which are non-functional due to frameshifts, nonsense mutations, deletions  
383 or truncations (O'Sullivan et al., 2009). For example, the dairy isolate *Lb. helveticus*  
384 DPC4571 is reported to have 217 pseudogenes, while *Lb. bulgaricus* ATCC 11842 carries a  
385 staggering 533 pseudogenes coding for proteins involved in regulating amino acid and  
386 nucleotide metabolism and bile salt hydrolysis (Callanan et al., 2008; O'Sullivan et al., 2009).  
387 In contrast, species mainly present in the gut, such as *Lb. acidophilus*, *Lb. gasseri*, *Lb. reuteri*  
388 and *Lb. johnsonii* have either pseudogenes or a low abundance of pseudogenes, which is  
389 likely the genetic basis supporting survival of these species in the gut environment  
390 (O'Sullivan et al., 2009).

391 Efforts have been made to find at least a partial correlation between genome characteristics  
392 and niche for such a versatile group as *Lactobacillus*. The study of O'Sullivan et al. (2009)  
393 compared the genomes of 11 LAB (ten *Lactobacillus* and one *Streptococcus thermophilus*)

394 arising from different sources. In total, nine genes were identified as niche determinative as  
395 they insured survival in the gut or dairy environments. These genes were grouped into four  
396 classes that could be used as niche-specific genes for gut and dairy LAB: sugar metabolism,  
397 the proteolytic system, restriction/modification systems and bile salt hydrolysis. In contrast to  
398 this study, Kant et al. (2011) did not reveal any niche-specific genes in a study that analysed  
399 20 genomes of 14 different *Lactobacillus* species. The possible cause of this observation is  
400 that the isolation source does not always correspond to the actual habitat, but rather a  
401 transient habitat (Fig. 1b), as some species, like *Lb. plantarum* can be found in various  
402 environments (Kant et al., 2011).

403 Correlation between gene loss and niche adaptation was examined by growing nine *Lb. casei*  
404 strains from various isolation sources in chemically defined amino acid media supplemented  
405 with one of the substrates representing plant, gut or dairy habitats (Broadbent et al., 2012).  
406 The two cheese specialists had the most restricted substrate profiles, with no genes for inulin,  
407 sucrose or cellobiose utilisation present in their genomes, while the other strains used a  
408 higher number of different substrates, with corn silage isolates growing on 26 different  
409 substrates (Broadbent et al., 2012). In the study of Smokvina et al. (2013), niche affinity of  
410 *Lb. paracasei* was examined through utilisation of carbon sources as growth factors for a set  
411 of strains with diverse origins: plant, mammalian and dairy. The analysis revealed the  
412 clustering of seven out of the 16 dairy isolates that could be considered as niche specialists,  
413 which had smaller genomes compared to the others (2.8 Mbp average), limited numbers of  
414 sugar cassettes and an absence of genes involved in utilisation of plant-derived sugars. This  
415 was expected, as the spectrum of sugars in the dairy environment is narrow with lactose  
416 dominating. On the other hand, no clear clustering pattern was revealed for plant and  
417 mammalian isolates. Plant isolates originate from a broad range of ecosystems that differ in  
418 environmental and nutritional conditions, while mammalian isolates come from the gut where

419 they are exposed to constantly changing surroundings due to the presence of food and other  
420 microorganisms, and this complicates their precise grouping (Smokvina et al., 2013).  
421 Lactobacilli occupy habitats that differ considerably in environmental conditions. The dairy  
422 niche bacteria have to be robust enough to survive manufacture and storage conditions  
423 encountered during industrial production. In the gut, strains need to be able to survive in the  
424 presence of other intestinal microbiota and resist bile salts and other harsh conditions found  
425 in the gut (Senan et al., 2014). A genome-scale study based on genes involved in stress  
426 responses of the *Lb. helveticus* strains MTCC 5463 (probiotic strain isolated from a vaginal  
427 swab of a healthy volunteer, Senan et al. (2015)) and DPC4571 (a dairy isolate, Callanan et  
428 al. (2008)) gave an insight into genes responsible for adaptation to various environments  
429 (Senan et al., 2014). When comparing these two genomes for the ability of the strains to  
430 survive in a bile-rich environment, it was shown that the MTCC 5463 genome exhibited  
431 multiple coding sequences for bile salt hydrolase (bsh). However, the cheese starter  
432 DPC4571, adapted to a dairy niche, displayed a total lack of active *bsh* genes. The probiotic  
433 strain is exposed to other gut microbiota and in constant competition for successful  
434 colonisation and available nutrients. In order to survive in these conditions, it carries a higher  
435 number of starvation-induced genes. By contrast, while the dairy strain possessed some genes  
436 for starvation proteins, such as phosphate starvation inducible stress-related protein, it was  
437 deficient in the gene for the carbon starvation protein CstA. Both strains carried a substantial  
438 number of genes that allow response to heat and cold shock, but the molecular chaperones  
439 were far more prevalent in the probiotic genome (Senan et al., 2014). Another study  
440 performed on *Lb. helveticus* strains confirmed loss of genes encoding mucus-binding proteins  
441 from strains adapted to the milk environment, but confirmed their maintenance in probiotic  
442 strain R0052, where they are essential for survival and residence of the strain in the gut  
443 (Cremonesi et al., 2012).

444 Another noteworthy conclusion regarding niche adaptability was made when genome  
445 sequences of two strains, *Lb. helveticus* DPC4571 and *Lb. acidophilus* NCFM, were  
446 compared. The remarkable level of identity of 98% for 16S rRNA sequences was observed.  
447 Additionally, 75% of ORFs in DPC4571 were found in NCFM, which confirmed a close  
448 relationship between the two strains that inhabit significantly different environments (milk  
449 and gut). The genetic differences between these two strains were examined and they  
450 explained the genetic basis for niche specialisation. It was shown that the dairy strain lacked  
451 many genes that were retained in the probiotic strain, such as PTS systems, cell-wall  
452 anchoring proteins and the already mentioned mucus binding proteins (Callanan et al., 2008).  
453 In the previously mentioned study that analysed 100 *Lb. rhamnosus* strains, interesting  
454 observations regarding niche adaptability and clustering were made. Most dairy isolates  
455 clustered together, while intestinal and probiotic strains shared similarities with other human  
456 isolates. When both the phenotypic and genomic data of each strain were joined, two geno-  
457 phenotypes were identified. Firstly, the strains in group A were characterised by the absence  
458 of SpaCBA pili, lactose, maltose and rhamnose metabolism all of which point to dairy  
459 adaptation. Secondly, group B strains were bile resistant, pili possessing and L-fucose  
460 utilising, all characteristics important for intestinal tract survival. Although isolates of the  
461 same origin could be found in both groups, cheese isolates mainly belonged to group A,  
462 while intestinal isolates belonged mainly to group B. Intestinal isolates in group A may have  
463 originated from the consumption of food and represent rather a transient flora, while isolates  
464 from group B represent typical GIT residents (Fig. 1b). Interestingly, vaginal and oral isolates  
465 shared geno-phenotype A, which suggests a connection with dairy isolates (Douillard et al.,  
466 2013). Another study attempted to link genotypes and carbohydrate utilisation profiles of 65  
467 *Lb. rhamnosus* strains isolated from diverse habitats, such as human, baby and goat feces,  
468 cheese and fermented milk (Ceapa et al., 2015). Genomic fingerprinting was performed by

469 amplified fragment length polymorphism (AFLP) genotyping, and 11 genotypic groups were  
470 determined. Although not seen as a strict rule, strains of the same origin clustered together.  
471 Some clusters contained strains from various origins, indicating that these strains frequently  
472 change habitats (Fig. 1b). Conversely, some clusters had members of a single isolation niche,  
473 such as dairy. Following on from this, 25 isolates that represent all 11 clusters obtained by  
474 AFLP were tested for the carbon sources they could potentially use. Based on 72 carbon  
475 sources, three metabolic groups were determined, with group A including strains that could  
476 use plant derived carbohydrates, group B including strains with no ability to use lactose and  
477 group C containing strains that could use various carbohydrates. Although group B had no  
478 ability to use lactose, some strains isolated from cheese did belong to this group, where they  
479 were present as non-starter flora and had a role in proteolysis in the later stages of ripening.  
480 Interestingly, there was no direct correlation between metabolic groups and niche isolation,  
481 but strains coming from the same AFLP cluster appeared in the same metabolic group. This  
482 work again confirms that origin of isolation gives only an indication of potential metabolic  
483 capacity of the strain, but other approaches also have to be employed to fully understand  
484 strain fitness. For example, *Lb. rhamnosus* strain HN001 is present as a cheese isolate, but it  
485 has the ability to use 53 different carbon sources, which contradicts the general tendency of  
486 niche specialists to use a more narrow range of carbohydrates indicating that this strain was  
487 most probably very recently introduced into cheese environment. On the other hand, strain  
488 ATCC 53103 (GG) which originated from the intestine, belongs to a metabolically specialist  
489 group, possibly because it was transferred from dynamic environment such as GIT to more  
490 stable industrial habitat, which may have led to the metabolic simplification (Ceapa et al.,  
491 2015).

492 Finally, the effect of niche adaptation could be seen even within different dairy products. In  
493 the multi locus sequence typing (MLST) study of 11 housekeeping genes in 245 *Lb.*

494 *helveticus* isolates from natural fermented products, particular branches of isolates could be  
495 associated with the dairy product from which they originated (koumiss group, tarag group  
496 and coumiss-tarag group). These results suggest that even ecological niches representing  
497 different dairy environments may impact evolution of *Lb. helveticus* strains because genetic  
498 relationships are generally correlated with the ecological niches (Sun et al., 2015b).



#### 499 **4. A genomic perspective on key dairy traits: flavour formation and phage resistance**

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

#### 507 **4.1 Diverse proteolytic and flavour formation abilities of dairy lactobacilli**

508 Flavour formation in dairy products is the result of a complex network of processes which  
509 ends in specific combinations of flavour compounds and aroma development. Three major  
510 processes contribute to flavour development: glycolysis, lipolysis and proteolysis (van  
511 Kranenburg et al., 2002; Smit et al., 2005; Settanni and Moschetti, 2010). Glycolysis refers  
512 mainly to the metabolism of lactose and citrate. While lactose, the primary milk sugar, is  
513 mostly metabolised to lactic acid, a proportion of it can be converted to flavour compounds  
514 such as diacetyl, acetoin, acetaldehyde, or acetic acid, depending on the organism (van  
515 Kranenburg et al., 2002). Certain organisms also have the ability to metabolise citrate. Citrate  
516 is generally metabolised to pyruvate, which can be further metabolised to acetoin in the final  
517 product (Medina de Figueroa et al., 2001; Mortera et al., 2013). Lipolysis in fermented milk  
518 products arises mainly from the activity of microbial lipolytic enzymes (Collins et al., 2003).  
519 Esterases hydrolyse hydrosoluble ester chains between 2 and 8 C atoms, and lipases are more  
520 active on longer ester chains (10 C atoms). Free fatty acids contribute to cheese flavour,  
521 particularly short and intermediate chain fatty acids, which represent the starting molecules  
522 for catabolic reactions resulting in the production of numerous flavour and aromatic  
523 compounds (Collins et al., 2003). Of all the metabolic processes responsible for flavour

524 development in dairy products, proteolysis is considered the most important and complex  
525 one, affecting texture, hardness, elasticity and the overall flavour of the fermented product  
526 (Savijoki et al., 2006). The proteolysis cascade starts with casein degradation by cell-  
527 envelope proteinases (CEP, Prt). The peptides released in this processes are then transported  
528 in the cell, where peptidases with varying specificities cleave them, releasing amino acids.  
529 These amino acids are the substrates for various metabolic reactions, with aminotransferases  
530 being the first enzymes in the subsequent catabolic cascade. Diverse and numerous aromas  
531 are released in these reactions (aldehydes, ketones, carboxylic acids and volatile sulfur  
532 compounds) (Marilley and Casey, 2004). In this section, the genomics of the components of  
533 proteolytic system of *Lactobacillus* will be discussed, as proteolysis represents a critical  
534 process in flavour development in dairy products.

