Chapter 3 Genomic Characterization of Starter and Probiotic Cultures

Giovanna E. Felis, Sandra Torriani, Ana Belén Flórez, & Baltasar Mayo

Contributors:

Giovanna E. Felis
Dept. of Biotechnology
University of Verona
Strada le Grazie, 15
I-37134 Verona
phone: +39 045 8027930
email: giovanna.felis@univr.it

Sandra Torriani
Dept. of Biotechnology
University of Verona
Strada le Grazie, 15
I-37134 Verona
phone: +39 045 8027930
email: sandra.torriani@univr.it

Baltasar Mayo
Dpto. of Microbiology and Biochemistry
Instituto de Productos Lácteos de Asturias (IPLA-CSIC)
Paseo Río Linares, s/n
33300-Villaviciosa
Spain
phone: +34985892131
email: baltasar.mayo@ipla.csic.es

Ana Belén Flórez
Dpto. of Microbiology and Biochemistry
Instituto de Productos Lácteos de Asturias (IPLA-CSIC)
Paseo Río Linares, s/n
33300-Villaviciosa
Spain
phone: +34985892131
email: abflorez@ipla.csic.es

3.1 Introduction

The first publicly available genome sequence of a lactic acid bacteria (LAB) strain commonly used as a starter in dairy fermentations is that of *Lactococcus lactis* subsp. *lactis* IL1403 (Bolotin *et al.*, 2001). Since then, a huge number of genome sequences of starter and probiotic LAB has been determined. Indeed, genome sequencing has become the starting point of most microbial studies, since it can provide, at a reasonable cost, the framework of the biology of any strain of interest. Comparative genome analyses can provide insights on the genetic make-up, and thus on the metabolic potential, of the strains, as well as on many aspects of their genome evolution and divergence. Further, sequencing of genomes can reveal horizontal gene transfer (HGT) events between species and strains of starter and probiotic cultures.

Knowledge of the complete genome sequence is not only useful in discovering novel genes and new properties, but also in determining the absence at the genetic level of undesirable traits. In addition, it is of much help in characterizing the variants of modified/improved strains for both cultures with a technological role, i.e., necessary/useful for food fermentation processes, and a functional role, i.e., able to confer health benefits to the consumer/host (FAO/WHO, 2001; Hill *et al.*, 2014). Genome sequencing can further allow precise identification down to the strain level, while providing strategies for its specific detection and quantification, important aspects for commercially-relevant bacteria.

As an example, comparative sequence analysis has led, at the beginning of the century, to the recognition of Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR), and their CRISPR-associated (*cas*) genes (Jansen *et al.*, 2002). In slightly more than a decade, these rudimentary immune systems of bacteria have also been identified in several industrial LAB strains (Horvath *et al.*, 2009). CRISPR-*cas* systems seem to be related with the prevention of lateral HGT processes, which are involved in for instance phage resistance (Barrangou and Horvath 2012); thus, they are considered as a desirable trait for strain components of starter cultures. Moreover, CRISPR-*cas* loci could be used as a strain-level identification tool (Barrangou and Horvath 2012) and have become a method for genome editing in a variety of genetic engineering applications (Barrangou, 2014).

In general, comparative genomics, i.e. the comparison of genome sequence data for many strains, which can be performed with dedicated software platforms and using specific databases, is a very important strategy to reveal the peculiarities of specific strains/groups of strains. Therefore, genome sequencing efforts continuing to date on one side might provide redundant information, but, on the other side, they are necessary to continuously improve our knowledge of the microbial biodiversity.

Since the first edition of this book chapter (Dellaglio et al., 2005), several important reports have appeared providing new data on genome analyses of strains of interest for the food industry. First, in 2006, the comparative genomics of a bunch of LAB strains belonging to several genera and species was published (Makarova et al., 2006), and, more in detail for probiotics, the term *probiogenomics* was introduced by Ventura et al. (2009) to indicate that genome sequence analyses could provide insights into the genetic background related to probiosis (see below), which has been further investigated by Lukjancenko et al. (2011). Genome-wide analyses can indeed provide scientists with clues on the genetic background for strains to sense and adapt to their ecological niche, where they exert specific functions (Siezen and Wilson, 2010), or to reveal the genetic background related to physiological properties of specific interest in industrial applications, as in the case of the proto-cooperation between *Streptococcus thermophilus* and *Lactobacillus delbrueckii* subsp. *bulgaricus* for yogurt production (Hao et al., 2011). This approach could be applied to all starter cultures (Garrigues et al., 2013), making it possible a genotype-phenotype matching for specific strains, opening the way for targeted strain improvements, as well as the basis for in depth safety assessments. In fact, the term *pangenomics*, namely the analysis of the pan-genome of a species, devised from the genome sequences of different strains of the same species (Medini et al., 2005), implies the importance of comparative analyses. With this respect, the GOLD database, i.e. the Genomes OnLine Database (<http://www.genomesonline.org>), at its fifth version (Reddy et al., 2015), is like a gold mine, since it is a comprehensive online resource, regularly updated, cataloguing and monitoring genome studies ongoing worldwide. Information on genomes (and metagenomes) present in the database is classified in a four-level system: (i) *Studies*, which group one or more related (ii) *Biosamples*, individual samples of genetic material, from which the organism DNA was isolated for downstream sequencing, that are analogous to the NCBI's BioProject; (iii) *Sequencing Projects*, the sequencing deliverables from the Biosamples, and (iv) *Analysis projects*, methods of data processing which are applied to sequencing projects (Reddy et al., 2015), all in compliance with the indications of the Genomic Standards Consortium about Minimum Information Standards (Field et al., 2011; Yilmaz et al., 2011).

Performing a GOLD database search, when “dairy industry” is selected among the filters for the relevance of the study, the website (October 2016) showed that 2,578 sequencing projects were ongoing worldwide for 2,537 organisms, mainly related to the domain Bacteria.

Within bacteria, the most important species for dairy applications are of course LAB strains belonging to the genera *Lactococcus* and *Lactobacillus*, together with the species *S. thermophilus*, which they all belong to the phylum *Firmicutes*. Several LAB species also include probiotic strains, together with many strains of *Bifidobacterium*, belonging this latter genus to the phylum *Actinobacteria*. This has led to the accumulation of sequence data, which numbers are reported in **Table 1**. It has to be pointed out that suppliers of commercial and probiotic cultures all around the world may have additional data on larger numbers of industrially-proprietary strains of these genera and species.

In the last two years, two important papers have appeared reporting the efforts to determine the genome sequences of the type strains of *Bifidobacterium* (Milani et al., 2014) and *Lactobacillus* species (Sun et al., 2015), as well as LAB strains of related genera, including those relevant for dairy applications. This constitutes the framework for genome-based analyses at various taxonomic levels and provides a reference scheme to investigate physiology, biochemistry, evolution and diversity of dairy-related starter and probiotics.

Since information on identification will be reported elsewhere in this book (chapter 6), here we report on the insights obtained, from genome sequence data, on diversity and evolution at taxonomic ranks above the species level, as well as on safety aspects and on specific technologically and functionally important features.

Table 1. Summary of genome sequences available to date for dairy starter and probiotic dairy strains retrieved using their species names as organisms' names. Studies, Organisms, Sequencing projects and Analysis projects correspond to the four-level classification system used in GOLD database (Reddy et al., 2015), which is briefly explained in the text. Names of genera are abbreviated as follows: *B.*, *Bifidobacterium*; *Lb.*; *Lactobacillus*; *Lc.*, *Lactococcus*; *S.*, *Streptococcus*.

Species	Studies	Organisms	Sequencing projects	Analysis projects
<i>B. animalis</i> (including <i>B. animalis</i> subsp. <i>lactis</i>)	26	192	35	28
<i>B. bifidum</i>	18	198	32	25
<i>B. breve</i>	22	185	42	35
<i>B. longum</i> (including <i>B. longum</i> subsp. <i>infantis</i>)	40	439	107	75
<i>Lb. acidophilus</i>	15	388	26	19
<i>Lb. casei</i>	26	389	44	35
<i>Lb. delbrueckii</i> (including <i>Lb. delbrueckii</i> subsp. <i>bulgaricus</i>)	20	837	30	29
<i>Lb. fermentum</i>	17	400	20	17
<i>Lb. gasseri</i>	12	361	26	17
<i>Lb. helveticus</i>	19	297	30	22
<i>Lb. johnsonii</i>	10	61	12	11
<i>Lb. paracasei</i>	15	581	52	49
<i>Lb. plantarum</i>	70	1106	118	100
<i>Lb. reuteri</i>	15	184	20	17
<i>Lb. salivarius</i>	18	119	20	15
<i>Lc. lactis</i>	53	1632	95	89
<i>S. thermophilus</i>	34	330	34	31

3.2 Genome sequencing and comparative genomics: insights into evolution and adaptation to dairy environments

In this section, the most recent information gained from genome sequence analyses on the evolution and adaptation to dairy environments of the main microbial genera and species will be briefly reviewed. Due to the large amount of information available, an effort of synthesis and selection of information has been made. Information is structured following a taxonomic framework and focusing mainly on *Firmicutes* first, which include *Lactococcus lactis*, the most frequently employed starter cultures, and *S. thermophilus*, as well as lactobacilli, which could be used as either starter cultures, adjunct and probiotic cultures (Kelleher et al., 2015). Moreover, some information of *Actinobacteria*, in particular bifidobacteria and dairy propionibacteria, will be provided.

It can be underlined that adaptation to milk, which is a nutritionally rich medium containing carbohydrates, proteins, fats, vitamins and minerals at a neutral pH (Marshall, 1991), seems to have generally determined trends of loss or inactivation (mostly by mutation) of genes encoding non-essential properties for growth in milk. The loss of some phenotypic traits has further been complemented by the acquisition through horizontal transfer of key genes coding for critical functions in the milk environment, e.g., lactose and casein utilization, phage resistance, exopolysaccharide production, and so on.

3.2.1. Phylum *Firmicutes*

3.2.1.1. Family *Streptococcaceae*: *Lactococcus* and *Streptococcus*

The taxonomic family of *Streptococcaceae* includes the three genera *Lactococcus*, *Lactovum* and *Streptococcus*, which share a high degree of sequence similarity (considering 16S rRNA gene sequences), and can be found in a variety of environmental niches, including the dairy-related environments (Cavanagh et al., 2015).

Lc. lactis and *S. thermophilus* constitute the most economically-important dairy species. Consequently, they have been investigated in several studies from a genomic viewpoint; also, the impact of genome sequencing for the selection of dairy starters in these two taxa has been recently reviewed (Kelleher et al., 2015). A common element in industrial dairy *Streptococcaceae* genomes is the process of gene loss, along with a relatively high percentage of pseudogenes caused by

nonsense mutations, deletions, truncations and/or frameshifts, which are thought to contribute to cell economy, and thus to milk adaptation (Cavanagh et al., 2015).

At present, eleven species are recognisedrecognized in the genus *Lactococcus* (<http://www.bacterio.net/lactococcus.html>). Of these, the most important is *Lc. lactis*. Taxonomically speaking, *Lc. lactis* is divided into four subspecies, i.e., subsp. *cremoris*, *hordniae*, *lactis* and *tructae* (Pérez et al., 2011), with *lactis* and *cremoris* being the only subspecies of industrial relevance. The evolution and adaptation of *Lc. lactis* to milk have been recently reviewed by Cavanagh et al. (2015), and will be briefly updated here. Besides *Lc. lactis*, there is a long list of newly described species, including *Lc. chungangensis*, *Lc. formosensis*, *Lc. fujiensis*, *Lc. garvieae*, *Lc. hircilactis*, *Lc. laudensis*, *Lc. piscium*, *Lc. plantarum*, *Lc. raffinolactis*, and *Lc. taiwanensis*.

