Food Science and Biotechnology (2023) 32:389–412
https://doi.org/10.1007/s10068-022-01142-8

INVITED REVIEW

Bioinformatics and its role in the study of the evolution and probiotic potential of lactic acid bacteria

Remilyn M. Mendoza[1] · Sang Hoon Kim[1] · Robie Vasquez[1] · In-Chan Hwang[1] · Young-Seo Park[2] · Hyun-Dong Paik[3] · Gi-Seong Moon[4] · Dae-Kyung Kang[1]

Received: 14 March 2022 / Revised: 30 June 2022 / Accepted: 13 July 2022 / Published online: 10 August 2022

Abstract

Due to their numerous well-established applications in the food industry, there have been many studies regarding the adaptation and evolution of lactic acid bacteria (LAB) in a wide variety of hosts and environments. Progress in sequencing technology and continual decreases in its costs have led to the availability of LAB genome sequence data. Bioinformatics has been central to the extraction of valuable information from these raw genome sequence data. This paper presents the roles of bioinformatics tools and databases in understanding the adaptation and evolution of LAB, as well as the bioinformatics methods used in the initial screening of LAB for probiotic potential. Moreover, the advantages, challenges, and limitations of employing bioinformatics for these purposes are discussed.

Keywords Bioinformatics · Lactic acid bacteria · Comparative genomics · Adaptation and evolution · Probiotic screening

Introduction

Lactic acid bacteria (LAB) are a group of beneficial bacteria that are found in diverse environments, including the gastrointestinal tracts (GITs) of humans (Wang et al., 2020; Zhou et al., 2020) and other animals (Lee et al., 2017; Son et al., 2020), the urogenital tract (Pan et al., 2020; van der Veer et al., 2019), and in dairy (Koryszewska-Bagińska et al., 2019; Quilodrán-Vega et al., 2020; Schmid et al., 2018) and fermented foods (Eisenbach et al., 2018; Eisenbach et al., 2019). LAB are generally gram-positive, catalase-negative, non-sporulating bacteria that can be homofermentative, facultative heterofermentative, or obligate heterofermentative (O'Sullivan et al., 2009). The known genera of LAB include *Lactobacillus*, *Lactococcus*, *Leuconostoc*, *Pediococcus*, and a few species of *Streptococcus* (Goel et al., 2020). LAB have "generally regarded as safe" (GRAS) status and have been defined as probiotics by the World Health Organization, as "live microorganisms that when administered in adequate amounts give health benefits to the host" (Tarrah et al., 2020). With the numerous applications of LAB, this has resulted in a great deal of interest in the research community.

✉ Dae-Kyung Kang
dkkang@dankook.ac.kr

Remilyn M. Mendoza
remilyn.mendoza@gmail.com

Sang Hoon Kim
giooman@naver.com

Robie Vasquez
robie.vasquez1@gmail.com

In-Chan Hwang
hec777@naver.com

Young-Seo Park
ypark@gachon.ac.kr

Hyun-Dong Paik
hdpaik@konkuk.ac.kr

Gi-Seong Moon
gsmoon@ut.ac.kr

[1] Department of Animal Resources Science, Dankook University, 119 Dandae-ro, Cheonan 31116, Republic of Korea

[2] Department of Food Science and Biotechnology, Gachon University, Seongnam 13120, Republic of Korea

[3] Department of Food Science and Biotechnology of Animal Resource, Konkuk University, Seoul 05029, Republic of Korea

[4] Division of Food Science and Biotechnology, Korea National University of Transportation, Jeungpyeong 27909, Republic of Korea

Genome sequencing has revolutionized the ways in which the biology, physiology, ecology, evolution, and applications of organisms are studied. The first prokaryotic genome was sequenced in 1995, with the 1000th prokaryotic genome completed 15 years later in 2009 (Lagesen, 2010). The first complete genome of a *Lactobacillus* species was reported in 2003 when Kleerebezem et al. published the complete genome sequence of *Lactobacillus plantarum* (present name, *Lactiplantibacillus plantarum* subsp. *plantarum*) WCFS1 (Kleerebezem et al., 2003). At the time of writing, about 95,511 genomes of organisms belonging to order Lactobacillales alone are present in the National Center for Biotechnology Information (NCBI) database, of which 12,259 (1639 complete) genomes belong to the Lactobacillaceae (https://www.ncbi.nlm.nih.gov/labs/data-hub/taxonomy/186826/). The continued innovation in sequencing technologies, with increased computational speed and storage, and continued improvement of software, as well as the decreasing costs of sequencing, have contributed to this rapid growth of genome sequence data (Bansal, 2005; Gauthier et al., 2018; Lagesen, 2010). However, the expansion of these massive amounts of genomic data is accompanied by the challenge of interpretation to derive meaningful biological information.