535 Cell envelope proteinases (CEPs) are multi-subunit, cell wall associated proteinases and their  
536 main role during growth in milk is degradation of casein into smaller peptides (Sun et al.,  
537 2015a). The importance of surface proteinases is made clear in studies that showed that  
538 knock-out strains lack the ability to grow in milk (Mayo et al., 2010).

539 In an extensive study performed on the genomes of 213 *Lactobacillus* and associated genera,  
540 intriguing diversity in CEP characteristics was revealed (Sun et al., 2015a). In total, genes for  
541 60 CEPs were identified and presence of genes for CEPs was highly correlated with  
542 phylogenetic clades. Three different anchoring mechanisms were observed: a SLAP domain  
543 (S-layer anchoring domain) responsible for non-covalent interactions was present,  
544 particularly in the *Lb. delbrueckii* sub-clade; a LPXTG motif for covalent linkage to  
545 peptidoglycan and a derivative of the LPXTG motif. In thirteen cases, no anchoring domain  
546 for CEP was identified as sequences were terminated exactly before the typical start of the  
547 anchoring domain sequence. Multiple alignments indicated the sequences of these 13 CEPs  
548 differ from other CEPs along the entire length of the protein. Besides this, the possibilities of

549 various domain combinations in the CEPs enable a diversity of potential substrates to be  
550 utilised, resulting in a range of final products, which could contribute to improvement of  
551 dairy products flavour (Sun et al., 2015a).

552 The vast majority of LAB have only one CEP, but for certain strains of *Lb. helveticus*, it has  
553 been confirmed through multiplex PCR analysis that at least four different proteinases exist  
554 (Broadbent et al., 2011) and four *prt* genes were described in the genome of *Lb. helveticus*  
555 CNRZ32 (Broadbent et al., 2013). The presence of a higher number of proteinases with  
556 different substrate and cleavage specificities could explain the efficiency of the *Lb. helveticus*  
557 proteolytic system. CEPs have different and complimentary properties and some strains could  
558 have acquired additional genes because they provide an adaptive advantage regarding milk  
559 protein hydrolysis (Genay et al., 2009). In the study by Broadbent et al. (2011), 51 *Lb.*  
560 *helveticus* strains were tested for presence of *prt* paralogs. The distribution of *prt* genes  
561 varied among *Lb. helveticus* strains and the most abundant gene was *prtH3*, which contradicts  
562 the study by (Genay et al., 2009) who found that *prtH2* was in fact a ubiquitous gene in *Lb.*  
563 *helveticus* strains. The reasons for this contradiction are that sequences for *prtH4* were not  
564 available, and *prtH3* gene from DPC4571 strain was described as an allele of *prtH2*  
565 (Broadbent et al., 2011). From the dairy industry perspective, the diverse proteinase gene  
566 content in *Lb. helveticus* may be a crucial factor in determining the function and behaviour of  
567 these strains with regard to desired flavour formation (Broadbent et al., 2011).

568 The correct maturation of CEP depends on the presence of the maturation proteins, PrtM. For  
569 instance, while *Lb. helveticus* CNRZ32 has 2 *prtM* paralogs designated as *prtM* and *prtM2*, in  
570 other analysed *Lb. helveticus* strains *prtM* was found only in strains that possessed *prtH*, and  
571 *prtM2* was encoded in genomes of all tested strains. It has been proposed that *prtM* is needed  
572 for activation of *prtH*, and *prtM2* is responsible for folding and activation of other *prt*  
573 paralogs (Broadbent et al., 2011). On the other hand, no *prtM* gene for this protein was found

574 in any of the 4 completely sequenced *Lb. delbrueckii* strains (Liu et al., 2012). However, the  
575 foldase protein (PrsA) involved in maturation of extracellular proteinase and folding and  
576 stability of subtilisins in *Bacillus subtilis* was detected. PrsA might be involved in maturation  
577 of PrtB, as PrsA from four *Lb. delbrueckii* strains were homologous with known PrtM  
578 proteins (Liu et al., 2012).

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

588 The diversity in peptidase content is observed on the same species level, where strains differ  
589 in numbers of peptidases and transport system components. Upon analysis of four fully  
590 sequenced genomes of *Lactobacillus delbrueckii* (ATCC 11842, BAA-365, 2038 and ND02),  
591 strain ND02 possessed the highest number of proteinase and peptidase genes, as well as the  
592 highest number of peptidase and amino acid transport systems. Intracellular peptidases  
593 showed some differences between the four strains, such as three unique peptidases in strain  
594 ND02. In the case of strain 2038, two cell surface peptidases En1A and Pep-D4 were present  
595 as complete genes, indicating that this strain has a more powerful proteolytic capability and  
596 potentially produces more free amino acids than the other strains (Liu et al., 2012). All four  
597 sequenced strains possessed two complete Opp systems, but they differed in numbers and  
598 organisation of substrate binding protein OppA. The highest number of OppA genes was

599 found in the industrial strain 2038 and their products enable transport of different  
600 oligopeptides (Liu et al., 2012).

601 The next step in the protein degradation cascade is the metabolism of free amino acids,  
602 following which a large number of flavour compounds arise. Aminotransferases are the first  
603 enzymes in the cascade, transferring amino groups from amino acids to alpha-keto acids,  
604 most often alpha-keto glutarate. In a comparative study of enzymes involved in amino acid  
605 metabolism contributing to generation of flavour compounds in 21 genomes of different LAB  
606 species, (12 of which were lactobacilli), a homolog of the *bcaT* gene, coding for branched-  
607 chain aminotransferase activity, was present in all *Lactobacillus* strains considered as  
608 important in dairy production, while a larger number of homologs for the *araT* gene, coding  
609 for aromatic aminotransferase activity, were usually present (Liu et al., 2008). The  
610 distribution of amino acid metabolising enzymes amongst starter and NSLAB including the  
611 species discussed in this review, were compared by Gobbetti et al. (2015), and it confirmed  
612 the diversity of the metabolic capability of lactobacilli and underlined the importance of  
613 genomic analysis as part of a knowledge-based approach to strain selection.

614 Cysteine and methionine are precursors for the production of volatile sulfur compounds  
615 (VSCs) which are important flavour compounds that are found in many cheese varieties. The  
616 metabolism of sulfur containing amino acids is complex as multiple alternative metabolic  
617 pathways exist (Mayo et al., 2010). One of the enzymes involved in metabolism of  
618 methionine is cystathionine gamma lyase (CGL), which was found in several *Lb. casei* strains  
619 isolated from cheese and milk (Irmeler et al., 2008). Two variants of the gene encoding CGL  
620 shared 81% of similarity and were named *ctl1* and *ctl2*. Homologs of *ctl1* and *ctl2* were found  
621 in other LAB: *Lb. helveticus*, *Lb. bulgaricus* *Lb. rhamnosus* and *S. thermophilus*, but they  
622 were not present in three publicly available genomes of *Lb. casei* (ATCC 334, Zhang and  
623 BL23) and it is likely that these strains uptake sulfur-containing peptides and amino acids

624 from the environment (Irmeler et al., 2009). Analysis of nucleotides upstream from a *ctl* gene  
625 cluster found an ORF encoding for a putative transposase, supporting the possibility of  
626 horizontal transfer of the cluster to *Lb. casei* strains. The gene cluster forms an operon  
627 important in cysteine biosynthesis, as its expression was downregulated when L-cysteine is  
628 added to the medium (Bogicevic et al., 2012). Furthermore, when these strains were used in  
629 cheese production, significantly higher levels of VSC were detected at the end of ripening  
630 (Bogicevic et al., 2013).

631 Glutamate dehydrogenase (GDH) is an enzyme that acts as a cofactor for aminotransferase  
632 function, as it enables recycling of alpha-ketoglutarate, the receptor of the amino group  
633 during transamination. When genomes of 12 species of *Lactobacillus* were analysed, the  
634 presence of a *gdh* gene was confirmed only in *Lb. plantarum* WCFS1 and *Lb. salivarius*  
635 UCC118 (Liu et al., 2008), which agrees with the strain dependency of *gdh* presence and  
636 higher prevalence in natural strains commonly found in cheese manufacture (Tanous et al.,  
637 2002). However, the majority of *Lb. casei*, *Lb. rhamnosus* and *Lb. plantarum* genomes  
638 possess the *gdh* gene (Gobbetti et al., 2015), but no *gdh* gene was found in any of the  
639 sequenced *Lb. delbrueckii* strains (Liu et al., 2012; Gobbetti et al., 2015). Nevertheless, two  
640 genes encoding proteins homologous to aspartate aminotransferase were found in *Lb.*  
641 *delbrueckii* and which could potentially catalyse the formation of glutamate from 2-  
642 oxoglutarate and L-aspartate (Liu et al., 2012).

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

#### 649 **4.2 CRISPR regions of dairy-related lactobacilli**

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

654 Several mechanisms of phage resistance were previously described for lactic acid bacteria  
655 and they include prevention of phage adsorption, blocking the entry of phage DNA, cutting  
656 phage nucleic (restriction/modification systems) acid and abortive infection (Garneau and  
657 Moineau, 2011). However, recently, a new system that enables effective resistance to phage  
658 attacks was discovered, and it was shown that this system was almost universally present in  
659 bacteria, including LAB. CRISPR (clustered regularly interspaced short palindromic repeats),  
660 together with CRISPR-associated genes (*cas*) form a bacterial immune system against foreign  
661 DNA, such as phage or plasmids (Barrangou and Horvath, 2012). The typical CRISPR locus,  
662 located behind the leader sequence, contains a string of DNA repeats and spacers, which  
663 represent short sequences corresponding to foreign DNA inserted between two repeats  
664 (Deveau et al., 2010). The efficient defence from foreign DNA attack involves the  
665 incorporation of short sequences of foreign DNA in CRISPR loci (acquisition) (Fig. 2a). In  
666 the event of foreign DNA being present in the cell, these short sequences are transcribed into  
667 small interfering RNAs, called CRISPR RNA (crRNA), which guide multifunctional protein  
668 complexes to recognise and cleave matching foreign DNA (Fig. 2b) (Barrangou and Horvath,  
669 2012).

670 Two genes, *cas1* and *cas2*, are regularly present in CRISPR-Cas systems, and they are  
671 involved in the acquisition process (Barrangou, 2013). Based on the signature genes which  
672 confer interference, three types of CRISPR-cas systems are well described. Type I systems  
673 have *cas3* as the signature gene, which encodes an endonuclease involved in the cleavage of

674 DNA. Another feature of this type is the Cascade complex, participating in processing of  
675 crRNA and recognition of target DNA. The signature gene of Type II systems is *cas9*, which  
676 encodes a protein important for the crRNA synthesis and target DNA cleavage. Specificity of  
677 Type II systems is trans-activating CRISPR RNA (tracrRNA) that hybridizes to crRNA and  
678 enables its maturation by endoribonuclease RNase III. Type III systems are defined by the  
679 signature gene *cas10* and they are mechanistically diverse, with IIIA systems cleaving DNA  
680 and IIIB systems cleaving RNA molecules (Barrangou, 2013; Selle and Barrangou, 2015).  
681 Besides these three systems, novel types (IV, V and VI) were discovered more recently  
682 (Wright et al., 2016).

683 In LAB, eight different families of CRISPR loci were found and these families did not  
684 correlate with phylogeny of LAB indicating their independent evolution from other elements  
685 on the chromosome. The analysis of CRISPR loci at the level of the LAB showed that highly  
686 similar loci were found in distant genera and species. This could be explained by HGT and  
687 indeed, these loci have different GC content compared to the rest of the host genome.

688 Interestingly, the comparison of CRISPRs of two closely related species, *Lb. helveticus* and  
689 *Lb. casei*, showed that they belong to different families, once again confirming the high level  
690 of variability of these regions (Horvath et al., 2009).