At the time of writing this review (Oct. 2016), 85 genomes for different strains of *Lc. lactis* are deposited in the GenBank database (<https://www.ncbi.nlm.nih.gov/genome/?term=Lactococcus+lactis>). Genomes of lactococci are usually relatively small, ranging from ~2.2 to 2.5 Mb (Kelleher et al., 2015), with a protein-encoding gene content between ~2000 and ~2800 (Makarova et al., 2006). At the genus level, the analysis of 5 *L. lactis* genomes, 1 *Lc. raffinolactis* genome and 19 *Lc. garvieae* genomes has shown that 70% of the genes of the latter species was shared with other lactococci, constituting, most probably, the core genome of the genus (Ferrario et al., 2013). In most strains, chromosomes are often complemented by a large array of 4-7 plasmids (Kelleher et al., 2015). In fact, dairy lactococci of industrial use are characterized by the presence of large plasmids carrying genes for all pivotal properties of technological significance mentioned above (lactose utilization and casein breakdown, bacteriophage resistance, etc.) (Ainsworth et al. 2014; Siezen et al., 2005).

Dairy starters of *Lc. lactis* have been defined as ‘domesticated’ strains (Passerini et al., 2010). Most probably, these domesticated strains derive from plant-associated microorganisms (Kelly et al., 2010), which became adapted to the dairy environment. In fact, adaptation has led to a reductive evolution of their genomes, mainly with respect to the biosynthesis of amino acids and the ability to ferment plant derived carbohydrates (Kelly et al., 2010; Cavanagh et al., 2015). Comparison of a dairy strain and a sourdough strain, IL1403 vs A12, respectively, has revealed that about a quarter of the A12 genes were strain-specific and mainly responsible for niche specialisation (Passerini et al., 2013), indicating the importance of the ‘dispensable genome’; i.e., of genes present only in a

particular strain or group of strains. Those genes are most commonly associated with phages, transposons/mobile elements, and plasmids and could be horizontally transferred, generating diversity within the species. Interestingly, lactose utilization and citrate metabolism, the latter is a characteristic of *Lc. lactis* subsp. *lactis* biovar *diacetylactis*, are plasmid-encoded traits in *Lc. lactis* (Kelleher et al., 2015; Van Hylckama Vlieg et al., 2006). This reductive evolution seems to be common not only to the other industrially important dairy bacterium of the same taxonomic family, i.e., *S. thermophilus* (Hols et al., 2005), but also to *L. delbrueckii* subsp. *bulgaricus* (Van De Guchte et al., 2006).

As already mentioned, lactose utilization in starter strains of *Lc. lactis*, is plasmid encoded and could be instable (Ainsworth et al. 2014). At least two alternatives for lactose metabolism have been described in this species, showing that phenotypic growth on lactose is not always exclusive of strains bearing a plasmid encoded *lac* operon. These alternatives can easily be revealed by whole genome sequencing, demonstrating that this technique is more appropriate to completely characterize the strains compared to gene-specific detection and analysis (Kelleher et al., 2015).

As for *S. thermophilus*, to date (Oct. 2016) 30 genome sequences are available in GenBank, (<https://www.ncbi.nlm.nih.gov/genome/?term=Streptococcus+thermophilus>), and the species belongs to a genus which includes 85 species including several pathogens (<http://www.bacterio.net/streptococcus.html>). More in detail, *S. thermophilus* belongs to the *salivarius* group of streptococci, which includes also *Streptococcus salivarius* and *Streptococcus vestibularis*, two commensals that may occasionally cause opportunistic infections in humans (Delorme et al., 2015). Despite the genetic relatedness of these three species, they can be readily distinguished by Multilocus Sequence Typing (MLST) and comparative genome analysis. These techniques confirm the status of separate species for *S. thermophilus*, which was temporarily classified as a subspecies of *S. salivarius* (Schleifer et al., 1991). In fact, according to recent genome analyses, *S. thermophilus* seems to have evolved relatively recently, from a common ancestor of *S. vestibularis*, in parallel with the development of the agriculture (Delorme et al., 2015).

S. thermophilus also possesses a relatively small genome (about 1.7-1.8 Mb), with respect to strains of the other two species, and its genome is characterized by a relatively high number of pseudogenes, which is in line with adaptation to milk (Delorme et al., 2015; Bolotin et al., 2004). Moreover, as observed for *Lc. lactis*, additional genes coding for thermotolerance are frequently carried on plasmids (Kelleher et al., 2015). Regarding lactose metabolism, so important for milk

adaptation, it differs from that of *Lc. Lactis*, as lactose utilization in *S. thermophilus* is more similar to that of *Lactobacillus delbrueckii* (Kelleher et al., 2015).

Interestingly, two other species of the genus *Streptococcus*, belonging to a distinct phylogenetic group, the *Streptococcus bovis*/*Streptococcus equinus* complex, i.e., *Streptococcus gallolyticus* subsp. *macedonicus* (Schlegel et al., 2003; Whiley & Kilian, 2003) and *Streptococcus infantarius* subsp. *infantarius* (Jans et al., 2013), appear to be dominant in specific fermented (dairy) products (Papadimitriou et al., 2014). In particular, *S. gallolyticus* subsp. *macedonicus* seems to have all the characteristics indicative of reductive evolution, i.e., reduced genome size, increased number of pseudogenes, as well as loss of genes/pathways related to plant carbohydrates catabolism. In addition, its genome makeup is integrated, among others, by extra genes for lactose and galactose metabolism, casein hydrolysis, supporting once again its adaptation to milk (Papadimitriou et al., 2014). However, data reported by Papadimitriou et al. (2014) suggest that neither *S. gallolyticus* subsp. *macedonicus* nor *S. infantarius* subsp. *infantarius* are specialized dairy bacteria like *S. thermophilus*, and most probably represent intermediate evolutionary stages which could be compared to the ancestors of *S. thermophilus*, and not to the highly dairy-specialized species that we know today.

3.2.1.2. Families *Lactobacillaceae* and *Leuconostocaceae*

Members of those two taxonomic families include strains that may be used as starter and adjunct cultures for dairy application (Kelleher et al., 2015).

Considering the genus *Lactobacillus*, at present it includes about 180 validly described species (<http://www.bacterio.net/lactobacillus.html>), while 14 are described in the genus *Leuconostoc* (<http://www.bacterio.net/leuconostoc.html>). Recently, a genome sequencing initiative has been completed focusing on 185 type strains of species of 11 genera related to genus *Lactobacillus*, including *Oenococcus* and *Leuconostoc* genomes (Sun et al., 2015). The report showed that, when analyzing different parameters such as average nucleotide identity (ANI) and total nucleotide identity (TNI), the genus *Lactobacillus* appears to be more diverse than a well-defined ‘standard’ taxonomic family (Sun et al., 2015). Moreover, the genus is paraphyletic, i.e., its evolutionary origin is not unique, and the genera *Pediococcus*, *Weissella*, *Leuconostoc*, *Oenococcus* and *Fructobacillus*, appear to be sub-clades within the ‘real’ lactobacilli. For those reasons, therefore, the name Lactobacillus Genus Complex (LGC), which would include all these genera, has recently

been proposed (Sun et al., 2015). This implies that the separation of these bacteria into two families, namely *Lactobacillaceae* and *Leuconostocaceae*, should be revised. In contrast, the family *Streptococcaceae* appears clearly separated from the LGC.

Genome sequences available to date are more than 900 (NCBI Genome search with 'Lactobacillus[orgn]' as a query, Oct. 2016). Of these, *Lb. rhamnosus* and *Lb. plantarum* strains account for 225 sequences (104 and 121, respectively). Due to the remarkable diversity of the genus, sequences are very variable in size and in many different parameters. Sun et al. (2015) reported that the smallest and the largest genomes for lactobacilli were 1.23Mb (*Lb. sanfranciscensis*) 4.91 Mb (*Lb. parakefiri*), respectively. Also, genome GC content appeared to be very variable, between 31.93% to 57.02%, and the core genome very small, estimated in 73 genes, including 213 strains. Remarkably, the pan-genome appears to be very large (44,668 gene families) and continuously increasing when novel genomes are added (Sun et al., 2015).

Among dairy-relevant species, the smallest genome appeared to be that of *Lb. delbrueckii* (about 1.8 Mbp), and the largest that of *Lb. plantarum* (about 3.2 Mbp) (Stefanovic et al., 2017). Comparative genomics analyses have revealed that a mobile insertion element (IS), IS91, seems to be dairy specific (found in dairy *Lb. casei* and *Lb. paracasei* subsp. *tolerans*) (Sun et al., 2015).

Lb. delbrueckii and *Lb. helveticus* are the most important dairy starter cultures among lactobacilli and belong to the same phylogenetic group, the *Lb. delbrueckii* group, which can also be considered the genus *Lactobacillus sensu stricto*, whilst other useful adjunct cultures, like *Lb. casei*, *Lb. paracasei*, *Lb. plantarum*, and *Lb. rhamnosus* are phylogenetically and metabolically distinct from *Lb. delbrueckii* clade. Notably, *Lb. casei*, *Lb. paracasei*, and *Lb. rhamnosus* form a group of closely related facultatively heterofermentative species, while *Lb. plantarum* belongs to another phylogenetic clade and is most similar to *Lb. fabifermentans*, *Lb. paraplantarum*, *Lb. pentosus*, and *Lb. xiangfangensis* (Sun et al., 2015).. Probiotic strains most commonly associated to dairy products belong to either *Lb. acidophilus*, a species within the *Lb. delbrueckii* clade, or to *Lb. rhamnosus* (Stefanovic et al., 2017). Although oOther relevant strains belonging to *Lb. fermentum*, *Lb. gasseri*, *Lb. johnsonii*, *Lb. plantarum*, *Lb. reuteri*, and *Lb. salivarius* species are also used with the same purposes (Table 1).

Dairy specialists have shown an abundance of genes encoding components of proteolytic systems, as well as sugar and amino acid transporters. These might result from duplications, which make it possible for organisms to uptake nutrients from milk, while substantial gene loss has been

described for coding sequences not necessary in dairy ecosystems (Stefanovic et al., 2017). Information on dairy-related genes is not always consistent among studies (e.g., O'Sullivan et al., 2009 and Kant et al., 2011), but the studies by Broadbent et al. (2012) and Smokvina et al. (2013) on strains of *Lb. casei* group (*Lb. casei* and *Lb. paracasei*, respectively) have shown that dairy-specialized strains undergo reduction of the genome, which appears to be associated to a restricted capacity of carbohydrate degradation. On the contrary, strains able to survive in dairy products but not dairy specialists, could present genes useful for being flexible and able to proliferate in different niches, such as those of the gut or the environment (Senan et al., 2015).

In summary, dairy adaptation has had deep consequences at the genomic level and, although a general trend towards genome reduction with respect to several functions, strains belonging to different species could show a different set of genes, either vertically or horizontally transmitted, that account for their performance in milk fermentation.

3.2.2. Phylum *Actinobacteria*

3.2.2.1 Genus *Bifidobacterium* and related genera

The genus *Bifidobacterium* includes bacterial strains highly relevant for their health-promoting properties. At present, the genus comprises 50 validly described species (<http://www.bacterio.net/bifidobacterium.html>), and the most important in the probiotic perspective are *B. animalis* (including *B. animalis* subsp. *lactis*), *B. bifidum*, *B. breve*, and *B. longum* (including *B. longum* subsp. *infantis*).