Bioinformatics has been the major driving force behind deriving such information from genome sequence data. Initially used in the field of biochemistry to determine the structures of proteins following Edman degradation peptide sequencing, bioinformatics has become a key player in DNA analysis (Gauthier et al., 2018). From the basic processing of raw sequencing reads to more complex analyses, such as comparative genomics, metagenomics, evolution studies, and structural and metabolic modeling, bioinformatics tools have been widely used and developed in this era of "omics", "big data", and modern biology (Siezen et al., 2004). Figure 1 shows the general methodology from DNA extraction to downstream bioinformatics analyses that is commonly undertaken for comparative genomics studies of LAB. Depending on the goals of a study, many different sequencing technologies and bioinformatics programs may be applied (Carriço et al., 2018; Dominguez et al., 2018). This mini-review presents some of the bioinformatics tools and databases commonly used in genomic studies of LAB.

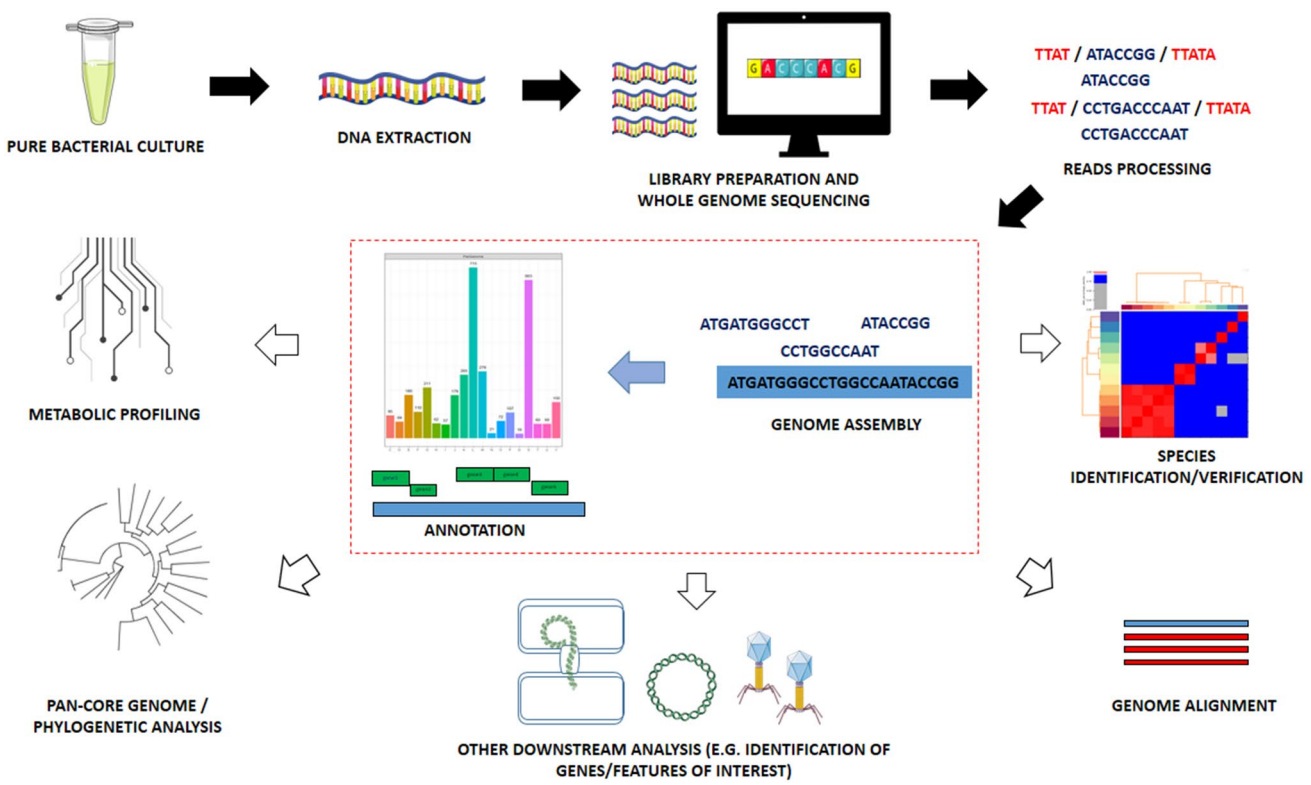

Fig. 1 The general workflow for the comparative genomics of LAB from DNA isolation from pure culture to the downstream analyses of the genome after assembly and annotation. Filled arrows designate the common order of process while the unfilled arrows designate the analyses that can be done after genome assembly or annotation without specific order. Genome assembly and annotation in the dashed box as central processes as most of the downstream analyses require either genome assembly, annotation outputs (e.g. protein fasta file, genbank files), or both as input data to be processed depending on the tool and the purpose of analysis. For detailed guidance on genome assembly, annotation, and microbial bioinformatics readers are referred to the papers of Carriço et al. (2018) and Dominguez et al. (2018)

Moreover, the recent information gathered on LAB, particularly the former genus *Lactobacillus*—which is presently divided into 25 genera (Zheng et al., 2020)—where bioinformatics analyses have been mainly applied, are presented. We also present an overview of the roles of bioinformatics in evolutionary studies and in the screening of the probiotic potential of LAB.

Bioinformatics in the study of the adaptation and evolution of LAB

There have been a great many studies regarding the ubiquity of LAB due to their ability to invade and inhabit highly diverse and complex environments. With their phylogenetic closeness, the ability of LAB to occupy a diverse set of ecological niches suggests considerable genetic adaptation during their evolution (O'Sullivan et al., 2009). Comparative genomics and phylogenomics have been the central approaches used in the study of the adaptation and evolution of LAB. The bioinformatics tools and databases that have been commonly used for comparative genomics and phylogenomics studies of LAB are presented in Table 1.

Pan-core genome analysis