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



698 CRISPR profiles of 100 *Lb. rhamnosus* strains were generated by spacer oligotyping, a  
699 method firstly described by Kamerbeek et al. (1997), and a considerable level of strain  
700 variety was revealed (Douillard et al., 2013). Additionally, in certain cases, correlation  
701 between CRISPR loci and specific niche was observed. In total, 24 spacers were identified  
702 from both plasmids and phage DNA. Spacers that corresponded to phages belonged to *Lb.*  
703 *rhamnosus* phages or *Lb. casei* phages. The study defined two general geno-phenotypes  
704 (discussed above) and the CRISPR locus profiles were substantially different in these two  
705 groups (Douillard et al., 2013). A comparative study of CRISPR in *Lactobacillus delbrueckii*  
706 ssp. *bulgaricus* that took into consideration 33 strains showed that these strains possessed  
707 either Type II or Type III CRISPR systems (Urshev and Ishlimova, 2015). However, in the  
708 genome of recently sequenced strain CFL1 both CRISPR types (II and III) were present  
709 simultaneously (Meneghel et al., 2016).

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

723 resistant to phage attack supports its wide and successful commercial application (Bull et al.,  
724 2014).

#### 725 **4.2.1 Applications of CRISPR systems**

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

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

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

748 crRNA and tracrRNA that leads the Cas9 to a specific DNA site (Fig. 2c). The DNA break  
749 can be followed by non-homologous end joining which induces indels, or homology-directed  
750 repair that introduces site-specific insertion from DNA donor templates (Sander and Joung,  
751 2014). This simple and highly specific approach has moved the boundaries of genetic and  
752 biochemical research, and it is almost ideal for genome editing applications due to its  
753 efficiency and affordability (Selle and Barrangou, 2015).

## 754 **5. Genome scale metabolic models and metabolic engineering of *Lactobacillus* species**

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

770 The metabolic network of an organism is based on genomic information, and this network  
771 connects the information of genes and the metabolic reactions they are involved in (Lewis et  
772 al., 2012). After detailed revision and correction of the (genome-scale) metabolic model, it is  
773 then transformed to a stoichiometric matrix, which is a mathematical representation of  
774 metabolic reactions. The purpose of this step is to convert GSMM to a computational one  
775 (O'Brien et al., 2015). Constraint-based reconstruction and analysis (COBRA) models are the  
776 most widely used in GSMM analysis (Lewis et al., 2012). Flux Balance Analysis (FBA) is  
777 the oldest, most basic and commonly used COBRA method (Lewis et al., 2012; O'Brien et  
778 al., 2015; Orth et al., 2010) for simulating GSMM. Detailed explanation of how FBA

779 operates can be found in Orth et al. (2010). Flux variability analysis (FVA), introduced by  
780 Mahadevan and Schilling (2003), modifies the FBA approach as it considers the effect of  
781 metabolic uncoupling. FVA determines, for each reaction in the model, the range of possible  
782 fluxes that correspond to experimental values of constraints (Smid and Hugenholtz, 2010).  
783 *Lc. lactis* was the first LAB to have a genome-scale model constructed (Oliveira et al., 2005),  
784 followed by *Lb. plantarum*, (Teusink et al., 2006) and *Streptococcus thermophilus* (Pastink et  
785 al., 2009) and most recently, *Lb. casei* (Vinay-Lara et al., 2014; Xu et al., 2015). Here, we  
786 will review the most important findings of models designed for some species of  
787 *Lactobacillus*.

788 The GSMM of *Lb. plantarum* WCFS1 was used to compare a traditional view of ATP  
789 production from lactate and acetate and ATP production based on the constraints approach  
790 when experimental constraints were applied. The traditional approach has certain  
791 disadvantages as it takes into account lactate and acetate production in other metabolic  
792 processes which do not contribute to ATP yield, like amino acid or citrate metabolism. After  
793 comparison of ATP production in both approaches, the same result was obtained in both  
794 cases, meaning that the effects of amino acid and citrate metabolism were not crucial.

795 Additionally, the model identified catabolic reactions such as transamination of aromatic and  
796 branch-chained amino acids to generate ATP. These reactions are seen as a major factor in  
797 flavour development, but have not been previously connected with ATP production. Further  
798 on, the model attempted to assess the effect of uncoupling on metabolic capacities. FVA was  
799 used to calculate the spectrum of flux values consistent with the experimental constraints and  
800 showed higher flexibility of the flux ranges for the uncoupled energy production and  
801 consumption. However, FBA was not able to correctly predict *Lb. plantarum* biomass  
802 production, as it did not take into account inefficient lactate production. FBA predicted  
803 higher growth, as it detected lactate production as incompatible with optimised growth. In

804 reality though, *Lb. plantarum* produces lactate and tends to utilise a route that is less efficient  
805 even under limited energy conditions, and this event cannot be predicted by FBA, which  
806 proposed higher yield as a result of mixed acid fermentation (Teusink et al., 2006).  
807 The study by Vinay-Lara et al. (2014) compared metabolic networks from two *Lb. casei*  
808 strains that are fully sequenced, ATCC 334 and 12A. FBA was used to analyse the properties  
809 and capabilities of both models. Both tested strains have similar amino acid requirements -  
810 branched-chain and aromatic amino acids and arginine are essential. It is most likely that the  
811 rich environment (cheese and corn silage) that these strains were isolated from reduced the  
812 need for synthesising all amino acids. Although models initially did not predict glutamate as  
813 an essential amino acid, excluding this amino acid from the culture medium significantly  
814 reduced the growth of ATCC 334 and resulted in no growth for 12A. However, in both  
815 metabolic models glutamine can be converted into glutamate, and the experimental studies  
816 suggested that this interconversion of glutamate to glutamine results in low yields of  
817 synthesised glutamate, thus explaining why glutamate is needed even in the presence of  
818 glutamine. A correction of the metabolic pathway was possible in the case of ATCC 334, but  
819 fixing the inconsistency in 12A was not successful, and the model was not unable to  
820 determine the strain's requirements for glutamate. Carbohydrate utilisation analysis of these  
821 strains once again confirmed the hypothesis of gene decay during adaptation to nutrient rich  
822 environments. Strain 12A, isolated from corn silage (Cai et al., 2007) possesses an ABC  
823 transporter for uptake of raffinose and enzymes needed for pullulan and panose degradation,  
824 sugars frequently present in plant material. On the contrary, ATCC 334, a cheese isolate,  
825 lacks these genes as they are most likely redundant in the dairy environment. Interestingly,  
826 the metabolic model for strain 12A shows that all the genes for converting myoinositol to  
827 glyceraldehyde-3-phosphate are present. Myoinositol can be used as phosphate storage  
828 molecule in plants. Although the majority of LAB cannot use this sugar as carbon source,

829 strain 12A has all the genes needed for conversion of myoinositol, but this metabolic pathway  
830 is not active in 12A probably due to regulatory effects (Vinay-Lara et al., 2014). In other *Lb.*  
831 *casei* models it was shown that, *in silico* growth of *Lb. casei* LC2W was improved by  
832 myoinositol under aerobic conditions, suggesting that this strain could utilise energy sources  
833 that seemed inappropriate under anaerobic conditions (Xu et al., 2015).

834 A genome-scale metabolic model of *Lb. casei* LC2W was used for the analysis of the oxygen  
835 effect on flavour compound synthesis and three new *in silico* knockout targets were selected  
836 for acetoin production. In *Lb. casei* LC2W, the main precursor of flavour compounds is  
837 alpha-acetolactate. Acetoin and diacetyl are produced from alpha-acetolactate by  
838 acetolactate-decarboxylase or through non-enzymatic processes. Although acetoin could  
839 accumulate in LC2W in both aerobic and anaerobic conditions, production of diacetyl was  
840 dependent on oxygen and it was possible to maintain diacetyl production at a high level with  
841 the increase of oxygen uptake. Additionally, FBA suggested three new *in silico* knockout  
842 targets for acetoin production: dihydrofolat-reductase, methylen-tetrahydrofolate-  
843 dehydrogenase and glycerol-phospho-transferase (Xu et al., 2015).

844 Regarding the flavour potential of LAB, a completely different approach was recently  
845 proposed. As seen, GSMM contain numerous gaps which cannot always be completed.  
846 Although there are many known pathways involved in flavour formation, the overall process  
847 of flavour development is highly complex. Compounds that are often seen as flavour  
848 contributors are products of amino acid metabolism: alcohols, aldehydes and acids, and  
849 especially sulfur compounds, products of methionine metabolism (Curioni and Bosset, 2002;  
850 Smit et al., 2005; Yvon, 2006). Reverse pathway engineering (RPE) (Liu et al., 2014) takes  
851 small molecules as a starting point and looks for enzymatic or chemical reactions that can  
852 track these compounds back to the known precursors. This method was used in LAB to  
853 predict so far unknown reactions in metabolic pathways by combining retrosynthesis and

854 genomic information. To confirm that the proposed approach is correct, the relatively well-  
855 known pathway of leucine degradation in LAB was tested in the model. Not only were the  
856 main branches confirmed, but it also suggested a novel route of generating 3-methyl  
857 butanoic acid, one of the most important flavour compounds of leucine metabolism. This  
858 novel route starts with the transamination product of leucine, alpha-keto-isocaproate, which is  
859 further reduced to alpha-hydroxy-isocaproate. The second step suggests formation of 3-  
860 methyl butanoic acid from alpha-hydroxy-isocaproate, and the related reaction found in the  
861 database was a lactate oxidation reaction catalysed by lactate-2-monoxygenase (LOX), so it  
862 was assumed that LOX could possibly catalyse oxidation of alpha-hydroxy-isocaproate.  
863 Broader activity of LOX seems to be dependent on the amino acid at position 95 and it could  
864 be obtained if alanine in position 95 was mutated to glycine (detailed explanation in Liu et  
865 al. (2014)). The RPE method also revealed a non-enzymatic reaction of converting alpha-  
866 keto-isocaproate to 2-methyl propanal, and this reaction connects valine and leucine  
867 catabolism. Regarding the methionine degradation, RPE discovered an enzymatic reaction  
868 responsible for the conversion of methanethiol to dimethyl-sulfide (DMS), using DMS as an  
869 input. Enzymes homocystein-S-methyltransferase, methionine synthase and thiol-S-methyl-  
870 transferase were proposed using the bioinformatics approach. The prediction of novel  
871 reactions using RPE opens up new possibilities for metabolic engineering. For example,  
872 hydroxy-isocaproate is often seen as an off-flavour in cheese products, but the proposed  
873 conversion to the flavour compound 3-methyl butanoic acid could be implemented in novel  
874 strategies for production of flavour by utilising off-flavours as precursors (Liu et al., 2014).

### 875 **5.1 Metabolic engineering as a future application of lactobacilli**

876 A vast amount of knowledge on genetics and metabolism of LAB opened the door for  
877 implementation of LAB in novel biotechnological applications (Gaspar et al., 2013).

878 Application of LAB is not limited only to classical food fermentation and the use of LAB as



879 cell factories is expected to increase (Gaspar et al., 2013). LAB are characterised by limited  
880 biosynthetic capacity and metabolic versatility and their physiology is relatively simple. They  
881 are characterised by relatively small genomes (2-3 Mbp), fast growth, high sugar uptake rates  
882 and less high-level control systems, all of which make them suitable candidates for metabolic  
883 engineering (Papagianni, 2012; Gaspar et al., 2013). Genetic engineering made possible the  
884 production of molecules not natively present in the host, but also enabled engineering of  
885 native genes (Keasling, 2012). Genetic engineering proved successful in the development of  
886 strains producing recombinant proteins and small chemicals, but development of tools that  
887 exceed genetic engineering is needed, as some molecules are synthesised in multiple  
888 reactions (Bution et al., 2015). Metabolic engineering summarizes previous knowledge  
889 regarding cell metabolic features and it uses molecular tools to deliberately change cellular  
890 metabolism for the purpose of the efficient production of target molecules (Bution et al.,  
891 2015) (Fig. 3). However, the host cell needs to meet several requirements to ensure efficient  
892 metabolic engineering occurs. Host cells should be genetically stable, not interfering with  
893 heterologous genes on the introduced vector, and have optimal traits for industrial  
894 applications. Apart from these, genomic information can help in the choice of host, as new  
895 pathways can induce stress response and impede gene expression (Keasling, 2012).