In the last two years, the sequencing initiative of the type strains for the to-date-described species was completed (Milani et al., 2014) and a comparative genome analyses on 62 genome sequences of Bifidobacteriales, the taxonomic order in which the genus *Bifidobacterium* is included, has been published (Zhang et al., 2016). The extensive detailed comparative analysis performed by Milani and colleagues (2014) showed an extensive trend of horizontal gene gain during evolution, involving mainly determinants for complex carbohydrate transport and degradation, which are most probably related to their adaptation to the gut niche. The analysis at the order level (Zhang et al., 2016) also revealed that, inside the family Bifidobacteriaceae, the genera *Bifidobacterium* and *Gardnerella* are not clearly separated; they appear more related to each other than to other genera (i.e., *Scardovia*, *Parascardovia* and *Alloscardovia*). Moreover, confirming previous findings,

several phylogenetic groups can be devised within the genus *Bifidobacterium*, and species of human probiotic interest were spread among three groups, i.e. *B. longum* group (*B. breve* and *B. longum*), *B. bifidum* group (*B. bifidum*), and *B. pseudolongum* group (*B. animalis*). Remarkably, authors highlighted the presence of 3 *Bifidobacterium*-specific conserved proteins (conserved signature protein, CSP), and 2 CSP specific for genus *Bifidobacterium* and *Gardnerella*, which all encode for unknown functions. This highlights the usefulness of genome sequencing and comparative analyses to reveal gene targets which are probably crucial for the biology of the microorganism but, at present, completely unknown. Besides those evolutionary considerations and insights into the biology of organism yet-to-be-discovered, complete genome sequences have also been the basis for the definition of an accurate PCR-based identification protocol (Ferrario et al., 2015).

3.2.2.2 Genus *Propionibacterium*

Propionibacteria are Gram positive, non-spore forming, non-motile rods with high GC content between 53 to 68 %, which includes at present 14 species (<http://www.bacterio.net/propionibacterium.html>).

Within the genus, several species, namely *P. freudenreichii* (subsp. *freudenreichii* and subsp. *shermanii*), *P. acidipropionici*, *P. jensenii*, and *P. thoenii* represent the dairy propionibacteria (Poonam et al., 2012); these can be used as starter and/or adjunct cultures. The first genome sequence of a *Propionibacterium* strain has only been determined in 2010 (Falentin et al. 2010). It consisted of 2.7 Mb with 67 % GC content. Six years later, 187 genome sequences of propionibacteria have been deposited in GenBank (<https://www.ncbi.nlm.nih.gov/genome/?term=Propionibacterium+%5Borgn%5D>). Of these, a large majority (120) focus on *P. acnes* and therefore are not relevant for the topic of this chapter. The relatively low number of sequences from dairy propionibacteria suggests that their study is far behind that of other dairy-associated genera. However, since propionibacteria could produce antimicrobial compounds (bacteriocins and organic acids) as well as vitamins of the B group, conjugated linoleic acid, trehalose, propionic acid, and bifidogenic factors (Poonam et al., 2012), it can be expected that study of this group of bacteria could go more in depth in the next years.

3.2.2.3 Other microorganisms

Some other starter and especially probiotic microorganisms could further be considered, including strains belonging to other LAB genera and species, as well as strains from other bacterial species and yeasts. High numbers of *Enterococcus* spp. and *Weissella* spp. have been repeatedly reported in many fermented foods, particularly in cheese (Franz et al., 2011; Abriouel et al., 2015). However, species of the genus *Enterococcus* have been proposed as starters (Giraffa, 2003) and some others (such as Symbioflor 1) have been used as probiotics since the 1950 (Domann et al., 2007). Their dual-faced status might be based on a few set of genes whose absence has to be demonstrated in species of this genus to be used in food and feed (EFSA, 2012). Strains of other species, such as *Escherichia coli* (Nissle 1917), *Bacillus subtilis* (Enterogermin), *Bacillus cereus* (Toyoi) and others have been widely used as human and animal probiotics. Further microorganisms could also include yeasts, such as *Geotrichum candidum*, the lactic mold for excellence, which has many different metabolic pathways that are of particular interest to the dairy industry (Boutrou and Guéguen, 2005), and *Saccharomyces cerevisiae* subsp. *bouardii*, the only yeast approved as a probiotic for human consumption (Hatoum et al., 2012). Representatives of all these species and strains have already been genome sequenced. However, in order of clarity and brevity in the precedent paragraphs we focused our attention into the typical LAB genera and species.

3.3 Application of genome analysis in LAB and bifidobacteria

The availability of genome sequences has significantly increased our capabilities to unravel the fermentation pathways and biochemical routes of LAB and bifidobacteria species involved in industrial and probiotic applications (Klaenhammer et al., 2005; Stanton et al., 2005). Comparative genomics is now being used to find core genes, niche-specific genes, and genes linked to specific probiotic traits, but can as well be used to find or exclude the presence of virulence or antibiotic resistance genes or to find indications of chromosomal integration of horizontally acquired DNA, which could indicate the potential of horizontal gene transfer (HGT) of such undesirable traits. The knowledge gathered through genome analysis will ultimately allow full exploitation of the biotechnological potential of LAB and bifidobacteria, facilitating, at the same time, their genetic manipulation. Engineering of LAB and bifidobacteria is pivotal in the discovery of gene functionality of key phenotypic traits, as well as in the use of model strains as a cell factory for the expression of heterologous proteins (Hanniffy et al., 2004), synthesis of food-grade additives and nutraceuticals (Hugenholtz et al., 2002), and the use of these bacteria as vaccine and therapeutic delivery systems (Wells and Mercenier, 2003). Genome sequence availability has recently allowed the replacement of former conventional genetic engineering techniques (Gasson and de Vos, 2004) by state-of-the-art genome editing methods, such as the Single-stranded DNA recombination

mediated genetic engineering (SSDNA recombineering) (van Pijkeren and Britton, 2014) and the Clustered regularly interspaced palindromic repeats (CRISPRs) and the CRISPR-associated (Cas) nuclease techniques (Oh and van Pijkeren, 2014).

3.3.1 *In silico* safety assessment of LAB and bifidobacteria

In silico safety assessment of starter and probiotics strains encompasses several aspects, such as the absence of transmissible antibiotic resistance (AR), the genetic make-up for virulence factors (VF), and other deleterious characteristics required to be met by regulatory agencies, such as those of the European Food Safety Authority (EFSA) (EFSA, 2013).

Determination of the presence/absence of these genetic traits has become more rapid and cost-effective thanks to the growing availability of whole genome sequences and the development of bioinformatics tools (Alkema et al., 2015). Thus, such determination could be considered as an indispensable criterion for (pre)-selection of strains with industrial and probiotic properties. Indeed, a deep and complete safety evaluation is required for each novel strain, but also for those strains that have been in use over a long period of time. As a consequence, whole genome analyses could serve as the standard in the safety evaluation process, and influence regulatory decisions on the commercial acceptability of the strain (Salveti et al., 2016).

In the last few years, a limited number of papers have already been published regarding the safety assessments of LAB and *Bifidobacterium* strains based on the complete genome sequences. A summary of such studies, that have considered mainly strains used or proposed to be used as probiotics, is reported in Table 2.

Table 2. Summary of studies on genome-based assessment of safety for LAB and *Bifidobacterium* strains. Genus names abbreviations are reported in Table 1, with the addition of *E.*, *Enterococcus*; *W.*, *Weissella*; *Le.*, *Leuconostoc*.

Species	Strain	Origin	Genes of concern	Reference
<i>Lc. lactis</i>	CHCC6005	Food/natural source	<i>tet</i> (S)	Bennedsen et al., 2011
<i>B. animalis</i> subsp. <i>lactis</i>	ATCC 27536	Animal (chicken feces)	<i>tet</i> (W)	
	CHCC13471	Food/natural source	<i>tet</i> (W)	
	IPLAIC4	Food (fermented milk)	<i>tet</i> (W)	
<i>Lb. plantarum</i>	JDM1	Chinese commercial probiotic	126 non-specific VF, 51 AR, 23 AM	Zhang et al., 2012
<i>B. longum</i>	JDM301	Chinese commercial probiotic	162 non-specific VF, 36 AR including <i>tet</i> (V), <i>tet</i> (W), <i>tet</i> (PB), <i>tet</i> (Q), 5 AM	Wei et al., 2012
<i>B. longum</i>	CECT 7347	Healthy infants (faeces)	No VF No AR	Chenoll et al., 2013
<i>S. salivarius</i>	NU10 and YU10	Healthy Malaysian subjects	No VF	Barbour and Philip, 2014
<i>E. faecium</i>	NRRL B-2354	Dairy utensils	<i>scm</i> , <i>sagA</i> , <i>efaA</i> , <i>pilA</i>	Kopit et al., 2014
<i>Lb. helveticus</i>	MTCC 5463	Indian healthy woman (vaginal tract)	44 non-specific VF 32 AR 8 AM	Senan et al., 2015
<i>W. cibaria</i>	KACC 11862	Food (kimchi)	2 VF, 4 AR	Abriouel et al., 2015
<i>W. confusa</i>	LBAE C39-2	Food (French wheat sourdough)	4 VF, 2 AR, <i>vanZ</i>	
<i>Le. mesenteroides</i>	LbE15	Food (Italian cheese)	<i>erm</i> (B)	Flórez et al., 2016
	LbE16	Food (Italian cheese)	<i>tet</i> (S), <i>aadE</i> , <i>aphA-3</i> , <i>sat4</i> , <i>vat</i> (E)	

VF: virulence factors; AR: antibiotic resistance; AM: adverse metabolites

Bennedsen et al. (2011) were the first to perform a screening for antimicrobial resistance (AR) genes and virulence factors (VF) via genome sequencing. Twenty-eight strains of LAB and bifidobacteria were tested for the presence of >250 AR genes and >400 toxin and VF genes. It was found that some strains contained AR genes, while no VF genes were detected. In particular, *Lc. lactis* CHCC6005, showing high-level resistance to tetracycline (MIC >256 µg/ml), carried the gene *tet(S)* on a medium-copy-number plasmid. Therefore, the authors indicated that this strain should be plasmid-cured before use. The gene *tet(W)*, encoding resistance to tetracycline, has been detected in all the three *B. animalis* subsp. *lactis* strains analyzed. This determinant is widespread in *B. animalis*. However, since transfer of *tet(W)* from *B. animalis* subsp. *lactis* to other bacteria has never been demonstrated, this gene can be considered non-transmissible (Gueimonde et al., 2010; Bennedsen et al., 2011).

A comprehensive safety assessment of the commercial probiotic strain *Lb. plantarum* JDM1 based on the whole genome sequence was performed by Zhang et al. (2012). In total, 51 nonspecific, AR-associated genes (as classified by the RAST annotating service; <http://rast.nmpdr.org/>), 126 no-offensive virulence-associated genes, and 23 adverse metabolism-associated genes were found in JDM1. However, there were no toxin or hemolysin encoding genes, and safety-associated genes were rarely transferable, thus the Generally Regarded As Safe (GRAS) status for *Lb. plantarum* JDM1 was confirmed. An aspect to be underlined is related to some discrepancies found between genotype and phenotype. For example, JDM1 was sensitive to the antibiotic chloramphenicol, but the genome contained a *cat* gene coding for a chloramphenicol O-acetyltransferase enzyme that is responsible for chloramphenicol resistance in bacteria. These data suggest that when the analysis of whole bacterial genome sequences reveals putative unfavorable genes, the biosafety of strains must be assessed more carefully and comprehensively through phenotypic analyses.

Whole genome sequencing was used by Wei et al. (2012) to assess the safety of *B. longum* JDM301, a commercial strain used widely in China with several probiotic functions. This strain is safe based on phenotype, however, bioinformatics analysis of its genome reveals several potential risk factors, i.e. 36 genes associated with AR, including a tetracycline resistance gene with certain risk of transfer, and 5 putative genes associated with production of harmful metabolites. In addition, *B. longum* JDM301 contains 162 non-specific VF, mainly associated with transcriptional regulation, adhesion, sugar and amino acid transport (Wei et al., 2012). The discrepancies between phenotype and genotype underline the necessity to further experimentally evaluate the potential risk factors found in the genome of *B. longum* JDM301, e.g. throughout *in vivo* assessment using animal models and/or clinical trials.

Chenoll et al. (2013) have demonstrated *in silico*, *in vitro* and *in vivo* the safety of *B. longum* CECT 7347, a probiotic strain able to ameliorate gluten-related damage in celiac disease. Robust arguments can be found in the genome analysis that supports the *in vitro* and *in vivo* results, confirming the safety status of CECT 7347. Indeed, its genome contains neither relevant VF, nor potential AR genes.