896 Metabolic engineering of lactic acid bacteria presents a novel approach for re-routing  
897 metabolic reactions in LAB so specific and desired compounds are produced in higher  
898 amounts. Several different types of molecules can be produced by LAB as cell-factories:  
899 lactic acid, flavour compounds (diacetyl, acetaldehyde), sweeteners (L-alanine, mannitol,  
900 sorbitol, xylitol), exopolysaccharide, vitamins etc. (Papagianni, 2012). Historically, the first  
901 attempt of engineering of LAB was oriented towards improving production of the bitter  
902 aroma compound diacetyl in *Lc. lactis*. Subsequently, many other studies expanded the  
903 species of LAB that were subject to engineering as well as the types of molecules produced.

904 Several recent review articles (Papagianni, 2012; Gaspar et al., 2013; Mazzoli et al., 2014)  
905 give detailed information about achievements in production of industrially important  
906 compounds in LAB. Production of food ingredients, commodity compounds, vitamins and  
907 ethanol are thoroughly reviewed with methods of engineering and future perspectives  
908 anticipated. Besides this, metabolic engineering is used as a tool for improvement of  
909 adherence and immunomodulatory properties of probiotic strains (described and reviewed in  
910 Yebra et al. (2012)). While most of results come from *Lc. lactis* as most widely used LAB,  
911 novel information comes from *Lactobacillus* species as well. Here we review studies  
912 performed on strains of *Lactobacillus* spp. mainly associated with dairy food.

913 Lactic acid is used as a preservative and flavour enhancing agent by the food industry, and  
914 also in cosmetic and pharmaceutical industries (Papagianni, 2012). In addition, L-lactic acid  
915 is used as the starting material in the production of biopolymers (Gaspar et al., 2013). Unlike  
916 chemical synthesis, which often leads to racemic mixture of L- and D-lactic acid, microbial  
917 fermentation can be optimised for production of a single enantiomer (Gaspar et al., 2013).

918 The L-isomer is a preferred for two reasons: D-isomer is not metabolised in humans and has a  
919 toxic effect and L-isomer polymerises which is important in polymers production (Kyla-  
920 Nikkila et al., 2000; Papagianni, 2012). The initial attempts to influence lactic acid  
921 production in lactobacilli date in 1990's, when the enhancement of L-lactic acid was achieved  
922 by the inactivation of *ldhD* in *Lb. helveticus* (Bhowmik and Steele, 1994), but the  
923 overexpression of *ldhL* in *Lb. plantarum* did not cause an increase of L-lactic acid synthesis,  
924 although increased activity of L-LDH was observed (Ferain et al., 1994). More recently,  
925 selective L-lactate production was tested in *Lb. helveticus* CNRZ32 and two approaches were  
926 used (Kyla-Nikkila et al., 2000). The promoter of the *ldhD* gene was deleted in the construct  
927 GRL86 while in the other construct, GRL89, the structural gene of *ldhD* was replaced with an  
928 additional copy of the structural gene of *ldhL*. Both constructs produced only L-lactic acid in

929 amounts that were on the level of total lactate produced by the wild type strain and no  
930 difference in growth profiles for either construct was observed compared to the wild strain.  
931 Additionally, the L-lactic acid production phase of mutant strains was prolonged compared to  
932 the wild strain (Kyla-Nikkila et al., 2000).

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

953 Sorbitol is a sugar alcohol largely used in the food industry as a sweetener (Gaspar et al.,  
954 2013). It is poorly absorbed in small intestine and as it has low calorie value, is used in  
955 diabetic appropriate foods (Ladero et al., 2007), but also as a softener and texturing agent  
956 (Yebara et al., 2012). An attempt to construct sorbitol-producing LAB was performed by  
957 introducing the *gutF* gene coding for sorbitol-6-phosphate-dehydrogenase, into the *lac*  
958 operon of *Lb. casei*. The strain with the integrated *gutF* was named BL232 and the expression  
959 was controlled as in other *lac* genes. Additionally, a L-lactate-dehydrogenase (*ldhL*) knockout  
960 of BL232 was constructed, and designated as BL233. Resting cells of both of these strains  
961 produced sorbitol from glucose, and the *ldhL* knockout showed higher production of sorbitol  
962 compared to BL232. It was proposed that *ldhL* inactivation leads to a higher NADH/NAD<sup>+</sup>  
963 ratio and the cell uses this for the sorbitol production (Nissen et al., 2005). In further studies,  
964 metabolic engineering of *Lb. casei* led to a strain that could produce sorbitol without  
965 consequent uptake after glucose exhaustion, by introducing a mutation in the sorbitol-specific  
966 phospho-transferase system. Sorbitol producing *Lb. casei* were constructed through a series  
967 of transformations of strain BL232: deletion of *ldh1* gene encoding the main lactate-  
968 dehydrogenase (BL251) followed by deletion of *gutB* gene (BL283) involved in transport of  
969 sorbitol and subsequent mutation of the mannitol-1-phosphate-dehydrogenase (*mtlD*) gene  
970 (BL300). While mutant BL251 used sorbitol after glucose consumption, BL283 was not able  
971 to transport sorbitol and levels of sorbitol did not drop after glucose exhaustion. To avoid  
972 synthesis of mixed polyols (sorbitol and mannitol, as occurred in the study of Nissen et al.  
973 (2005)), a gene encoding mannitol-1-phosphate dehydrogenase was inactivated (BL300) and  
974 this knockout strain did not produce mannitol, and sorbitol production was doubled compared  
975 to BL283. In addition, the resting cells of BL300 were able to produce sorbitol from lactose  
976 in 1% supplemented MRS, especially at pH 5.5 and 4.75, but this conversion was less  
977 efficient than the conversion of glucose. Additionally, BL300 cells were able to produce

978 sorbitol as a sole polyol from whey permeate, a by-product of the dairy industry (De Boeck et  
979 al., 2010).

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

989 Succinic acid is a starting block in synthesis of biodegradable plastic (Babu et al., 2013) and  
990 can be used as a food additive (Beauprez et al., 2010). In a study by Tsuji et al. (2013),  
991 production of succinic acid was examined in the previously described lactate-dehydrogenase  
992 deficient strain *Lb. plantarum* VL103. Three enzymes involved in succinic acid production:  
993 pyruvate-carboxylase (PC), phospho-enol-pyruvate (PEP) and malic enzyme (ME) were  
994 overexpressed in this strain, and all transformants showed increased activity of the  
995 corresponding enzyme, up to 2.4 fold in the case of PC. However, although PC  
996 overexpression was the most effective for succinic acid production in *Lb. plantarum*, a  
997 mutant with PEP enzyme overexpressed, exhibited a higher specific growth rate, compared to  
998 the two others, and seemed a better candidate for LAB succinic acid production, as PC  
999 overexpression was effective but slowed down the growth rate. Additionally, combined levels  
1000 of succinic acid production were observed in mutants displaying overproduction of the two  
1001 enzymes and the co-expression of PC and PEP increased succinic acid yield and biomass  
1002 (Tsuji et al., 2013).

1003 Engineered *Lb. casei* were used to increase the production of diacetyl and acetoin from whey  
1004 permeate (Nadal et al., 2009). These two compounds have a buttery flavour and are used as  
1005 additives in the food industry (Yebra et al., 2012). The presence of the lactococcal aceto-  
1006 hydroxy-acid synthase (*ilvBN*) gene and deletion of lactate-dehydrogenase gene (*ldh*) resulted  
1007 in an increase in diacetyl/acetoin synthesis from glucose, but strain with only *ldh* deletion  
1008 showed a similar result. By contrast, when the bacterial cells were exposed to lactose, strains  
1009 carrying the *ilvBN* gene showed four times higher production of the desired compounds. The  
1010 strain containing *ilvBN* and *ldh* mutations and a strain with additional *pdhC* (gene coding the  
1011 E2-dihydrolipoamide-acetyl-transferase, component of pyruvate-dehydrogenase complex  
1012 Pdh) mutation were used for whey permeate fermentations. Having found the most suitable  
1013 conditions for pH, the total amount of diacetyl/acetoin production was higher for the strain  
1014 with the *pdhC* mutation. Fed batch experiments with this strain were done with the addition  
1015 of whey permeate and yeast extract, but no further increase in diacetyl/acetoin concentrations  
1016 was observed, and it was proposed that higher concentrations of product might have  
1017 inhibitory effect. However, the amount of product obtained was still lower compared to  
1018 engineered *Lc. lactis* (Nadal et al., 2009).

1019 Exopolysaccharides (EPS) have been widely used in food industry, as they impact on the  
1020 texture of food products, but they have also been shown to possess prebiotic characteristics  
1021 (Papagianni, 2012). The EPS production levels in LAB are relatively low, and there have  
1022 been several attempts to increase its production, mainly in *Lc. lactis* (for review see Gaspar et  
1023 al. (2013)). In an attempt to increase EPS production in *Lb. casei*, the effects of cofactors  
1024 involved in EPS biosynthesis were investigated. The gene encoding NADH-oxidase (*nox*),  
1025 from *Streptococcus mutans*, was cloned and overexpressed in *Lb. casei* LC2W. The strain  
1026 obtained grew slower than the wild type, but showed 46% increase in EPS production (Li et  
1027 al., 2015b). Furthermore, several other genes believed to be involved in EPS production were

1028 chosen from different *Lactobacillus* strains (*Lb. plantarum*, *Lb. casei* and *Lb. rhamnosus*) and  
1029 their effect on EPS biosynthesis was tested. The genes *tga* (trans-glutaminase), *pfk* (phospho-  
1030 fructokinase), *pgm* (phospho-glucomutase), *galtf* (galacto-transferase), *rhatf* (rhamnosyl-  
1031 transferase), *rfbB* (dTDP-glucose-4,6-dehydratase) and *galt* (galactose-1-phosphate-urydil-  
1032 transferase), and previously described *nox* (NADH-oxidase), all involved in various steps of  
1033 EPS production were successfully cloned and overexpressed in *Lb. casei* LC2W. Although  
1034 recombinant strains had slower growth rates, some of them showed the positive effect of  
1035 overexpressed genes (*pfk*, *rfbB* and *galt*) on EPS production (Li et al., 2015a), but lower than  
1036 for the previously described *nox* mutant. Besides that, the *nox*-mutant was shown to produce  
1037 EPS in higher amounts in aerobic conditions, although growth was less than in anaerobic  
1038 conditions. In aerobic conditions, the strain with overexpression of NADH oxidase reduced  
1039 used more NADH and produced lower amounts of lactate, all of which led to the increased  
1040 EPS production (Li et al., 2015a).

1041 The question remains, however, would engineered bacteria be acceptable for direct use in  
1042 food production. According to the current EU legislation (Directive 2009/41/EC of the  
1043 European Parliament and of the Council), a genetically modified microorganism (GMM) is  
1044 any microorganism that has foreign DNA introduced in a way that does not occur naturally.  
1045 Many of these modified bacteria could potentially be used in dairy food production where  
1046 they could contribute to flavour and texture or fermented products containing these LAB  
1047 could be used as a vehicle for probiotic delivery. However, these foods would have GMO  
1048 status and fall under specific legislation, and guidelines for their applications have been  
1049 proposed (European Food Safety Authority, 2011). It also raises issues in applicability and  
1050 market potential as well as consumer acceptance of the modified LAB and careful analysis of  
1051 variations in legislatives as well as possibilities and limits in applying genetically modified  
1052 LAB in food, mainly in regard to consumers risk and benefits, should be taken into

1053 consideration (Pedersen et al., 2005; Sybesma et al., 2006). In addition, new approaches of  
1054 genome editing with employment of CRISPR-cas system would not be seen as GMM-  
1055 generating tools according to the current definition, as it was recently discussed in case of  
1056 genetically edited crops (Kanchiswamy et al., 2015), as only oligonucleotides that correspond  
1057 to native molecules are needed for this reaction and the complex that derives edition is further  
1058 degraded in the cell. This opens questions about redefining GMM and their use in the food  
1059 industry. One issue that has to be considered is the fact that although the CRISPR systems  
1060 have a high specificity level, the problem of unexpected negative effects remains a  
1061 possibility, which could have massive effect on global food market (Au, 2015).