In silico safety assessment of the *S. salivarius* strains NU10 and YU10 was performed by Barbour and Philip (2014). The two bacteriocin-producer strains were isolated from Malaysian subjects and have a potential as probiotics intended to be applied in the oral cavity. The *in vitro* tests and genome sequencing established the absence of virulence determinants known to be present in streptococcal pathogens. This finding nominates these strains as potential candidates for probiotic development, as they pass the initial safety assessment described previously for the *S. salivarius* strain K12 (Burton et al., 2006).

The safety of strain *E. faecium* NRRL B-2354 (ATCC 8459) based on its genomic and functional characteristics has been recently investigated (Kopit et al., 2014). This strain has a long history of use as a surrogate of pathogens in food products and thermal process validation, and the findings of this study support its continued use. Indeed, the inspection of the genome sequences revealed that NRRL B-2354 lacks AR genes, as well as enterococcal VF, including *acm*, *cyl*, the *ebp* operon, *esp*, *gelE*, *hyl*, and *IS16*. Accordingly, the strain is sensitive to clinically relevant antibiotics and does not present phenotypes associated to expression of VF. It contains complete copies of *scm*, *sagA*, *efaA*, and the *pilA* operon, but the roles of these genes in enterococcal virulence are not yet well understood. The technical guidance for establishing the safety of *E. faecium* strains intended as additives in animal nutrition of EFSA recommends investigating for the presence of *esp*, *hyl*, and *IS16* as well as sensitivity to ampicillin as exclusion criteria (EFSA, 2012). *E. faecium* NRRL B-2354 does not contain these marker genes typical of hospital-associated isolates responsible for clinical infections, and is susceptible to clinically relevant antibiotics, including ampicillin. Therefore, this strain meets the requirements for safety by the EFSA guidelines.

Data mining of the whole genome sequences of the established probiotic *Lb. helveticus* MTCC 5463, a strain of Indian origin, was carried out by Senan et al. (2015) for reassuring its safety. Genome sequences were screened for genetic determinants associated with AR, production of harmful metabolites, and VF. MTCC 5463 carried AR genes associated with beta-lactam and fluoroquinolone resistance. However, there is no threat of transferability of such resistance as the genome lacks the presence of transmissible elements, such as plasmids, transposons and complete prophages. A hemolysin gene was discovered in the genome, revealing a theoretical risk of virulence. Overall, the results of *in silico* analyses complemented the *in vitro* studies and human

clinical trials and provided safety assurance and QPS status to *Lb. helveticus* MTCC 5463 as a probiotic.

In the frame of a survey on the genus *Weissella*, Abriouel et al. (2015) examined the safety aspects of *W. cibaria* KACC 11862 and *W. confusa* LBAE C39-2 based on *in silico* analyses of their whole genome sequences. Strains belonging to these species are widespread in fermented foods and have been proposed as a starter culture, and also as probiotic (Fusco et al., 2015). In this survey, *W. cibaria* KACC11862 was found to harbor some VF (hemolysins) and AR genes (coding for a multidrug transporter involved in fosfomycin resistance, MDT-FosB; three methicillin resistance proteins; and *vanZ*, encoding a glycopeptide resistance protein). *W. confusa* LBAEC39-2 showed the presence of VF genes (encoding collagen adhesion, hemolysin, and mucus-binding proteins) and AR genes (MDT-FosB, MRP and *vanZ*). Further studies should elucidate the functionality of the AR genes of weissellas, and the transferability of these genes to other bacteria. Therefore, the safety of *Weissella* strains intended for industrial use should be investigated for each strain in detail, carefully selecting strains lacking pathogenic potential and which do not possess transferable AR genes.

Recently, genome analysis of three *Le. mesenteroides* strains isolated from Italian soft cheese samples provided a better understanding of the genetic bases of AR in such species and its transference capability among food-borne bacteria (Flórez et al., 2016). Indeed, in the genome of the multi-resistant strain *Le. mesenteroides* subsp. *mesenteroides* LbE16, genes that might be involved in tetracycline [*tet(S)*], aminoglycoside (*aadE*, *aphA-3*, *sat4*) and virginiamycin [*vat(E)*] resistance were found. *Le. mesenteroides* subsp. *dextranicum* LbE15, erythromycin and clindamycin resistant, harbours an *erm(B)* gene associated to a plasmid of ≈ 35 kbp. However, no known tetracycline resistance genes were detected in *Le. mesenteroides* subsp. *cremoris* LbT16, displaying atypical resistance to this antibiotic (MIC of 32 $\mu\text{g/ml}$) (EFSA, 2012), thus suggesting a new mechanism of resistance which can be due either to acquired genes or to a mutation of indigenous genes. Interestingly, analysis of the AR genes and their flanking regions revealed the potential of some determinants to be horizontally transferred. Indeed, the erythromycin resistance was transferred by conjugation between *Le. mesenteroides* and *E. faecalis* both *in vitro* and in cheese (Flórez et al., 2016), supplying novel proof that LAB can act as a reservoir of acquired AR genes, for which reason their safety should be carefully examined.

Globally, these publications document that whole genome sequences of LAB and bifidobacteria provide a wealth of *in silico* information related to strain safety, which cannot be obtained by any other approach. The genome-wide screening could be an effective and time-saving tool for identifying prognostic biomarkers of biosafety, thus providing prospective for answering the

important safety questions required also for marketing authorization and health claim requests (Miquel et al. 2015; Salvetti et al., 2016). Although many phenotypic tests can be re-placed by whole genome analyses, the overall physiology of a strain should be taken in consideration, as well. Since there is still a lack of homogeneity regarding the genetic and phenotypic traits to be assayed and the proper use of the available bioinformatics tools. Thus, harmonization of scientific procedures is needed to obtain an accurate characterization of each strain and a solid demonstration of its safety. This obvious need is underlined by recent publications, in which frameworks for appropriate evaluation schemes to determine the safety and efficacy of microorganisms intentionally added to foods are being proposed (Pariza et al., 2015; Miquel et al., 2015; Salvetti et al., 2016). The application of a comprehensive workflow, starting from the genome sequence of a probiotic or starter strain, is expected to increase the consistency of future safety assessment, ensuring stakeholders involved in this area (scientists, manufacturers, legislative bodies and consumers) the ability to obtain complete and easily comparable information to meet regulatory requirements.

3.3.2 Unravelling LAB and bifidobacteria properties

Only a few genetic traits appear to be universally conserved among the different LAB genomes, including enzymes involved in glycolysis in ‘true’ LAB or those of the bifidus shunt in bifidobacteria (Makarova et al., 2006; Klijn et al., 2005). Different genetic events (i.e., mutation, gene duplication, HGT, gene decay, gene loss, genome rearrangements, etc.) are considered to contribute to the present genome shape and structure of LAB and bifidobacteria species. As already mentioned, adaptation to nutritionally rich environments (e.g., milk, plant material, human and animal GIT) has promoted progressive gene decay (Callanan et al., 2008; van de Guchte et al., 2006), but also acquisition through horizontal gene transfer (HGT) events of genes involved in key properties for niche colonization or rapid growth in different environments (O’Sullivan et al., 2009; Schell et al., 2002). In addition to the genetic characterization of typical LAB and bifidobacteria features, sequencing and analysis of genomes has uncovered the genetic make-up of previously unnoticed technological- and probiotic-relevant phenotypic properties. As explained below, characterization of these “new” traits has occasionally driven to the development of novel industrial processes, enhancing applicability and/or benefits of the use of LAB and bifidobacteria in food systems.

3.3.2.1 Colonization and niche-competition strategies

In addition to its contribution to the metabolic and bioprocessing potential, availability of LAB and bifidobacteria genome sequences has also expanded our knowledge on the molecular basis of the mechanisms by which species colonize and persist on the ecological niches they occupy. As such, the genome analysis of well-known probiotic strains have given some clues on the mechanistic way by which probiotic organisms colonize the gut and the means by which they positively impact on host health (Klaenhammer et al., 2005). The properties should then have to be experimentally tested under laboratory conditions and in clinical trials. In this way, transcriptome analysis of the *B. breve* UCC2003 genome in a murine colonization model has revealed differential expression of a type IVb tight adherence (*tad*) pilus-encoding gene cluster, which proved to be essential for efficient *in vivo* murine (an probably human) gut colonization (O'Connell Motherway et al., 2011). The Tad pilus-encoding locus was shown to be conserved among different *Bifidobacterium* species, supporting the general involvement of pili in gut colonization and persistence (Turroni et al., 2011).

In a similar approach, pili-like structures were reported for the first time in lactobacilli (Kankainen et al., 2009). Pili are proteinaceous appendages (1-10 nm in diameter protruding 2-3 μ m) localized at the cell surface that have been well characterized in Gram-positive and Gram-negative pathogens (Proft and Baker, 2009). Unlike Gram-negative pili, each Gram-positive pilus is an assembly of multiple pilin subunits (proteins having LPXTG motives) coupled to each other by covalent bonds by the transpeptidase activity of a pilin-specific sortase. Typically, one of the proteins is called the pilin backbone, as 100-200 of these subunits are assembled head-to-tail to form the pilus shaft, while one or two other pilins, called ancillary pilins, may exist decorating the base, the cap or all along the pilus shaft. Analysis of the genome of the human-derived *Lb. rhamnosus* GG strain identified two separate pilus clusters in its genome (*spaCBA-srtC1* and *spaFED*). The former operon is encoded in a GG-specific genomic island, while the latter operon was also shown to be present in the genome of *L. rhamnosus* Lc705, a dairy strain that does not produce pili. The purified *spaC* and *spaB* components of the first cluster and *spaF* of the second have been shown to bind mucus (Kankainen et al., 2009; von Ossowski et al., 2010). Functional analysis using knockout mutants further proved that the SpaCBA pili, the only ones produced under *in vivo* conditions (Reunanen et al., 2012), were involved in biofilm formation and efficient adherence of *Lb. rhamnosus* GG to Caco-2 cells (Lebeer et al., 2012; Rasinkangas et al., 2014). The SrtC1 is the pilin-dedicated sortase recognizing and polymerizing the SpaA and SpaC pilin subunits. The sortase has also been found to be essential for pili formation (Rasinkangas et al., 2014). In addition, reduced expression of the pro-inflammatory interleukin-8 (IL-8) mRNA was induced when Caco-2 cells were incubated with the wild type strain as compared with the mutant.

This suggests that while providing mucus-binding ability, which may explain colonization and persistence of *L. rhamnosus* GG in the intestine, pili might also modulate IL-8 expression through interaction with surface molecules of the host cells (Lebeer et al., 2012). Therefore, in addition to a role in colonization, this system represents a previously undescribed mechanism for the interaction of probiotics with host immune tissue.

Genetic evidence for a putative pilus locus consisting in a sortase C gene flanked by 3 LPXTG protein encoding genes (*yhgD*, *yhgE* and *yhhB*) organized in an operon-like structure has been recently reported in the laboratory, plasmid-free derivative *Lc. lactis* IL 1403 (Dieye et al., 2010). Although *L. lactis* does not produce pili under standard growth conditions, overexpression of the pilus operon resulted in the display of pili on the surface of the cells (Oxaran et al., 2012). It is worth of mentioning that pilated IL 1403 exhibited an auto-aggregating phenotype in liquid cultures and formed a thicker biofilm than does the wild type, non-piliated strain. Functional analysis of the biogenesis machinery indicated that the pilus shaft was formed by oligomers of the YhgE pilin, the pilus cap was formed by YhgD and that YhhB was the basal pilin which allows tethering the pilus fibers to the cell wall (Oxaran et al., 2012). Analysis of wild isolates from plant and clinical environments showed that most strains could produce pili under normal culturing conditions, which suggests that these structures are spread on lactococci and might be pivotal for *L. lactis* to thrive in natural ecosystems.