1062 On the other hand, less restriction embraces the usage of modified LAB as potential cell  
1063 factories. The era of application of recombinant bacteria for molecules started with human  
1064 insulin production by recombinant *E. coli* developed in late 1970's (Goeddel et al., 1979). In  
1065 general, LAB are recognised as safe and non-pathogenic, which makes them suitable for  
1066 engineering projects. Even though these cells are engineered, the final product is purified and  
1067 separated from the bacterial producer and is used as a sole chemical in food or other  
1068 industries. However, the disposal of GMM in these cases presents a challenge, and optimal  
1069 destruction and prevention of environmental dissemination of engineered strains have to be  
1070 implemented in industrial strategies (Gautier, 2008).



1071 **6. Concluding remarks**

1072 The *Lactobacillus* genus represents a versatile group of LAB that continues to intrigue  
1073 scientists from different fields of microbiology. Their genetic characteristics are constantly  
1074 being supplemented with new data. The rising number of available genomes provides greater  
1075 opportunities for implementation of the data to give a better understanding of and improved  
1076 application of these microorganisms. Construction of pangenomes reveals genetic and  
1077 phenotypic diversity, and explains adaptability of lactobacilli to various habitats. Genetic data  
1078 can be also used to anticipate the potential of strains for application in various industrial  
1079 fields.

1080 The construction of genome scale computational models gives an indication of a strains  
1081 metabolic potential and facilitates identification of genes most suitable for engineering  
1082 studies (Bution et al., 2015). The introduction of next generation sequencing (NGS) methods  
1083 and metabolite profiling reveals new and unexpected features of LAB. The construction of  
1084 metabolic models of industrial microorganisms is becoming an essential step in the  
1085 development of fermented foods and food ingredients (Smid and Hugenholtz, 2010). The  
1086 overall knowledge obtained after deployment of all approaches described in this review  
1087 contributes to a better understanding of the physiology of *Lactobacillus* cultures during dairy  
1088 production, which encourages the development of novel production technologies that will  
1089 provide continuous product quality improvement (Steele et al., 2013).

1090 **Acknowledgments**

1091 The authors wish to thank to Dr Mary Rea and Dr Brian Healy for their critical reading of this  
1092 manuscript. Ewelina Stefanovic is in a receipt of Teagasc Walsh Fellowship.

1093 References:

- 1094 Altermann, E., Russell, W.M., Azcarate-Peril, M.A., Barrangou, R., Buck, B.L., McAuliffe,  
1095 O., Souther, N., Dobson, A., Duong, T., Callanan, M., Lick, S., Hamrick, A., Cano, R.,  
1096 Klaenhammer, T.R., 2005. Complete genome sequence of the probiotic lactic acid bacterium  
1097 *Lactobacillus acidophilus* NCFM. Proc. Natl. Acad. Sci. U S A 102, 3906-3912.
- 1098 Au, R., 2015. From genetic engineering to genome engineering: what impact has it made on  
1099 science and society. Adv. Biol. Biotechnol. Genet. 2, 1-8.
- 1100 Babu, R.P., O'Connor, K., Seeram, R., 2013. Current progress on bio-based polymers and  
1101 their future trends. Progr Biomater 2, 1-16.
- 1102 Barrangou, R., 2013. CRISPR-Cas systems and RNA-guided interference. Wiley Interdiscip.  
1103 Rev. RNA 4, 267-278.
- 1104 Barrangou, R., Horvath, P., 2012. CRISPR: new horizons in phage resistance and strain  
1105 identification. Annu. Rev. Food Sci. Technol. 3, 143-162.
- 1106 Beauprez, J.J., De Mey, M., Soetaert, W.K., 2010. Microbial succinic acid production:  
1107 Natural versus metabolic engineered producers. Process. Biochem. 45, 1103-1114.
- 1108 Bellanger, X., Payot, S., Leblond-Bourget, N., Guedon, G., 2014. Conjugative and  
1109 mobilizable genomic islands in bacteria: evolution and diversity. FEMS Microbiol. Rev. 38,  
1110 720-760.
- 1111 Berger, B., Pridmore, R.D., Barretto, C., Delmas-Julien, F., Schreiber, K., Arigoni, F.,  
1112 Brussow, H., 2007. Similarity and differences in the *Lactobacillus acidophilus* group  
1113 identified by polyphasic analysis and comparative genomics. J. Bacteriol. 189, 1311-1321.
- 1114 Bermudez-Humaran, L.G., Kharrat, P., Chatel, J.M., Langella, P., 2011. Lactococci and  
1115 lactobacilli as mucosal delivery vectors for therapeutic proteins and DNA vaccines. Microb.  
1116 Cell Fact. 10 Suppl 1, S4.

1117 Bhowmik, T., Steele, J.L., 1994. Cloning, characterization and insertional inactivation of the  
1118 *Lactobacillus helveticus* D(-) lactate dehydrogenase gene. Appl. Microbiol. Biotechnol. 41,  
1119 432-439.

1120 Bogicevic, B., Fuchsmann, P., Breme, K., Portmann, R., Guggenbuhl, B., Irmeler, S., 2013. A  
1121 preliminary study on the effect of *Lactobacillus casei* expressing *cystathionine*  
1122 *lyase1/cystathionine lyase2* on Cheddar cheese and the formation of sulphur-containing  
1123 compounds. Int. Dairy J. 33, 97-103.

1124 Bogicevic, B., Irmeler, S., Portmann, R., Meile, L., Berthoud, H., 2012. Characterization of  
1125 the *cysK2-ctl1-cysE2* gene cluster involved in sulfur metabolism in *Lactobacillus casei*. Int. J.  
1126 Food Microbiol. 152, 211-219.

1127 Broadbent, J.R., Cai, H., Larsen, R.L., Hughes, J.E., Welker, D.L., De Carvalho, V.G.,  
1128 Tompkins, T.A., Ardo, Y., Vogensen, F., De Lorentiis, A., Gatti, M., Neviani, E., Steele, J.L.,  
1129 2011. Genetic diversity in proteolytic enzymes and amino acid metabolism among  
1130 *Lactobacillus helveticus* strains. J. Dairy. Sci. 94, 4313-4328.

1131 Broadbent, J.R., Hughes, J.E., Welker, D.L., Tompkins, T.A., Steele, J.L., 2013. Complete  
1132 Genome Sequence for *Lactobacillus helveticus* CNRZ 32, an Industrial Cheese Starter and  
1133 Cheese Flavor Adjunct. Genome Announc. 1.

1134 Broadbent, J.R., Neeno-Eckwall, E.C., Stahl, B., Tandee, K., Cai, H., Morovic, W., Horvath,  
1135 P., Heidenreich, J., Perna, N.T., Barrangou, R., Steele, J.L., 2012. Analysis of the  
1136 *Lactobacillus casei* supragenome and its influence in species evolution and lifestyle  
1137 adaptation. BMC Genomics 13, 533.

1138 Bull, M.J., Jolley, K.A., Bray, J.E., Aerts, M., Vandamme, P., Maiden, M.C., Marchesi, J.R.,  
1139 Mahenthiralingam, E., 2014. The domestication of the probiotic bacterium *Lactobacillus*  
1140 *acidophilus*. Sci. Rep. 4, 7202.

1141 Buriti, F.C.A., da Rocha, J.S., Saad, S.M.I., 2005. Incorporation of *Lactobacillus acidophilus*  
1142 in Minas fresh cheese and its implications for textural and sensorial properties during storage.  
1143 Int. Dairy J. 15, 1279-1288.

1144 Bution, M.L., Molina, G., Abrahao, M.R., Pastore, G.M., 2015. Genetic and metabolic  
1145 engineering of microorganisms for the development of new flavor compounds from terpenic  
1146 substrates. Crit. Rev. Biotechnol. 35, 313-325.

1147 Cai, H., Rodriguez, B.T., Zhang, W., Broadbent, J.R., Steele, J.L., 2007. Genotypic and  
1148 phenotypic characterization of *Lactobacillus casei* strains isolated from different ecological  
1149 niches suggests frequent recombination and niche specificity. Microbiology 153, 2655-2665.

1150 Cai, H., Thompson, R., Budinich, M.F., Broadbent, J.R., Steele, J.L., 2009. Genome  
1151 sequence and comparative genome analysis of *Lactobacillus casei*: insights into their niche-  
1152 associated evolution. Genome Biol. Evol. 1, 239-257.

1153 Callanan, M., Kaleta, P., O'Callaghan, J., O'Sullivan, O., Jordan, K., McAuliffe, O.,  
1154 Sangrador-Vegas, A., Slattery, L., Fitzgerald, G.F., Beresford, T., Ross, R.P., 2008. Genome  
1155 sequence of *Lactobacillus helveticus*, an organism distinguished by selective gene loss and  
1156 insertion sequence element expansion. J. Bacteriol. 190, 727-735.

1157 Ceapa, C., Lambert, J., van Limpt, K., Wels, M., Smokvina, T., Knol, J., Kleerebezem, M.,  
1158 2015. Correlation of *Lactobacillus rhamnosus* Genotypes and Carbohydrate Utilization  
1159 Signatures Determined by Phenotype Profiling. Appl. Environ. Microbiol. 81, 5458-5470.

1160 Claesson, M.J., van Sinderen, D., O'Toole, P.W., 2007. The genus *Lactobacillus* - a genomic  
1161 basis for understanding its diversity. FEMS Microbiol. Lett. 269, 22-28.

1162 Claesson, M.J., van Sinderen, D., O'Toole, P.W., 2008. *Lactobacillus* phylogenomics-  
1163 towards a reclassification of the genus. Int. J. Syst. Evol. Microbiol. 58, 2945-2954.

1164 Cleveland, J., Montville, T.J., Nes, I.F., Chikindas, M.L., 2001. Bacteriocins: safe, natural  
1165 antimicrobials for food preservation. Int. J. Food Microbiol. 71, 1-20.

1166 Collins, R.E., Higgs, P.G., 2012. Testing the infinitely many genes model for the evolution of  
1167 the bacterial core genome and pangenome. *Mol. Biol. Evol.* 29, 3413-3425.

1168 Collins, Y.F., McSweeney, P.L.H., Wilkinson, M.G., 2003. Lipolysis and free fatty acid  
1169 catabolism in cheese: a review of current knowledge. *Int. Dairy J.* 13, 841-866.

1170 Corsetti, A., Settanni, L., 2007. Lactobacilli in sourdough fermentation. *Food Res. Int.* 40,  
1171 539-558.

1172 Cotter, P.D., Hill, C., Ross, R.P., 2005. Bacteriocins: developing innate immunity for food.  
1173 *Nat. Rev. Microbiol.* 3, 777-788.

1174 Cremonesi, P., Chessa, S., Castiglioni, B., 2012. Genome sequence and analysis of  
1175 *Lactobacillus helveticus*. *Front. Microbiol.* 3, 435.

1176 De Boeck, R., Sarmiento-Rubiano, L.A., Nadal, I., Monedero, V., Perez-Martinez, G., Yebra,  
1177 M.J., 2010. Sorbitol production from lactose by engineered *Lactobacillus casei* deficient in  
1178 sorbitol transport system and mannitol-1-phosphate dehydrogenase. *Appl. Microbiol.*  
1179 *Biotechnol.* 85, 1915-1922.

1180 De Vuyst, L., Leroy, F., 2007. Bacteriocins from lactic acid bacteria: production, purification,  
1181 and food applications. *J. Mol. Microbiol. Biotechnol.* 13, 194-199.

1182 Dellaglio, F., Felis, G.E., Torriani, S., 2002. The status of the species *Lactobacillus casei*  
1183 (Orla-Jensen 1916) Hansen and Lessel 1971 and *Lactobacillus paracasei* Collins et al. 1989.  
1184 Request for an opinion. *Int. J. Syst. Evol. Microbiol.* 52, 285-287.

1185 Deveau, H., Garneau, J.E., Moineau, S., 2010. CRISPR/Cas system and its role in phage-  
1186 bacteria interactions. *Annu. Rev. Microbiol.* 64, 475-493.

1187 Douglas, G.L., Klaenhammer, T.R., 2010. Genomic evolution of domesticated  
1188 microorganisms. *Annu. Rev. Food Sci. Technol.* 1, 397-414.