The genus *Lactobacillus* is usually defined as not motile, although most than half of the species belonging to the *Lb. salivarius* clade (based on 16 rRNA gene phylogeny) have officially been recognized as motile by means of a sophisticated molecular structure, the flagellum (Cousin et al., 2015). The motility of *Lb. ruminis* ATTC 27782 has been particularly well studied at both the genomic and phenotypic levels (Forde et al., 2011, Neville et al., 2012). In this strain, all 45 genes required to produce a functional flagellum have been shown to be organized in a single operon (Forde et al., 2011). The same structure has recently been identified in a single strain of *Lb. curvatus*, a species of the *Lb. sakei* clade (Cousin et al., 2015). Motility might confer competitive advantages for niche colonization, such as a superior acquisition of nutrients or superior biofilm formation capability, but it could also have an impact on the ecology in terms of host signaling and colonization (Neville et al., 2012).

3.3.2.2 Respiration in *Lc. lactis* and other LAB species

Sugar fermentation was long considered to be the sole means of energy metabolism available to LAB, producing organic acids (mainly lactic acid) as a final end product. While this description is still true, some LAB species exhibit a respiratory capacity in the presence of oxygen and an

exogenous heme supply (Pedersen et al., 2012). LAB fermentation results in the production of greatly reduced amounts of lactic acid, higher biomass yield and improved fitness of starters.

Strictly speaking, respiration is the coupling of a membrane potential to the reduction of oxygen. Early evidence for the respiratory capability of LAB was largely overlooked (Sijpesteijn, 1970), and the initial reports on the respiration in *Lc. lactis* (Duwat et al., 2001) were essentially based on the analysis of the complete genome sequence of *Lc. lactis* IL1403, which revealed the presence of genes encoding enzymes related to the aerobic (pyruvate oxidase, NADH-dependent oxidase, NADH-dependent peroxidase) and respiratory (cytochrome oxidase, ubiquinone/menaquinone biosynthesis C-methylase) metabolism (Bolotin et al., 2001). Subsequent research confirmed that this species had indeed the ability to respire in the presence of oxygen (Duwat et al., 2001; Gaudu et al., 2002), provided the growth medium contains heme because this bacterium lacks a functional biosynthetic pathway for this compound. Transcriptomic analysis of *Lc. lactis* MG1363 showed that the pyruvate dehydrogenase complex (encoded by the *pdhABCD* operon) was up-regulated 4-fold under respiratory conditions (Vido et al., 2004). Acetolactate synthase (*als*) and α -acetolactate decarboxylase (*aldC*) genes were also up-regulated, facilitating the synthesis of both diacetyl and acetoin. The most highly up-regulated gene under respiratory conditions was *ygfC*, which encodes a putative regulatory protein that increased almost 100-fold. In contrast, the expression of the pyruvate formate lyase (*pfl*) and the alcohol dehydrogenase (*adhE*) genes was reduced 2.5- and 50-fold, respectively (Vido et al., 2004). Altogether these gene expression changes under respiratory conditions cause a profound rerouting of the *Lc. lactis* metabolism, producing acetate, acetoin and diacetyl from pyruvate at the expense of lactic acid (Duwat et al., 2001).

These results are of industrial significance and allowed the development of a patented process for the production of LAB starters (Duwat et al., 1998). The patent was licensed to Chr. Hansen in 1999, and the initial results obtained with the well-characterized laboratory strain *Lc. lactis* IL1403 were optimized for different industrial *Lc. lactis* strains and also for some *Leuconostoc* species (Pedersen et al., 2005). Industrial *Lc. lactis* strains were also assayed in aeration in the absence of added heme to distinguish simply aeration from true respiration. Numerous genes were differentially expressed under these two different conditions. Approximately half of these genes have an unknown function, indicating that more research is still needed to fully understand the physiology of respiration. Starter cultures obtained by the respiration technology have been assayed in pilot scale tests for Cheddar cheese production. Manufacture parameters were all within the normal range, and statistically significant differences between analytical parameters of cheeses (moisture content, total soluble nitrogen and pH) made from respiration-grown cells or

fermentation-grown cells were not observed. Indeed, sensory differences of ripened cheeses were not perceived by two trained sensory panels. Industrial scale trials of Cheddar, Feta and cottage cheese have already been performed, and again no significant differences were revealed in manufacturing parameters, cheese microbiology, chemistry, texture or flavor development (Pedersen et al., 2005). More recently, other phenotypic properties of technological relevance, such as the production of the antimicrobial bacteriocin nisin, have been reported to be enhanced under respiratory conditions (Kördikanlıoğlu et al., 2015). This result might also have industrial significance for the development of nisin-based protective cultures.

Respiration in *Lb. plantarum* needs the exogenous addition of both heme and vitamin K₂ (which act as a source of menaquinone) (Brooijmans et al., 2009). As this bacterium lacks superoxide dismutase, high levels of manganese are also needed for its aerobic growth. Under respiratory conditions, growth of *Lb. plantarum* results in higher biomass yields. It also impacts on the robustness of the cells, as strains from aerobic cultures have been shown to be more resistant to industrially-relevant stress conditions (Watanabe et al., 2012). A similar respiration process has been reported for some *Lb. casei* strains, also requiring both hemin and menaquinone (Zotta et al., 2014). Respiration in this latter species resulted in the expression of phenotypes with enhanced technological properties, such as increased survival to stress and higher antioxidant capability (Ianniello et al., 2015), which can have application in the use of these bacterium as starter or probiotic. Remarkably, the genome of *Oenococcus oeni* also contains all genes for aerobic respiration (Borneman et al., 2012), but its functionality has yet to be tested.

Increased biomass after aerated incubation in the presence of heme has never been obtained for *S. thermophilus*, *Lb. delbrueckii* subsp. *bulgaricus* and *Lb. helveticus*. Congruently, analysis of the complete genome sequences of strains of these species (Bolotin et al., 2004; van de Guchte et al., 2006; Callanan et al., 2008, respectively) neither revealed the presence of genes for cytochrome oxidase nor for the biosynthesis of quinones (Pedersen et al., 2012), essential features for respiration.

3.3.2.3 Natural transformation of *S. thermophilus*

S. thermophilus is considered the second most important industrial dairy starter after *Lc. lactis* due to its extensive use for the manufacture of dairy products such as yogurt, hard cooked cheeses of the Italian and Swiss types, soft cheeses, etc. (Leroy and de Vuyst, 2004). As stated above, this bacterium is closely related to *S. salivarius* and neighbour of pathogenic species, such as *Streptococcus pyogenes* and *Streptococcus pneumoniae* (Bolotin et al., 2004). It is worth mentioning that the latter bacterium was the microorganism in which natural competence was first

discovered. Since then, this process has been described in many Gram positive and Gram negative bacteria, and -although most rarely- also in archaea (Johnston et al., 2014). Competence is defined as a transient physiological state that enables cells to take up exogenous, naked DNA and to stably integrate it into the genome by homologous recombination (Chen and Dubnau, 2004). Currently, the process is considered as a stress response process that may increase adaptability and fitness under adverse conditions (Charpentier et al., 2012). In response to specific environmental stresses, streptococci species synthesize and secrete peptide competence pheromones, which, at a critical extracellular concentration, activate the master competence regulator (ComX). This protein is a sigma factor that associates with the RNA polymerase to re-direct transcription towards genes required for DNA transport, processing and integration (Fontaine et al., 2015).

Genome analysis of the first *S. thermophilus* sequenced strains proved this species still share a substantial part of their overall genetic make-up with other streptococci, which included all genes necessary for competence (Bolotin et al., 2004). *S. thermophilus* contains a *comX*-like gene coding for a typical peptide pheromone-dependent two-component system that is similar to the competence control loci of *Streptococcus mutans* and *S. pyogenes*. The regulatory pathway controlling expression of key components of competence in *S. thermophilus* has recently been reported (Haustenne et al., 2015). The pheromone (called XIP peptide) is synthesized as a large peptide that is exported through an unknown transport system and matured by specific proteases. The resulting extracellular XIP is reimported by the oligopeptide transporter Opp/Ami, an essential component of the proteolytic system to grow in milk that also results necessary for in competence (Gardan et al., 2009). In the cytoplasm, the intracellular mature XIP binds to the transcriptional regulator ComR, and the XIP-ComR complex activates the transcription of most early competence genes (including *comS* and *comX*). Thereafter, ComX induces transcription of all late competence genes necessary for DNA transformation (Haustenne et al., 2015).

The mechanistic characterization of natural competence in *S. thermophilus* has allowed the development of genetic tools for introduction of heterologous genes in this bacterium by natural transformation without the need of antibiotic-resistance or other controversial markers for the selection of transformants (Blomqvist et al., 2010; Fontaine et al., 2010). Subsequently, insertion of the DNA into the chromosome takes place by double cross-over, homologous recombination. The ability of *S. thermophilus* to be naturally transformable allows obtaining recombinant strains easily (Lecomte et al., 2016). As a proof of concept, the gene encoding the cell-envelope-located proteinase PrtS, which is only present in certain fast milk-acidifying *S. thermophilus* strains (Delorme et al., 2010), has experimentally been transferred to a series of slow-acidifying starter strains (Dandoy et al., 2011).

3.4 Concluding remarks

In this book chapter a review of different aspects of genome-wide studies on strains relevant for dairy and probiotic products has been performed. This has shown that an incredibly high number of genome sequences is accumulating, which have been or still have to be analyzed in scientific studies. Those are extremely important to understand evolution and define diversity among strains at different taxonomic levels, but, most importantly could deeply enhance our knowledge on metabolic capabilities, e.g., lactose and citrate metabolism, proteolysis, lipolysis, bacteriocin production, and so on, that can eventually be linked to their genetic basis. This is the starting point of possible combinations with other omics data (proteomics, transcriptomics, metabolomics, etc.) that could be employed to devise genome scale metabolic models. This, coupled with traditional culture-based characterizations and small-scale fermentations, could enable a more precise strain selection for efficient fermentation, flavor development at industrial scale and probiotic actions (Stefanovic et al., 2017).

Remarkably, the need for large sequence data manipulation has led to the development of a novel area of competence for microbiology, i.e. bioinformatics, with many microbiologists operating in front of computers instead of a bench. In line with this observation, it can be noted that an EU project has recently been concluded (2015), funded in the FP7-SME framework (grant agreement n. 6054853 - FP7/2007-2013), aimed to develop a genomics toolbox to enhance business for SMEs in the market of starter cultures and probiotics (<http://www.genobox.eu/ns/>). Therefore, genome sequencing is recognized as a useful tool not only for scientific purposes, which could be pivotal for patent opportunities and reliable communication to authorities and customers/consumers (Danielsen & Johansen, 2009). Genome sequencing continues providing an incredible amount of genetic information for the characterization of starter and probiotics that will have to be deciphered in the coming decades.

References

Abriouel H, Lavilla Lerma L, Casado Muñoz M.C, Pérez Montoro B, Kabisch J, Pichner R, Cho G-S, Neve H, Fusco V, Franz CMAP, Gálvez A, Benomar N. 2015. The controversial nature of the *Weissella* genus: technological and functional aspects versus whole genome analysis-based pathogenic potential for their application in food and health. *Front. Microbiol.* 6: 1197.