1189 Douillard, F.P., Ribbera, A., Kant, R., Pietila, T.E., Jarvinen, H.M., Messing, M., Randazzo,  
1190 C.L., Paulin, L., Laine, P., Ritari, J., Caggia, C., Lahtinen, T., Brouns, S.J., Satokari, R., von

1191 Ossowski, I., Reunanen, J., Palva, A., de Vos, W.M., 2013. Comparative genomic and  
1192 functional analysis of 100 *Lactobacillus rhamnosus* strains and their comparison with strain  
1193 GG. PLoS Genet 9, e1003683.

1194 European Food Safety Authority, 2011. Guidance on the risk assessment of genetically  
1195 modified microorganisms and their products intended for food and feed use. EFSA Journal 9,  
1196 2193.

1197 Ejtahed, H.S., Mohtadi-Nia, J., Homayouni-Rad, A., Niafar, M., Asghari-Jafarabadi, M.,  
1198 Mofid, V., Akbarian-Moghari, A., 2011. Effect of probiotic yogurt containing *Lactobacillus*  
1199 *acidophilus* and *Bifidobacterium lactis* on lipid profile in individuals with type 2 diabetes  
1200 mellitus. J. Dairy Sci. 94, 3288-3294.

1201 El Kafsi, H., Binesse, J., Loux, V., Buratti, J., Boudebouze, S., Dervyn, R., Kennedy, S.,  
1202 Galleron, N., Quinquis, B., Batto, J.M., Moumen, B., Maguin, E., van de Guchte, M., 2014.  
1203 *Lactobacillus delbrueckii* ssp. *lactis* and ssp. *bulgaricus*: a chronicle of evolution in action.  
1204 BMC Genomics 15, 407.

1205 Ferain, T., Garmyn, D., Bernard, N., Hols, P., Delcour, J., 1994. *Lactobacillus plantarum*  
1206 *ldhL* gene: overexpression and deletion. J. Bacteriol. 176, 596-601.

1207 Garneau, J.E., Moineau, S., 2011. Bacteriophages of lactic acid bacteria and their impact on  
1208 milk fermentations. Microb. Cell Fact .10 Suppl 1, S20.

1209 Gaspar, P., Carvalho, A.L., Vinga, S., Santos, H., Neves, A.R., 2013. From physiology to  
1210 systems metabolic engineering for the production of biochemicals by lactic acid bacteria.  
1211 Biotechnol. Adv. 31, 764-788.

1212 Gautier, M., 2008. Ethical issues raised by genetically modified microorganisms.  
1213 [http://bioethics.agrocampus-ouest.eu/infoglueDeliverLive/digitalAssets/57484\\_41EN-ethical-](http://bioethics.agrocampus-ouest.eu/infoglueDeliverLive/digitalAssets/57484_41EN-ethical-issues-ogm.pdf)  
1214 [issues-ogm.pdf](http://bioethics.agrocampus-ouest.eu/infoglueDeliverLive/digitalAssets/57484_41EN-ethical-issues-ogm.pdf), Last accessed on 20<sup>th</sup> July 2016.

1215 Genay, M., Sadat, L., Gagnaire, V., Lortal, S., 2009. *prtH2*, not *prtH*, is the ubiquitous cell  
1216 wall proteinase gene in *Lactobacillus helveticus*. Appl. Environ. Microbiol. 75, 3238-3249.

1217 Giraffa, G., 2014. *Lactobacillus helveticus*: importance in food and health. Front. Microbiol.  
1218 5, 338.

1219 Giraffa, G., De Vecchi, P., Rossetti, L., 1998. Note: identification of *Lactobacillus*  
1220 *delbrueckii* subspecies *bulgaricus* and subspecies *lactis* dairy isolates by amplified rDNA  
1221 restriction analysis. J. Appl. Microbiol. 85, 918-924.

1222 Gobbetti, M., De Angelis, M., Di Cagno, R., Mancini, L., Fox, P.F., 2015. Pros and cons for  
1223 using non-starter lactic acid bacteria (NSLAB) as secondary/adjunct starters for cheese  
1224 ripening. Trends Food. Sci. Tech. 45, 167-178.

1225 Goeddel, D.V., Kleid, D.G., Bolivar, F., Heyneker, H.L., Yansura, D.G., Crea, R., Hirose, T.,  
1226 Kraszewski, A., Itakura, K., Riggs, A.D., 1979. Expression in *Escherichia coli* of chemically  
1227 synthesized genes for human insulin. Proc. Natl. Acad. Sci. U S A 76, 106-110.

1228 Gold, R.S., Meagher, M.M., Tong, S.X., Hutkins, R.W., Conway, T., 1996. Cloning and  
1229 expression of the *Zymomonas mobilis* "production of ethanol" genes in *Lactobacillus casei*.  
1230 Curr. Microbiol. 33, 256-260.

1231 Hao, P., Zheng, H., Yu, Y., Ding, G., Gu, W., Chen, S., Yu, Z., Ren, S., Oda, M., Konno, T.,  
1232 Wang, S., Li, X., Ji, Z.S., Zhao, G., 2011. Complete sequencing and pan-genomic analysis of  
1233 *Lactobacillus delbrueckii* subsp. *bulgaricus* reveal its genetic basis for industrial yogurt  
1234 production. PLoS One 6, e15964.

1235 Herias, M.V., Koninkx, J.F., Vos, J.G., Huis in't Veld, J.H., van Dijk, J.E., 2005. Probiotic  
1236 effects of *Lactobacillus casei* on DSS-induced ulcerative colitis in mice. Int. J. Food.  
1237 Microbiol. 103, 143-155.

1238 Horvath, P., Coute-Monvoisin, A.C., Romero, D.A., Boyaval, P., Fremaux, C., Barrangou,  
1239 R., 2009. Comparative analysis of CRISPR loci in lactic acid bacteria genomes. *Int. J. Food*  
1240 *Microbiol.* 131, 62-70.

1241 Iartchouk, O., Kozyavkin, S., Karamychev, V., Slesarev, A., 2015. Complete Genome  
1242 Sequence of *Lactobacillus acidophilus* FSI4, Isolated from Yogurt. *Genome Announc* 3.

1243 Innocente, N., Biasutti, M., Rita, F., Bricchese, R., Comi, G., Iacumin, L., 2016. Effect of  
1244 indigenous *Lactobacillus rhamnosus* isolated from bovine milk on microbiological  
1245 characteristics and aromatic profile of traditional yogurt. *LWT-Food Sci. Technol.* 66, 158-  
1246 164.

1247 Irmeler, S., Raboud, S., Beisert, B., Rauhut, D., Berthoud, H., 2008. Cloning and  
1248 characterization of two *Lactobacillus casei* genes encoding a cystathionine lyase. *Appl.*  
1249 *Environ. Microbiol.* 74, 99-106.

1250 Irmeler, S., Schafer, H., Beisert, B., Rauhut, D., Berthoud, H., 2009. Identification and  
1251 characterization of a strain-dependent cystathionine beta/gamma-lyase in *Lactobacillus casei*  
1252 potentially involved in cysteine biosynthesis. *FEMS Microbiol. Lett* 295, 67-76.

1253 Jiang, W., Marraffini, L.A., 2015. CRISPR-Cas: New Tools for Genetic Manipulations from  
1254 Bacterial Immunity Systems. *Annu. Rev. Microbiol.* 69, 209-228.

1255 Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., Charpentier, E., 2012. A  
1256 programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*  
1257 337, 816-821.

1258 Kamerbeek, J., Schouls, L., Kolk, A., van Agterveld, M., van Soolingen, D., Kuijper, S.,  
1259 Bunschoten, A., Molhuizen, H., Shaw, R., Goyal, M., van Embden, J., 1997. Simultaneous  
1260 detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and  
1261 epidemiology. *J. Clin. Microbiol.* 35, 907-914.



1262 Kanchiswamy, C.N., Malnoy, M., Velasco, R., Kim, J.S., Viola, R., 2015. Non-GMO  
1263 genetically edited crop plants. *Trends Biotechnol.* 33, 489-491.

1264 Kant, R., Blom, J., Palva, A., Siezen, R.J., de Vos, W.M., 2011. Comparative genomics of  
1265 *Lactobacillus*. *Microb. Biotechnol.* 4, 323-332.

1266 Kant, R., Rintahaka, J., Yu, X., Sigvart-Mattila, P., Paulin, L., Mecklin, J.P., Saarela, M.,  
1267 Palva, A., von Ossowski, I., 2014. A comparative pan-genome perspective of niche-adaptable  
1268 cell-surface protein phenotypes in *Lactobacillus rhamnosus*. *PLoS One* 9, e102762.

1269 Keasling, J.D., 2012. Synthetic biology and the development of tools for metabolic  
1270 engineering. *Metab. Eng.* 14, 189-195.

1271 King, Z.A., Lloyd, C.J., Feist, A.M., Palsson, B.O., 2015. Next-generation genome-scale  
1272 models for metabolic engineering. *Curr. Opin. Biotechnol.* 35, 23-29.

1273 Kleerebezem, M., Boekhorst, J., van Kranenburg, R., Molenaar, D., Kuipers, O.P., Leer, R.,  
1274 Turchini, R., Peters, S.A., Sandbrink, H.M., Fiers, M.W., Stiekema, W., Lankhorst, R.M.,  
1275 Bron, P.A., Hoffer, S.M., Groot, M.N., Kerkhoven, R., de Vries, M., Ursing, B., de Vos,  
1276 W.M., Siezen, R.J., 2003. Complete genome sequence of *Lactobacillus plantarum* WCFS1.  
1277 *Proc .Natl. Acad. Sci. U S A* 100, 1990-1995.

1278 Kyla-Nikkila, K., Hujanen, M., Leisola, M., Palva, A., 2000. Metabolic engineering of  
1279 *Lactobacillus helveticus* CNRZ32 for production of pure L-(+)-lactic acid. *Appl. Environ.*  
1280 *Microbiol.* 66, 3835-3841.

1281 Ladero, V., Ramos, A., Wiersma, A., Goffin, P., Schanck, A., Kleerebezem, M., Hugenholtz,  
1282 J., Smid, E.J., Hols, P., 2007. High-level production of the low-calorie sugar sorbitol by  
1283 *Lactobacillus plantarum* through metabolic engineering. *Appl. Environ. Microbiol.* 73, 1864-  
1284 1872.

1285 Lebeer, S., Vanderleyden, J., De Keersmaecker, S.C., 2008. Genes and molecules of  
1286 lactobacilli supporting probiotic action. *Microbiol Mol Biol Rev* 72, 728-764.

1287 Lewis, N.E., Nagarajan, H., Palsson, B.O., 2012. Constraining the metabolic genotype-  
1288 phenotype relationship using a phylogeny of *in silico* methods. Nat. Rev. Microbiol. 10, 291-  
1289 305.

1290 Li, N., Huang, Y., Liu, Z., You, C., Guo, B., 2015a. Regulation of EPS production in  
1291 *Lactobacillus casei* LC2W through metabolic engineering. Lett. Appl. Microbiol. 61, 555-  
1292 561.

1293 Li, N., Wang, Y., Zhu, P., Liu, Z., Guo, B., Ren, J., 2015b. Improvement of  
1294 exopolysaccharide production in *Lactobacillus casei* LC2W by overexpression of NADH  
1295 oxidase gene. Microbiol. Res. 171, 73-77.

1296 Liu, E., Hao, P., Konno, T., Yu, Y., Oda, M., Zheng, H., Ji, Z.-S., 2012. Amino Acid  
1297 Biosynthesis and Proteolysis in *Lactobacillus Bulgaricus* Revisited: A Genomic Comparison.  
1298 Comput. Mol. Biosci. 02, 61-77.

1299 Liu, M., Bayjanov, J.R., Renckens, B., Nauta, A., Siezen, R.J., 2010. The proteolytic system  
1300 of lactic acid bacteria revisited: a genomic comparison. BMC Genomics 11, 36.

1301 Liu, M., Bienfait, B., Sacher, O., Gasteiger, J., Siezen, R.J., Nauta, A., Geurts, J.M., 2014.  
1302 Combining chemoinformatics with bioinformatics: *in silico* prediction of bacterial flavor-  
1303 forming pathways by a chemical systems biology approach "reverse pathway engineering".  
1304 PLoS One 9, e84769.