- Ainsworth S, Stockdale S, Bottacini F, Mahony J, Sinderen D. 2014. The *Lactococcus lactis* plasmidome: much learnt, yet still lots to discover. *FEMS Microbiol. Rev.* 38: 1066-1088.
- Alkema W, Boekhorst J, Wels M, van Hijum SAFT. 2015. Microbial bioinformatics for food safety and production. *Brief. Bioinform.* 17: 283-292.
- Barbour A, Philip K. 2014. Variable characteristics of bacteriocin-producing *Streptococcus salivarius* strains isolated from Malaysian subjects. *PLoS One* 9: e100541.
- Barrangou R, Horvath P. 2012. CRISPR: new horizons in phage resistance and strain identification. *Annu. Rev. Food Sci. Technol.* 3:143-62.
- Barrangou R. 2014. Cas9 targeting and the CRISPR revolution. *Science* 344: 707-8.
- Bennedsen M, Stuer-Lauridsen B, Danielsen M, Johansen E. 2011. Screening for antimicrobial resistance genes and virulence factors via genome sequencing. *Appl. Environ. Microbiol.* 77: 2785-2787.
- Blomqvist T, Steinmoen H, Håvarstein LS. 2010. A food-grade site-directed mutagenesis system for *Streptococcus thermophilus* LMG 18311. *Lett. Appl. Microbiol.* 50: 314-319.
- Bolotin A, Quinquis B, Renault P, Sorokin A, Ehrlich SD, Kulakauskas S, Lapidus A, Goltsman E, Mazur M, Pusch GD, Fonstein M, Overbeek R, Kyprides N, Purnelle B, Prozzi D, Ngui K, Masuy D, Hancy F, Burtreau S, Boutry M, Delcour J, Goffeau A, Hols P. 2004. Complete sequence and comparative genome analysis of the dairy bacterium *Streptococcus thermophilus*. *Nat. Biotechnol.* 22: 1554-1558.
- Bolotin A, Wincker P, Mauger S, Jaillon O, Malarne K, Weissenbach J, Ehrlich SD, Sorokin A. 2001. The complete genome sequence of the lactic acid bacterium *Lactococcus lactis* ssp. *lactis* IL1403. *Genome Res.* 11: 731-753.
- Borneman AR, McCarthy JM, Chambers PJ, and Bartowsky EJ. 2012. Comparative analysis of the *Oenococcus oeni* pan genome reveals genetic diversity in industrially-relevant pathways. *BMC Genomics* 13: 373.

- Boutrou R, Guéguen M. 2005. Interests in *Geotrichum candidum* for cheese technology. Int. J. Food Microbiol. 102: 1-20.
- Broadbent JR, Neeno-Eckwall EC, Stahl B, Tandee K, Cai H, Morovic W, Horvath P, Heidenreich J, Perna NT, Barrangou R, Steele JL. 2012. Analysis of the *Lactobacillus casei* supragenome and its influence in species evolution and lifestyle adaptation. BMC Genomics 13: 533.
- Brooijmans RJ, de Vos WM, Hugenholtz J. 2009. *Lactobacillus plantarum* WCFS1 electron transport chains. Appl. Environ. Microbiol. 75: 3580-3585.
- Burton JP, Wescombe PA, Moore CJ, Chilcott CN, Tagg JR. 2006. Safety assessment of the oral cavity probiotic *Streptococcus salivarius* K12. Appl. Environ. Microbiol. 72: 3050-3053.
- Callanan M, Kaleta P, O'Callaghan J, O'Sullivan O, Jordan K, McAuliffe O, Sangrador-Vegas A, Slattery L, Fitzgerald GF, Beresford T, Ross RP. 2008. Genome sequence of *Lactobacillus helveticus*, an organism distinguished by selective gene loss and insertion sequence element expansion. J. Bacteriol. 190: 727-735.
- Cavanagh D, Fitzgerald GF, McAuliffe O. 2015. From field to fermentation: The origins of *Lactococcus lactis* and its domestication to the dairy environment. Food Microbiol. 47: 45-61.
- Charpentier X, Polard P, Claverys JP. 2012. Induction of competence for genetic transformation by antibiotics: convergent evolution of stress responses in distant bacterial species lacking SOS? Curr. Opin. Microbiol. 15: 570-576.
- Chen I, Dubnau D. 2004. DNA uptake during bacterial transformation. Nat. Rev. Microbiol. 2: 241-249.
- Chenoll E, Codoñer FM, Silva A, Ibáñez A, Martinez-Blanch JF, Bollati-Fogolín M, Crispo M, Ramírez S, Sanz Y, Ramón D, Genovés S. 2013. Genomic sequence and pre-clinical safety assessment of *Bifidobacterium longum* CECT 7347, a probiotic able to reduce the toxicity and inflammatory potential of gliadin-derived peptides. J. Prob. Health 1: 2.
- Cousin FJ, Lynch SN, Harris HMB, McCann A, Lynch DB, Neville BA, Irisawa, T, Okada, S, Endo A, O'Toole PW. 2015. Detection and genomic characterization of motility in *Lactobacillus*

- curvatus*: confirmation of motility in a species outside of the *Lactobacillus salivarius* clade. Appl. Environ. Microbiol. 81: 1297-1308.
- Dandoy D, Fremaux C, de Frahan MH, Horvath P, Boyabal P, Hols P, Fontaine L. 2011. The fast milk acidifying phenotype of *Streptococcus thermophilus* can be acquired by natural transformation of the genomic island encoding the cell-envelope proteinase PrtS. Microb. Cell Fact. 10: S21.
- Danielsen M, Johansen E. 2009. Functional genomics of dairy micro-organisms and probiotics in the era of low-cost DNA sequencing. Aust. J. Dairy Technol. 64: 102-105.
- Dellaglio F, Felis GE, Torriani S, Sørensen K, Johansen E. 2005. Genomic Characterisation of Starter Cultures. In: SDT Book - Probiotic Dairy Products. Editor A. Tamime. Blackwell Publishing Ltd, 2005, (Chapter 2) pp. 16-38. ISBN 1-4051-2124-6.
- Delorme C, Abraham AL, Renault P, Guédon E. 2015. Genomics of *Streptococcus salivarius*, a major human commensal. Infect. Genet. Evol. 33: 381-392.
- Delorme C, Bartholini C, Bolotine A, Ehrlich SD, Renault P. 2010. Emergence of a cell wall protease in the *Streptococcus thermophilus* population. Appl. Environ. Microbiol. 76: 451-460.
- Dieye Y, Oxaran V, Ledue-Clier F, Alkhalaf W, Buist G, Juillard V, Lee CW, Piard JC. 2010. Functionality of sortase A in *Lactococcus lactis*. Appl. Environ. Microbiol. 76: 7332-7337.
- Domann E, Hain T, Ghai R, Billion A, Kuenne C, Zimmermann K, Chakraborty T. 2007. Comparative genomic analysis for the presence of potential enterococcal virulence factors in the probiotic *Enterococcus faecalis* strain Symbioflor 1. Int. J. Med Microbiol. 297: 533-539.
- Duwat P, Sourice S, Cesselin B, Lamberet G, Vido K, Gaudu P, Le Loir Y, Violet F, Loubière P, Gruss A. 2001. Respiration capacity of the fermenting bacterium *Lactococcus lactis* and its positive effects on growth and survival. J. Bacteriol. 183: 4509-4516.
- Duwat P, Sourice S, Gruss, A. 1998. Process for preparing starter cultures of lactic acid bacteria. French Patent Application FR9809463.

- EFSA. 2007. Opinion of the Scientific Committee on a request from EFSA on the introduction of a Qualified Presumption of Safety (QPS) approach for assessment of selected microorganisms referred to EFSA. EFSA J. 587: 1-16.
- EFSA. 2012. Guidance on the safety assessment of *Enterococcus faecium* in animal nutrition. EFSA J. 10: 2682-2692.
- EFSA. 2013. Scientific opinion on the maintenance of the list of QPS biological agents intentionally added to food and feed (2013 update). EFSA J. 11: 3449.
- Falentin H, Deutsch SM, Jan G, Loux V, Thierry A, Parayre S, Maillard MB, Dherbecourt J, Cousin FJ, Jardin J (2010) The complete genome of *Propionibacterium freudenreichii* CIRMBIA1^T, a hardy *Actinobacterium* with food and probiotic applications. PLoS One 5: e11748
- FAO/WHO. 2001. Health and nutritional properties of probiotics in food including powder milk with live lactic acid bacteria. World Health Organization [online], http://www.who.int/foodsafety/publications/fs_management/en/probiotics.pdf.
- Ferrario C, Milani C, Mancabelli L, Lugli GA, Turrone F, Duranti S, Mangifesta M, Viappiani A, Sinderen D, Ventura M. 2015. A genome-based identification approach for members of the genus *Bifidobacterium*. FEMS Microbiol. Ecol. 91: fiv009.
- Ferrario C, Ricci G, Milani C, Lugli GA, Ventura M, Eraclio G, Borgo F, Fortina MG. 2013. *Lactococcus garvieae*: where is it from? A first approach to explore the evolutionary history of this emerging pathogen PLoS One 8: e84796.
- Field D, Amaral-Zettler L, Cochrane G, Cole JR, Dawyndt P, Garrity GM, Gilbert J, Glöckner FO, Hirschman L, Karsch-Mizrachi I, Klenk HP, Knight R, Kottmann R, Kyrpides N, Meyer F, San Gil I, Sansone SA, Schriml LM, Sterk P, Tatusova T, Ussery DW, White O, Wooley J. 2011. The Genomic Standards Consortium. PLoS Biol. 9: e1001088.

- Flórez AB, Campedelli I, Delgado S, Alegría Á, Salvetti E, Felis GE, Mayo B, Torriani S. 2016. Antibiotic susceptibility profiles of dairy leuconostoc, analysis of the genetic basis of atypical resistances and transfer of genes *in vitro* and in a food matrix. PLoS One 11: e0145203.
- Fontaine L, Dandoy D, Boutry C, Delplace B, de Frahan MH, Fremaux C, Horvath P, Boyaval P, Hols P. 2010. Development of a versatile procedure based on natural transformation for marker-free targeted genetic modification in *Streptococcus thermophilus*. Appl. Environ. Microbiol. 76: 7870-7877.
- Fontaine L, Wahl A, Flécharde M, Mignolet J, Hols P. 2015. Regulation of competence for natural transformation in streptococci. Infect. Genet. Evol. 33: 343-360.
- Forde BM, Neville BA, O'Donnell MM, Riboulet-Bisson E, Claesson MJ, Coghlan A, Ros RP, O'Toole PW. 2011. Genome sequences and comparative genomics of two *Lactobacillus ruminis* strains from the bovine and human intestinal tract. Microb. Cell Fact. 10: S13.
- Franz CM, Huch M, Abriouel H, Holzapfel W, Gálvez A. 2011. Enterococci as probiotics and their implications in food safety. Int. J. Food Microbiol. 151: 125-140.
- Fusco V, Quero GM, Cho G, Kabisch J, Meske D, Neve H, Bockelmann W, Franz CMAP. 2015. The genus *Weissella*: taxonomy, ecology and biotechnological potential. Front. Microbiol. 6: 155.
- Gardan R, Besset C, Guillot A, Gitton C, Monnet V. 2009. The oligopeptide transport system is essential for the development of natural competence in *Streptococcus thermophilus*. J. Bacteriol. 191: 4647-4655.
- Garrigues C, Johansen E, Crittenden R. 2013. Pangenomics – an avenue to improved industrial starter cultures and probiotics. Curr. Opin. Biotechnol. 24: 187-191.
- Gasson, M.J., de Vos, W.M. 2004. *Genetics and Biotechnology of Lactic Acid Bacteria*. Springer, Berlin. ISBN: 978-94-010-4584-1.