1305 Liu, M., Nauta, A., Francke, C., Siezen, R.J., 2008. Comparative genomics of enzymes in  
1306 flavor-forming pathways from amino acids in lactic acid bacteria. Appl. Environ. Microbiol.  
1307 74, 4590-4600.

1308 Liu, S., Nichols, N.N., Dien, B.S., Cotta, M.A., 2006. Metabolic engineering of a  
1309 *Lactobacillus plantarum* double *ldh* knockout strain for enhanced ethanol production. J. Ind.  
1310 Microbiol. Biotechnol. 33, 1-7.

1311 Ljungh, A., Wadstrom, T., 2006. Lactic acid bacteria as probiotics. *Curr. Issues Intest.*  
1312 *Microbiol.* 7, 73-90.

1313 Mahadevan, R., Schilling, C.H., 2003. The effects of alternate optimal solutions in constraint-  
1314 based genome-scale metabolic models. *Metab. Eng.* 5, 264-276.

1315 Makarova, K., Slesarev, A., Wolf, Y., Sorokin, A., Mirkin, B., Koonin, E., Pavlov, A.,  
1316 Pavlova, N., Karamychev, V., Polouchine, N., Shakhova, V., Grigoriev, I., Lou, Y., Rohksar,  
1317 D., Lucas, S., Huang, K., Goodstein, D.M., Hawkins, T., Plengvidhya, V., Welker, D.,  
1318 Hughes, J., Goh, Y., Benson, A., Baldwin, K., Lee, J.H., Diaz-Muniz, I., Dosti, B., Smeianov,  
1319 V., Wechter, W., Barabote, R., Lorca, G., Altermann, E., Barrangou, R., Ganesan, B., Xie,  
1320 Y., Rawsthorne, H., Tamir, D., Parker, C., Breidt, F., Broadbent, J., Hutkins, R., O'Sullivan,  
1321 D., Steele, J., Unlu, G., Saier, M., Klaenhammer, T., Richardson, P., Kozyavkin, S., Weimer,  
1322 B., Mills, D., 2006. Comparative genomics of the lactic acid bacteria. *Proc. Natl. Acad. Sci.*  
1323 *U S A* 103, 15611-15616.

1324 Makarova, K.S., Koonin, E.V., 2007. Evolutionary genomics of lactic acid bacteria. *J.*  
1325 *Bacteriol.* 189, 1199-1208.

1326 Marco, M.B., Moineau, S., Quiberoni, A., 2012. Bacteriophages and dairy fermentations.  
1327 *Bacteriophage* 2, 149-158.

1328 Marilley, L., Casey, M.G., 2004. Flavours of cheese products: metabolic pathways, analytical  
1329 tools and identification of producing strains. *Int. J. Food Microbiol.* 90, 139-159.

1330 Mayo, B., Aleksandrak-Piekarczyk, T., Fernández, M., Kowalczyk, M., Álvarez-Martín, P.,  
1331 Bardowski, J., 2010. Updates in the Metabolism of Lactic Acid Bacteria, in: *Mozzi, F., Raya,*  
1332 *R., Vignolo, G. (Eds.), Biotechnology of Lactic Acid Bacteria. Wiley-Blackwell, pp. 3-33.*

1333 Mayo, B., van Sinderen, D., Ventura, M., 2008. Genome analysis of food grade lactic Acid-  
1334 producing bacteria: from basics to applications. *Curr. Genomics* 9, 169-183.

1335 Mazzoli, R., Bosco, F., Mizrahi, I., Bayer, E.A., Pessione, E., 2014. Towards lactic acid  
1336 bacteria-based biorefineries. *Biotechnol. Adv.* 32, 1216-1236.

1337 McAuliffe, O. 2017. Genetics of Lactic Acid Bacteria, in: Fox, P.F., McSweeney, P.L.H.,  
1338 Cotter, P., Everett, D.W. (Eds), *Cheese: Chemistry, Physics And Microbiology*, Fourth  
1339 Edition, Elsevier, The Netherlands - Currently in press

1340 Medina de Figueroa, R., Oliver, G., Benito de Cádiz, I.L., 2001. Influence of temperature  
1341 on flavour compound production from citrate by *Lactobacillus rhamnosus* ATCC 7469.  
1342 *Microbiol. Res.* 155, 257-262.

1343 Medini, D., Donati, C., Tettelin, H., Masignani, V., Rappuoli, R., 2005. The microbial pan-  
1344 genome. *Curr. Opin. Genet. Dev.* 15, 589-594.

1345 Meneghel, J., Dugat-Bony, E., Irlinger, F., Loux, V., Vidal, M., Passot, S., Beal, C., Layec,  
1346 S., Fonseca, F., 2016. Draft Genome Sequence of *Lactobacillus delbrueckii* subsp. *bulgaricus*  
1347 CFL1, a Lactic Acid Bacterium Isolated from French Handcrafted Fermented Milk. *Genome*  
1348 *Announc.* 4.

1349 Milesi, M.M., Wolf, I.V., Bergamini, C.V., Hynes, E.R., 2010. Two strains of nonstarter  
1350 lactobacilli increased the production of flavor compounds in soft cheeses. *J. Dairy Sci.* 93,  
1351 5020-5031.

1352 Molenaar, D., Bringel, F., Schuren, F.H., de Vos, W.M., Siezen, R.J., Kleerebezem, M.,  
1353 2005. Exploring *Lactobacillus plantarum* genome diversity by using microarrays. *J.*  
1354 *Bacteriol.* 187, 6119-6127.

1355 Mortera, P., Pudlik, A., Magni, C., Alarcon, S., Lolkema, J.S., 2013. Ca<sup>2+</sup>-citrate uptake and  
1356 metabolism in *Lactobacillus casei* ATCC 334. *Appl. Environ. Microbiol.* 79, 4603-4612.

1357 Nadal, I., Rico, J., Perez-Martinez, G., Yebra, M.J., Monedero, V., 2009. Diacetyl and  
1358 acetoin production from whey permeate using engineered *Lactobacillus casei*. *J. Ind.*  
1359 *Microbiol. Biotechnol.* 36, 1233-1237.

1360 Nichols, N.N., Dien, B.S., Bothast, R.J., 2003. Engineering lactic acid bacteria with pyruvate  
1361 decarboxylase and alcohol dehydrogenase genes for ethanol production from *Zymomonas*  
1362 *mobilis*. J. Ind. Microbiol. Biotechnol. 30, 315-321.

1363 Nissen, L., Perez-Martinez, G., Yebra, M.J., 2005. Sorbitol synthesis by an engineered  
1364 *Lactobacillus casei* strain expressing a sorbitol-6-phosphate dehydrogenase gene within the  
1365 lactose operon. FEMS Microbiol. Lett. 249, 177-183.

1366 O'Brien, E.J., Monk, J.M., Palsson, B.O., 2015. Using Genome-scale Models to Predict  
1367 Biological Capabilities. Cell 161, 971-987.

1368 O'Sullivan, O., O'Callaghan, J., Sangrador-Vegas, A., McAuliffe, O., Slattery, L., Kaleta, P.,  
1369 Callanan, M., Fitzgerald, G.F., Ross, R.P., Beresford, T., 2009. Comparative genomics of  
1370 lactic acid bacteria reveals a niche-specific gene set. BMC Microbiol. 9, 50.

1371 Oliveira, A.P., Nielsen, J., Forster, J., 2005. Modeling *Lactococcus lactis* using a genome-  
1372 scale flux model. BMC Microbiol. 5, 39.

1373 Ong, L., Henriksson, A., Shah, N.P., 2007. Chemical analysis and sensory evaluation of  
1374 Cheddar cheese produced with *Lactobacillus acidophilus*, *Lb. casei*, *Lb. paracasei* or  
1375 *Bifidobacterium* sp. Int. Dairy J. 17, 937-945.

1376 Orth, J.D., Thiele, I., Palsson, B.O., 2010. What is flux balance analysis? Nat. Biotechnol. 28,  
1377 245-248.

1378 Papagianni, M., 2012. Metabolic engineering of lactic acid bacteria for the production of  
1379 industrially important compounds. Comput. Struct. Biotechnol. J 3, e201210003.

1380 Pastink, M.I., Teusink, B., Hols, P., Visser, S., de Vos, W.M., Hugenholtz, J., 2009. Genome-  
1381 scale model of *Streptococcus thermophilus* LMG18311 for metabolic comparison of lactic  
1382 acid bacteria. Appl. Environ. Microbiol. 75, 3627-3633.

1383 Pedersen, M.B., Iversen, S.L., Sorensen, K.I., Johansen, E., 2005. The long and winding road  
1384 from the research laboratory to industrial applications of lactic acid bacteria. FEMS  
1385 Microbiol. Rev. 29, 611-624.

1386 Pfeiler, E.A., Klaenhammer, T.R., 2007. The genomics of lactic acid bacteria. Trends  
1387 Microbiol. 15, 546-553.

1388 Pridmore, R.D., Berger, B., Desiere, F., Vilanova, D., Barretto, C., Pittet, A.C., Zwahlen,  
1389 M.C., Rouvet, M., Altermann, E., Barrangou, R., Mollet, B., Mercenier, A., Klaenhammer,  
1390 T., Arigoni, F., Schell, M.A., 2004. The genome sequence of the probiotic intestinal  
1391 bacterium *Lactobacillus johnsonii* NCC 533. Proc. Natl. Acad. Sci. U S A 101, 2512-2517.

1392 Ramachandran, P., Lacher, D.W., Pfeiler, E.A., Elkins, C.A., 2013. Development of a tiered  
1393 multilocus sequence typing scheme for members of the *Lactobacillus acidophilus* complex.  
1394 Appl. Environ. Microbiol. 79, 7220-7228.

1395 Rossi, F., Rizzotti, L., Felis, G.E., Torriani, S., 2014. Horizontal gene transfer among  
1396 microorganisms in food: current knowledge and future perspectives. Food Microbiol. 42,  
1397 232-243.

1398 Sander, J.D., Joung, J.K., 2014. CRISPR-Cas systems for editing, regulating and targeting  
1399 genomes. Nat. Biotechnol. 32, 347-355.

1400 Savijoki, K., Ingmer, H., Varmanen, P., 2006. Proteolytic systems of lactic acid bacteria.  
1401 Appl. Microbiol. Biotechnol. 71, 394-406.

1402 Selle, K., Barrangou, R., 2015. Harnessing CRISPR-Cas systems for bacterial genome  
1403 editing. Trends. Microbiol. 23, 225-232.

1404 Senan, S., Prajapati, J.B., Joshi, C.G., 2014. Comparative genome-scale analysis of niche-  
1405 based stress-responsive genes in *Lactobacillus helveticus* strains. Genome 57, 185-192.

1406 Senan, S., Prajapati, J.B., Joshi, C.G., 2015. Whole-genome based validation of the adaptive  
1407 properties of Indian origin probiotic *Lactobacillus helveticus* MTCC 5463. J. Sci. Food  
1408 Agric. 95, 321-328.

1409 Settanni, L., Moschetti, G., 2010. Non-starter lactic acid bacteria used to improve cheese  
1410 quality and provide health benefits. Food Microbiol. 27, 691-697.

1411 Sgarbi, E., Lazzi, C., Tabanelli, G., Gatti, M., Neviani, E., Gardini, F., 2013. Nonstarter lactic  
1412 acid bacteria volatiles produced using cheese components. J. Dairy Sci. 96, 4223-4234.

1413 Siezen, R.J., Tzeneva, V.A., Castioni, A., Wels, M., Phan, H.T., Rademaker, J.L.,  
1414 Starrenburg, M.J., Kleerebezem, M., Molenaar, D., van Hylckama Vlieg, J.E., 2010.  
1415 Phenotypic and genomic diversity of *Lactobacillus plantarum* strains isolated from various  
1416 environmental niches. Environ. Microbiol. 12, 758-773.

1417 Siezen, R.J., van Hylckama Vlieg, J.E., 2011. Genomic diversity and versatility of  
1418 *Lactobacillus plantarum*, a natural metabolic engineer. Microb. Cell Fact. 10 Suppl 1, S3.