- Gaudu P, Vido K, Cesselin B, Kulakauskas S, Tremblay J, Rezaïki L, Lamberret G, Sourice S, Duwat P, Gruss A. 2002. Respiration capacity and consequences in *Lactococcus lactis*. *Antonie van Leeuwenhoek* 82: 263-269.
- Giraffa G. 2003. Functionality of enterococci in dairy products. *Int. J. Food Microbiol.* 88:215-222.
- Gueimonde M, Florez AB, van Hoek AH, Stuer-Lauridsen B, Stroman P, de los Reyes-Gavilan CG, Margolles A. 2010. Genetic basis of tetracycline resistance in *Bifidobacterium animalis* subsp. *lactis*. *Appl. Environ. Microbiol.* 76: 3364–3369.
- Hanniffy S, Wiedermann U, Repa A, Mercenier A, Daniel C, Fioramonti J, Tlaskolova H, Kozakova H, Israelsen H, Madsen S, Vrang A, Hols P, Delcour J, Bron P, Kleerebezem M, Wells J. 2004. Potential and opportunities for use of recombinant lactic acid bacteria in human health. *Adv. Appl. Microbiol.* 56: 1-64.
- Hao P, Zheng H, Yu Y, Ding G, Gu W, Chen S, Yu Z, Ren S, Oda M, Konno T, Wang S, Li X, Ji ZS, Zhao G. 2011. Complete sequencing and pan-genomic analysis of *Lactobacillus delbrueckii* subsp. *bulgaricus* reveal its genetic basis for industrial yogurt production. *PLoS One* 6: e15964.
- Hatoum R, Labrie S, Fliss I. 2012. Antimicrobial and probiotic properties of yeasts: from fundamental to novel applications. *Front. Microbiol.* 3: 421.
- Haustenne L, Bastin G, Hols P, Fontaine L. 2015. Modeling of the ComRS signaling pathway reveals the limiting factors controlling competence in *Streptococcus thermophilus*. *Front. Microbiol.* 6: 1413.
- Hill C, Guarner F, Reid G, Gibson GR, Merenstein DJ, Pot B, Morelli L, Canani RB, Flint HJ, Salminen S, Calder PC, Sanders ME. 2014. Expert consensus document. The International Scientific Association for Probiotics and Prebiotics consensus statement on the scope and appropriate use of the term probiotic. *Nat. Rev. Gastroenterol. Hepatol.* 11: 506-514.
- Hols P, Hancy F, Fontaine L, Grossiord B, Prozzi D, Leblond-Bourget N, Decaris B, Bolotin A, Delorme C, Dusko ES. 2005. New insights in the molecular biology and physiology of

- Streptococcus thermophilus* revealed by comparative genomics FEMS Microbiol. Rev. 29: 435-463.
- Horvath P, Couté-Monvoisin AC, Romero DA, Boyaval P, Fremaux C, Barrangou R. 2009. Comparative analysis of CRISPR loci in lactic acid bacteria genomes. Int. J. Food Microbiol. 131: 62-70.
- Hugenholtz J, Sybesma W, Groot MN, Wisselink W, Ladero V, Burgess K, van Sinderen D, Piard JC, Eggink G, Smid EJ, Savoy G, Sesma F, Jansen T, Hols P, Kleerebezem M. 2002. Metabolic engineering of lactic acid bacteria for the production of nutraceuticals. Antonie van Leeuwenhoek 82: 217-235.
- Ianniello RG, Ricciardi A, Parente E, Tramutola A, Reale A, Zotta T. 2015. Aeration and supplementation with heme and menaquinone affect survival to stresses and antioxidant capability of *Lactobacillus casei* strains. LWT Food Sci. Technol. 60: 817-824.
- Jans C, Follador R, Hochstrasser M, Lacroix C, Meile L, Stevens MJA. 2013. Comparative genome analysis of *Streptococcus infantarius* subsp. *infantarius* CJ18, an African fermented camel milk isolate with adaptations to dairy environment. BMC Genomics 14: 200.
- Jansen R, Embden JD, Gaastra W, Schouls LM. 2002. Identification of genes that are associated with DNA repeats in prokaryotes. Mol. Microbiol. 43: 1565-1575.
- Johnston C, Martin B, Fichant G, Polard P, Claverys JP. 2014. Bacterial transformation: distribution, shared mechanisms and divergent control. Nat. Rev. Microbiol. 12: 181-196.
- Kankainen M, Paulin L, Tynkkynen S, von Ossowski I, Reunanen J, Partanen P, Satokari R, Vesterlund S, Hendrickx AP, Lebeer S, De Keersmaecker SC, Vanderleyden J, Hämäläinen T, Laukkanen S, Salovuori N, Ritari J, Alatalo E, Korpela R, Mattila-Sandholm T, Lassig A, Hatakka K, Kinnunen KT, Karjalainen H, Saxelin M, Laakso K, Surakka A, Palva A, Salusjärvi T, Auvinen P, de Vos WM. 2009. Comparative genomic analysis of *Lactobacillus rhamnosus* GG reveals pili containing a human- mucus binding protein. Proc. Natl. Acad. Sci. USA 106: 17193-17198.

- Kant R, Blom J, Palva A, Siezen RJ, de Vos WM. 2010. Comparative genomics of *Lactobacillus*. Microbiol. Biotechnol. 4: 323-332.
- Kelleher P, Murphy J, Mahony J, van Sinderen D. 2015. Next-generation sequencing as an approach to dairy starter selection. Dairy Sci. Technol. 95: 545-568.
- Kelly WJ, Ward LJH, Leahy SC. 2010. Chromosomal diversity in *Lactococcus lactis* and the origin of dairy starter cultures. Genome Biol. Evol. 2: 729-744.
- Klaenhammer TR, Barrangou R, Buck BL, Azcaráte-Peril MA, Altermann E. 2005. Genomic features of lactic acid bacteria effecting bioprocessing and health. FEMS Microbiol. Rev. 29: 393-409.
- Klijn A, Mercenier A, Arigoni F. 2005. Lessons from the genomes of bifidobacteria. FEMS Microbiol. Rev. 29: 491-509.
- Kopit LM, Kim EB, Siezen RJ, Harris LJ, Marco ML. 2014. Safety of the surrogate microorganism *Enterococcus faecium* NRRL B-2354 for use in thermal process validation. Appl. Environ. Microbiol. 80: 1899-1909.
- Kördikanlıoğlu B, Şimşek Ö, Saris PE. 2015. Nisin production of *Lactococcus lactis* N8 with hemin-stimulated cell respiration in fed-batch fermentation system. Biotechnol. Prog. 31: 678-685.
- Lebeer S, Claes I, Tytgat HL, Verhoeven TL, Marien E, von Ossowski I, Reunanen J, Palva A, Vos WM, Keersmaecker SC, Vanderleyden J. 2012. Functional analysis of *Lactobacillus rhamnosus* GG pili in relation to adhesion and immunomodulatory interactions with intestinal epithelial cells. Appl. Environ. Microbiol. 78: 185-193.
- Lecomte X, Gagnaire V, Lortal S, Dary A, Genay M. 2016. *Streptococcus thermophilus*, an emerging and promising tool for heterologous expression: Advantages and future trends. Food Microbiol. 53: 2-9.
- Leroy F, de Vuyst L. 2004. Lactic acid bacteria as functional starter cultures for the food fermentation industry. Trends Food Sci. Technol. 15: 67-78.

- Lukjancenko O, Ussery DW, Wassenaar TM. 2012. Comparative genomics of *Bifidobacterium*, *Lactobacillus* and related probiotic genera. *Microbiol. Ecol.* 63: 651-73.
- Makarova K, Slesarev A, Wolf Y, Sorokin A, Mirkin B, Koonin E, Pavlov A, Pavlova N, Karamychev V, Polouchine N, Shakhova V, Grigoriev I, Lou Y, Rohksar D, Lucas S, Huang K, Goodstein DM, Hawkins T, Plengvidhya V, Welker D, Hughes J, Goh Y, Benson A, Baldwin K, Lee JH, Díaz-Muñoz I, Dosti B, Smeianov V, Wechter W, Barabote R, Lorca G, Altermann E, Barrangou R, Ganesan B, Xie Y, Rawsthorne H, Tamir D, Parker C, Breidt F, Broadbent J, Hutkins R, O'Sullivan D, Steele J, Unlu G, Saier M, Klaenhammer T, Richardson P, Kozyavkin S, Weimer B, Mills D. 2006. Comparative genomics of the lactic acid bacteria. *Proc. Natl. Acad. Sci. USA.* 103: 15611-15616.
- Marshall VM. 1991. Inoculated ecosystem in a milk environment. *J. Appl. Bacteriol.* 73: 127S-135S.
- Medini D, Donati C, Tettelin H, Massignani V, Rappuoli R. 2005. The microbial pan-genome. *Curr. Opin. Genet. Dev.* 15: 589-594.
- Milani C, Lugli GA, Duranti S, Turrone F, Bottacini F, Mangifesta M, Sanchez B, Viappiani A, Mancabelli L, Taminiau B, Delcenserie V, Barrangou R, Margolles A, van Sinderen D, Ventura M. 2014. Genomic encyclopedia of type strains of the genus *Bifidobacterium*. *Appl. Environ. Microbiol.* 80: 6290-6302.
- Miquel S, Beaumont M, Martín R, Langella P, Braesco V, Thomas M. 2015. A proposed framework for an appropriate evaluation scheme for microorganisms as novel foods with a health claim in Europe. *Microb. Cell Fact.* 14: 48.
- Neville BA, Forde BM, Claesson MJ, Darby T, Coghlan A, Nally K, Ros RP, O'Toole PW. 2012. Characterization of pro-inflammatory flagellin proteins produced by *Lactobacillus ruminis* and related motile strains. *PLoS One* 7: e40592.
- O'Connell Motherway M, Zomer A, Leahy SC, Reunanen J, Bottacini F, Claesson MJ, O'Brien F, Flynn K, Casey PG, Munoz JA, Kearney B, Houston AM, O'Mahony C, Higgins DG, Shanahan

- F, Palva A, de Vos WM, Fitzgerald GF, Ventura M, O'Toole PW, van Sinderen D. 2011. Functional genome analysis of *Bifidobacterium breve* UCC2003 reveals type IVb tight adherence (Tad) pili as an essential and conserved host-colonization factor. *Proc. Natl. Acad. Sci. USA* 108: 11217-11222.
- Oh JH, van Pijkeren JP. 2014. CRISPR-Cas9-assisted recombineering in *Lactobacillus reuteri*. *Nucleic Acids Res.* 42: e131.
- O'Sullivan O, O'Callaghan J, Sangrador-Vegas A, McAuliffe O, Slattery L, Kaleta P, Callanan M, Fitzgerald GF, Ross RP, Beresford T. 2009. Comparative genomics of lactic acid bacteria reveals a niche-specific gene set. *BMC Microbiol.* 9: 50.
- Oxaran V, Ledue-Clier F, Dieye Y, Herry JM, Péchoux C, Meylheuc T, Briandet R, Juillard V, Piard JC. 2012. Pilus biogenesis in *Lactococcus lactis*: molecular characterization and role in aggregation and biofilm formation. *PLoS One.* 7: e50989.
- Papadimitriou K, Anastasiou R, Mavrogonatou E, Blom J, Papandreou NC, Hamodrakas SJ, Ferreira S, Renault P, Supply P, Pot B, Tsakalidou E. 2014. Comparative genomics of the dairy isolate *Streptococcus macedonicus* ACA-DC 198 against related members of the *Streptococcus bovis*/*Streptococcus equinus* complex. *BMC Genomics* 15: 272.
- Pariza MW, Gillies KO, Kraak-Ripple SF, Leyer G, Smith AB. 2015. Determining the safety of microbial cultures for consumption by humans and animals. *Regul. Toxicol. Pharmacol.* 73: 164-171.
- Passerini D, Beltramo C, Coddeville M, Quentin Y, Ritzenthaler P, Daveran-Mingot ML, Le Bourgeois P. 2010. Genes but not genomes reveal bacterial domestication of *Lactococcus lactis*. *PLoS One* 5: e15306.
- Passerini D, Coddeville M, Le Bourgeois P, Loubière P, Ritzenthaler P, Fontagné-Faucher C, Daveran-Mingot ML, Coccagn-Bousquet M. 2013. The carbohydrate metabolism signature of *Lactococcus lactis* strain A12 reveals its sourdough ecosystem origin *Appl. Environ. Microbiol.* 79: 5844-5852.

- Pedersen MB, Gaudu P, Lechardeur D, Petit MA, Gruss A. 2012. Aerobic respiration metabolism in lactic acid bacteria and uses in biotechnology. *Annu. Rev. Food Sci. Technol.* 3: 37-58.
- Pedersen MB, Iversen SL, Sorensen KI, Johansen E. 2005. The long and winding road from the research laboratory to industrial applications of lactic acid bacteria. *FEMS Microbiol. Rev.* 29: 611-624.
- Pérez T, Balcazar JL, Peix A, Valverde A, Velazquez E, De Bleas I, Ruiz-Zarzuela I. 2011. *Lactococcus lactis* subsp. *truttae* subsp. nov. isolated from the intestinal mucus of brown trout (*Salmo trutta*) and rainbow trout (*Oncorhynchus mykiss*). *Int. J. Syst. Evol. Microbiol.* 61: 1894-1898.
- Poonam, Pophaly SD, Tomar SK, De S, Singh R. 2012. Multifaceted attributes of dairy propionibacteria: a review. *World J. Microbiol. Biotechnol.* 28: 3081-3095.
- Proft T, Baker EN. 2009. Pili in Gram-negative and Gram-positive bacteria. Structure, assembly and their role in disease. *Trends Microbiol.* 16: 33-40.
- Rasinkangas P, Reunanen J, Douillard FP, Ritari J, Uotinen V, Palva A, de Vos WM. 2014. Genomic characterization of non-mucus-adherent derivatives of *Lactobacillus rhamnosus* GG reveals genes affecting pilus biogenesis. *Appl. Environ. Microbiol.* 80: 7001-7009.
- Reddy TBK, Thomas A, Stamatis D, Bertsch J, Isbandi M, Jansson J, Mallajosyula J, Pagani I, Lobos E, Kyrpides N. 2015. The Genomes OnLine Database (GOLD) v.5: a metadata management system based on a four level (meta)genome project classification. *Nucleic Acids Res.* 43: D1099-D1106.
- Reunanen J, von Ossowski I, Hendrickx AP, Palva A, de Vos WM. 2012. Characterization of the SpaCBA pilus fibers in the probiotic *Lactobacillus rhamnosus* GG. *Appl. Environ. Microbiol.* 78: 2337-2344.
- Richards VP, Palmer SR, Pavinski Bitar PD, Qin X, Weinstock GM, Highlander SK, Town CD, Burne RA, Stanhope MJ. 2014. Phylogenomics and the dynamic genome evolution of the genus *Streptococcus*. *Genome Biol. Evol.* 6: 741-753.

- Salvetti E, Orrù L, Capozzi V, Martina A, Lamontanara A, Keller D, Cash H, Felis GE, Cattivelli L, Torriani S, Spano G. 2016. Integrate genome-based assessment of safety for probiotic strains: *Bacillus coagulans* GBI-30, 6086 as a case study. *Appl. Microbiol. Biotechnol.* 100: 4595-4605.
- Schell MA, Karmirantzou M, Snel B, Vilanova D, Berger B, Pessi G, Zwahlen MC, Desiere F, Bork P, Delley M, Pridmore RD, Arigoni F. 2002. The genome sequence of *Bifidobacterium longum* reflects its adaptation to the human gastrointestinal tract. *Proc. Natl. Acad. Sci. USA* 99: 14422-14427.
- Schlegel L, Grimont F, Ageron E, Grimont PA, Bouvet A. 2003. Reappraisal of the taxonomy of the *Streptococcus bovis*/*Streptococcus equinus* complex and related species: description of *Streptococcus gallolyticus* subsp. *gallolyticus* subsp. nov., *S. gallolyticus* subsp. *macedonicus* subsp. nov. and *S. gallolyticus* subsp. *pasteurianus* subsp. nov. *Int. J. Syst. Evol. Microbiol.* 53: 631-645.
- Schleifer KH, Ehrmann M, Krusch U, Neve H. 1991. Revival of the species *Streptococcus thermophilus* (ex Orla-Jensen, 1919) nom. rev. *Syst. Appl. Microbiol.* 14: 386-388.
- Senan S, Prajapati JB, Joshi CG. 2015a. Whole-genome based validation of the adaptive properties of Indian origin probiotic *Lactobacillus helveticus* MTCC 5463. *J. Sci. Food Agric.* 95: 321-328.
- Senan S, Prajapati JB, Joshi CG. 2015b. Feasibility of genome-wide screening for biosafety assessment of probiotics: a case study of *Lactobacillus helveticus* MTCC 5463. *Probiotics Antimicrob. Proteins* 7: 249-258.
- Siezen RJ, Renckens B, Van Swam I, Peters S, van Kranenburg R, Kleerebezem M, de Vos WM. 2005. Complete sequences of four plasmids of *Lactococcus lactis* subsp. *cremoris* SK11 reveal extensive adaptation to the dairy environment *Appl. Environ. Microbiol.* 71: 8371-82.
- Siezen RJ, Wilson G. 2010. Probiotics Genomics. *Microb. Biotechnol.* 3: 1-9.
- Sijpesteijn AK. 1970. Induction of cytochrome formation and stimulation of oxidative dissimilation by hemin in *Streptococcus lactis* and *Leuconostoc mesenteroides*. *Antonie van Leeuwenhoek* 36: 335-348.

- Smokvina T, Wels M, Polka J, Chervaux C, Brisse S, Boekhorst J, van Hylckama Vlieg JE, Siezen RJ. 2013. *Lactobacillus paracasei* comparative genomics: towards species pan-genome definition and exploitation of diversity. PLoS One 8: e68731.
- Stanton C, Ross RP, Fitzgerald GF, van Sinderen D. 2005. Fermented functional foods. Curr. Opin. Biotechnol. 16: 198-203.
- Stefanovic E., Fitzgerald G., McAuliffe O. 2017. Advances in the genomics and metabolomics of dairy lactobacilli: A review. Food Microbiol. 61: 33-49.
- Sun Z, Harris HM, McCann A, Guo C, Argimón S, Zhang W, Yang X, Jeffery IB, Cooney JC, Kagawa TF, Liu W, Song Y, Salvetti E, Wrobel A, Rasinkangas P, Parkhill J, Rea MC, O'Sullivan O, Ritari J, Douillard FP, Paul Ross R, Yang R, Briner AE, Felis GE, de Vos WM, Barrangou R, Klaenhammer TR, Caufield PW, Cui Y, Zhang H, O'Toole PW. 2015. Expanding the biotechnology potential of lactobacilli through comparative genomics of 213 strains and associated genera. Nat. Commun. 6: 8322.
- Turroni F, Milani C, van Sinderen D, Ventura M. 2011. Genetic strategies for mucin metabolism in *Bifidobacterium bifidum* PRL2010: an example of possible human-microbe co-evolution. Gut Microbes 2: 183-189.
- van de Guchte M, Penaud S, Grimaldi C, Barbe V, Bryson K, Nicolas P, Robert C, Oztas S, Mangenot S, Couloux A, Loux V, Dervyn R, Bossy R, Bolotin A, Batto JM, Walunas T, Gibrat JF, Bessires P, Weissenbach J, Ehrlich SD, Maguin E. 2006. The complete genome sequence of *Lactobacillus bulgaricus* reveals extensive and ongoing reductive evolution. Proc. Natl. Acad. Sci. USA 103: 9274-9279.
- van Hylckama Vlieg JE, Rademaker JL, Bachmann H, Molenaar D, Kelly WJ, Siezen RJ. 2006. Natural diversity and adaptive responses of *Lactococcus lactis* Curr. Opin. Biotechnol. 17: 183-190.
- van Pijkeren JP, Britton RA. 2014. Precision genome engineering in lactic acid bacteria. Microb. Cell Fact. 13: S10.

- Ventura M, O'Flaherty S, Claesson MJ, Turrone F, Klaenhammer TR, van Sinderen D, O'Toole PW. 2009. Genome-scale analyses of health-promoting bacteria: probiogenomics. *Nat. Rev. Microbiol.* 7: 61-71.
- Vido K, Le Bars D, Mistou MY, Anglade P, Gruss A, Gaudu P. 2004. Proteome analyses of heme-dependent respiration in *Lactococcus lactis*: involvement of the proteolytic system. *J. Bacteriol.* 186: 1648-1657.
- von Ossowski I, Reunanen J, Satokari R, Vesterlund S, Kankainen M, Huhtinen H, Tynkkynen S, Salminen S, de Vos WM, Palva A. 2010. Mucosal adhesion properties of the probiotic *Lactobacillus rhamnosus* GG SpaCBA and SpaFED pilin subunits. *Appl. Environ. Microbiol.* 76: 2049-2057.
- Watanabe M, van der Veen S, Nakajima H, Abee T. 2012. Effect of respiration and manganese on oxidative stress resistance of *Lactobacillus plantarum* WCFS1. *Microbiol.* 158: 293-300.
- Wei Y-X, Zhang Z-Y, Liu C, Malakar PK, Guo X-K. 2012. Safety assessment of *Bifidobacterium longum* JDM301 based on complete genome sequences. *World J. Gastroenterol.* 18: 479-488.
- Wells JM, Mercenier A. 2003. Lactic acid bacteria as mucosal delivery system. In, *Genetics of Lactic Acid Bacteria*. B.J.B. Wood and P.J. Warner (eds.), pp. 261-290. Kluwer Academic/Plenum Publishers, New York. ISBN 978-0-306-47290-9.
- Whiley RA, Kilian M. 2003. International committee on systematics of prokaryotes subcommittee on the taxonomy of staphylococci and streptococci: minutes of the closed meeting, 31 July 2002, paris France. *Int. J. Syst. Evol. Microbiol.* 53: 915-917.
- Yilmaz P, Kottmann R, Field D, Knight R, Cole JR, Amaral-Zettler L, Gilbert JA, Karsch-Mizrachi I, Johnston A, Cochrane G, Vaughan R, Hunter C, Park J, Morrison N, Rocca-Serra P, Sterk P, Arumugam M, Bailey M, Baumgartner L, Birren BW, Blaser MJ, Bonazzi V, Booth T, Bork P, Bushman FD, Buttigieg PL, Chain PS, Charlson E, Costello EK, Huot-Creasy H, Dawyndt P, DeSantis T, Fierer N, Fuhrman JA, Gallery RE, Gevers D, Gibbs RA, San Gil I, Gonzalez A, Gordon JI, Guralnick R, Hankeln W, Highlander S, Hugenholtz P, Jansson J, Kau AL, Kelley

ST, Kennedy J, Knights D, Koren O, Kuczynski J, Kyrpides N, Larsen R, Lauber CL, Legg T, Ley RE, Lozupone CA, Ludwig W, Lyons D, Maguire E, Methé BA, Meyer F, Muegge B, Nakielny S, Nelson KE, Nemergut D, Neufeld JD, Newbold LK, Oliver AE, Pace NR, Palanisamy G, Peplies J, Petrosino J, Proctor L, Priesse E, Quast C, Raes J, Ratnasingham S, Ravel J, Relman DA, Assunta-Sansone S, Schloss PD, Schriml L, Sinha R, Smith MI, Sodergren E, Spo A, Stombaugh J, Tiedje JM, Ward DV, Weinstock GM, Wendel D, White O, Whiteley A, Wilke A, Wortman JR, Yatsunenko T, Glöckner FO. 2011. Minimum information about a marker gene sequence (MIMARKS) and minimum information about any (x) sequence (MIxS) specifications. Nat. Biotechnol. 29: 415-420.

Zhang G, Gao B, Adeolu M, Khadka B, Gupta RS. 2016. Phylogenomic analyses and comparative studies on genomes of the *Bifidobacteriales*: identification of molecular signatures specific for the order *Bifidobacteriales* and its different subclades. Front. Microbiol. 7: 978.

Zhang Z-Y, Liu C, Zhu Y-Z, Wei Y-X, Tian F, Zhao G-P, and Guo X-K. 2012. Safety assessment of *Lactobacillus plantarum* JDM1 based on the complete genome. Int. J. Food Microbiol. 153: 166-170.

Zotta T, Ricciardi A, Ianniello RG, Parente E, Reale A, Rossi F, Lucilla I, Comi G, Coppola R. 2014. Assessment of aerobic and respiratory growth in the *Lactobacillus casei* group. PLoS One 9: e99189.

Figure 1.- Compass rose of different levels and aspects of knowledge provided by genome sequence and data analysis.