1419 Smid, E.J., Hugenholtz, J., 2010. Functional genomics for food fermentation processes.  
1420 Annu. Rev. Food. Sci. Technol. 1, 497-519.

1421 Smit, G., Smit, B.A., Engels, W.J., 2005. Flavour formation by lactic acid bacteria and  
1422 biochemical flavour profiling of cheese products. FEMS Microbiol. Rev. 29, 591-610.

1423 Smokvina, T., Wels, M., Polka, J., Chervaux, C., Brisse, S., Boekhorst, J., van Hylckama  
1424 Vlieg, J.E., Siezen, R.J., 2013. *Lactobacillus paracasei* comparative genomics: towards  
1425 species pan-genome definition and exploitation of diversity. PLoS One 8, e68731.

1426 Stahl, B., Barrangou, R., 2013. Complete Genome Sequence of Probiotic Strain *Lactobacillus*  
1427 *acidophilus* La-14. Genome Announc. 1, e00376-00313.

1428 Steele, J., Broadbent, J., Kok, J., 2013. Perspectives on the contribution of lactic acid bacteria  
1429 to cheese flavor development. Curr. Opin. Biotechnol. 24, 135-141.

1430 Strahinic, I., Lozo, J., Terzic-Vidojevic, A., Fira, D., Kojic, M., Golic, N., Begovic, J.,  
1431 Topisirovic, L., 2013. Technological and probiotic potential of BGRA43 a natural isolate of  
1432 *Lactobacillus helveticus*. Front. Microbiol. 4, 2.

1433 Sun, Z., Harris, H.M., McCann, A., Guo, C., Argimon, S., Zhang, W., Yang, X., Jeffery, I.B.,  
1434 Cooney, J.C., Kagawa, T.F., Liu, W., Song, Y., Salvetti, E., Wrobel, A., Rasinkangas, P.,  
1435 Parkhill, J., Rea, M.C., O'Sullivan, O., Ritari, J., Douillard, F.P., Paul Ross, R., Yang, R.,  
1436 Briner, A.E., Felis, G.E., de Vos, W.M., Barrangou, R., Klaenhammer, T.R., Caufield, P.W.,  
1437 Cui, Y., Zhang, H., O'Toole, P.W., 2015a. Expanding the biotechnology potential of  
1438 lactobacilli through comparative genomics of 213 strains and associated genera. Nat.  
1439 Commun. 6, 8322.

1440 Sun, Z., Liu, W., Song, Y., Xu, H., Yu, J., Bilige, M., Zhang, H., Chen, Y., 2015b.  
1441 Population structure of *Lactobacillus helveticus* isolates from naturally fermented dairy  
1442 products based on multilocus sequence typing. J. Dairy Sci. 98, 2962-2972.

1443 Sybesma, W., Hugenholtz, J., de Vos, W.M., Smid, E.J., 2006. Safe use of genetically  
1444 modified lactic acid bacteria in food. Bridging the gap between consumers, green groups, and  
1445 industry. Electron. J. Biotechn. 9, 424-448.

1446 Tanous, C., Kieronczyk, A., Helinck, S., Chambellon, E., Yvon, M., 2002. Glutamate  
1447 dehydrogenase activity: a major criterion for the selection of flavour-producing lactic acid  
1448 bacteria strains. Antonie Van Leeuwenhoek 82, 271-278.

1449 Taverniti, V., Guglielmetti, S., 2012. Health-Promoting Properties of *Lactobacillus*  
1450 *helveticus*. Front. Microbiol. 3, 392.

1451 Teusink, B., Bachmann, H., Molenaar, D., 2011. Systems biology of lactic acid bacteria: a  
1452 critical review. Microb. Cell Fact. 10 Suppl 1, S11.



1453 Teusink, B., Wiersma, A., Molenaar, D., Francke, C., de Vos, W.M., Siezen, R.J., Smid, E.J.,  
1454 2006. Analysis of growth of *Lactobacillus plantarum* WCFS1 on a complex medium using a  
1455 genome-scale metabolic model. J. Biol. Chem. 281, 40041-40048.

1456 Toh, H., Oshima, K., Nakano, A., Takahata, M., Murakami, M., Takaki, T., Nishiyama, H.,  
1457 Igimi, S., Hattori, M., Morita, H., 2013. Genomic adaptation of the *Lactobacillus casei* group.  
1458 PLoS One 8, e75073.

1459 Tsuji, A., Okada, S., Hols, P., Satoh, E., 2013. Metabolic engineering of *Lactobacillus*  
1460 *plantarum* for succinic acid production through activation of the reductive branch of the  
1461 tricarboxylic acid cycle. Enzyme Microb. Technol. 53, 97-103.

1462 Tuo, Y., Zhang, W., Zhang, L., Ai, L., Zhang, Y., Han, X., Yi, H., 2013. Study of probiotic  
1463 potential of four wild *Lactobacillus rhamnosus* strains. Anaerobe 21, 22-27.

1464 Urshev, Z., Ishlimova, D., 2015. Distribution of clustered regularly interspaced palindrome  
1465 repeats CRISPR2 and CRISPR3 in *Lactobacillus delbrueckii* ssp. *bulgaricus* strains.  
1466 Biotechnol. Biotec. Eq. 29, 541-546.

1467 Van de Guchte, M., Penaud, S., Grimaldi, C., Barbe, V., Bryson, K., Nicolas, P., Robert, C.,  
1468 Oztas, S., Mangenot, S., Couloux, A., Loux, V., Dervyn, R., Bossy, R., Bolotin, A., Batto,  
1469 J.M., Walunas, T., Gibrat, J.F., Bessieres, P., Weissenbach, J., Ehrlich, S.D., Maguin, E.,  
1470 2006. The complete genome sequence of *Lactobacillus bulgaricus* reveals extensive and  
1471 ongoing reductive evolution. Proc. Natl. Acad. Sci. U S A 103, 9274-9279.

1472 Van Hoorde, K., Van Leuven, I., Dirinck, P., Heyndrickx, M., Coudijzer, K., Vandamme, P.,  
1473 Huys, G., 2010. Selection, application and monitoring of *Lactobacillus paracasei* strains as  
1474 adjunct cultures in the production of Gouda-type cheeses. Int. J. Food Microbiol. 144, 226-  
1475 235.

1476 Van Kranenburg, R., Kleerebezem, M., van Hylckama Vlieg, J., Ursing, B.M., Boekhorst, J.,  
1477 Smit, B.A., Ayad, E.H.E., Smit, G., Siezen, R.J., 2002. Flavour formation from amino acids  
1478 by lactic acid bacteria: predictions from genome sequence analysis. *Int. Dairy J.* 12, 111-121.  
1479 Villena, J., Oliveira, M.L., Ferreira, P.C., Salva, S., Alvarez, S., 2011. Lactic acid bacteria in  
1480 the prevention of pneumococcal respiratory infection: future opportunities and challenges.  
1481 *Int. Immunopharmacol.* 11, 1633-1645.  
1482 Vinay-Lara, E., Hamilton, J.J., Stahl, B., Broadbent, J.R., Reed, J.L., Steele, J.L., 2014.  
1483 Genome-scale reconstruction of metabolic networks of *Lactobacillus casei* ATCC 334 and  
1484 12A. *PLoS One* 9, e110785.  
1485 Wright, A.V., Nunez, J.K., Doudna, J.A., 2016. Biology and Applications of CRISPR  
1486 Systems: Harnessing Nature's Toolbox for Genome Engineering. *Cell* 164, 29-44.  
1487 Wyszynska, A., Kobierecka, P., Bardowski, J., Jagusztyn-Krynicka, E.K., 2015. Lactic acid  
1488 bacteria-20 years exploring their potential as live vectors for mucosal vaccination. *Appl.*  
1489 *Microbiol. Biotechnol.* 99, 2967-2977.  
1490 Xu, N., Liu, J., Ai, L., Liu, L., 2015. Reconstruction and analysis of the genome-scale  
1491 metabolic model of *Lactobacillus casei* LC2W. *Gene* 554, 140-147.  
1492 Ya, T., Zhang, Q., Chu, F., Merritt, J., Bilige, M., Sun, T., Du, R., Zhang, H., 2008.  
1493 Immunological evaluation of *Lactobacillus casei* Zhang: a newly isolated strain from  
1494 koumiss in Inner Mongolia, China. *BMC Immunol.* 9, 68.  
1495 Yebra, M.J., Pérez-Martínez, G., Rodríguez-Díaz, J., Monedero, V., 2012. Genetically  
1496 Engineered Lactobacilli for Technological and Functional Food Applications, in Valdez, B.  
1497 (Ed), *Food Industrial Processes –Methods and Equipment*. InTech, Shanghai, China, pp. 143-  
1498 168.

1499 Yebra, M.J., Zuniga, M., Beaufils, S., Perez-Martinez, G., Deutscher, J., Monedero, V., 2007.  
1500 Identification of a gene cluster enabling *Lactobacillus casei* BL23 to utilize myo-inositol.  
1501 Appl. Environ. Microbiol. 73, 3850-3858.

1502 Yu, S., Peng, Y., Zheng, Y., Chen, W., 2015. Comparative Genome Analysis of  
1503 *Lactobacillus casei*: Insights into Genomic Diversification for Niche Expansion. Indian J.  
1504 Microbiol. 55, 102-107.

1505 Zhu, Y., Zhang, Y., Li, Y., 2009. Understanding the industrial application potential of lactic  
1506 acid bacteria through genomics. Appl. Microbiol. Biotechnol. 83, 597-610.

1507 Table 1: General genomic features of the most important dairy related *Lactobacillus* species.

1508 All data were obtained at <http://www.ncbi.nlm.nih.gov/>, last assessed in July 2016.

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

1509

1510 Figures captions:

1511 Figure 1: Process of niche adaptation. (a) Ancestor of *Lactobacillus* spp. had undergone  
1512 multiple genome changes, such as decay of superfluous genes and acquisition of genes that  
1513 support survival in specific environmental conditions, which all led to niche specialisation for  
1514 various habitats, three of which have been depicted here (dairy, environment, human and  
1515 animal GIT). However, strains of *Lactobacillus* could change their habitat (b), for instance  
1516 during human consumption of dairy or plant food, and this is why isolation source does not  
1517 always correspond to the strains' natural environment. This has to be kept in mind while  
1518 analysing characteristics of strains isolated from different ecological niches, as origin of  
1519 isolation gives only an indication of metabolic capacity of an organism.

1520 Figure 2: (a) CRISPR-Cas system of bacteria enables efficient resistance to phage attack. For  
1521 example, in case of dairy lactobacilli, when the cells encounter the dairy phage for the first  
1522 time, its DNA is cleaved and a sequence that includes repeater (black box) and spacer (blue  
1523 box) is integrated in CRISPR-cas locus, directly behind the leader sequence. (b) In the event  
1524 of repeated attack by the same phage, its DNA sequence corresponding to an existing spacer  
1525 induces transcription and maturation of CRISPR RNA (crRNA), which activates Cas  
1526 complex and efficiently cleaves the foreign DNA. Further stages of phage reproduction are  
1527 terminated, and there are no newly assembled phage particles. As the dairy strain combats the  
1528 phage, normal fermentation process occurs. (c) CRISPR systems mechanism initiated  
1529 development of genome editing tool. Here, Cas 9 nuclease interacts with chimeric guide  
1530 RNA, that provides the enzyme to the specific site in DNA, after which precise double  
1531 stranded break (DBS) occurs. After DBS, breaks can be either nonhomologously joined  
1532 leading to an indel mutation, or, in presence of a donor DNA, this sequence is precisely  
1533 inserted in a homology directed repair event.

1534 Figure 3: Schematic view of range of applications of available genome sequences. The whole  
1535 genome sequencing (WGS) data provides the basis for genomic characterisation of species or  
1536 genera, as well as evolutionary studies, such as niche adaptability. Insight in genetic content  
1537 of a strain can predict the presence of metabolic machinery that could generate flavour  
1538 compounds. Additionally, they enable the construction of genome scale metabolic models,  
1539 which coupled to genetic information and biochemical data lead to the development of  
1540 metabolic engineering studies. Results of these studies reveal strains capacity for plausible  
1541 industrial applications.