Common themes in the comparative genomics of LAB involve the identification of the pangenome, core genome, accessory genes, and unique genes of LAB species (Eisenbach et al., 2019; Schmid et al., 2018; Wang et al., 2020). The pangenome is the whole set of genes of a species (Zhang et al., 2019a); the core genome is the set of genes present in all strains and is usually composed of the housekeeping genes or those that are needed for basic survival (Zhang et al., 2019a; Zhou et al., 2020); accessory genes are those shared by two or more species in the sample (Zhang et al., 2020); and the unique or specific genes, sometimes called the variable genome (Zhou et al., 2020) or the dispensable genome (Tettelin et al., 2008), are those genes present only in specific strains (Zhang et al., 2020; Zhou et al., 2020). The accessory genes and the dispensable genome contribute to species diversity and provide functional features that confer selective advantages for survival in a specific niche, such as antibiotic resistance genes and genes that may contribute to adaptation and colonization (Tettelin et al., 2008; Zheng et al., 2020). Comparative genomics has been used to gain insights into the genomic characteristics of single- or multi-niche LAB (Evanovich et al., 2019; Feyereisen et al., 2019; Jia et al., 2020; Koryszewska-Bagińska et al., 2019; Lee et al., 2017; Mao et al., 2021; Martino et al., 2016; O'Sullivan et al., 2009; Petrova et al., 2018; Senan et al., 2014; Son et al., 2020; Verce et al., 2020; Xing et al., 2017; Zheng et al., 2015).

ERGO™ is one of the earliest tools used for comparative genomics analysis of LAB (Siezen et al., 2004). ERGO, designed by Integrated Genomics Inc. and only accessible through subscription, is a genome analysis and discovery suite that is aimed at providing the tools necessary for comparative analysis of genomes and the integration of sophisticated metabolic and cellular reconstructions (Overbeek et al., 2003). More recent tools for comparative genomics, however, have been accessible for free and were designed for more rapid and high-throughput processing of larger datasets. These tools include the Efficient Database framework for comparative Genome Analyses using BLAST score Ratios (EDGAR) (Blom et al., 2009), Pan-genome analysis pipeline, (PGAP) (Zhao et al., 2012), and Roary (Page et al., 2015), which have been used extensively in recent pangenome studies of LAB. Among the pangenome analysis tools mentioned, EDGAR is available online as a web server and access can be requested from the developer team if working with unpublished genomes. Visualizations such as Venn diagrams, synteny plots, and phylogenetic trees can also be generated from EDGAR (3.0) (Dieckmann et al., 2021). Meanwhile, PGAP (and PGAP-X (Zhao et al., 2018)) and Roary are standalone tools that can be run on a local computer with the Linux environment (PGAP and Roary), and in windows and MacOS platforms (PGAP-X). PGAP-X can provide a pangenome profile curve, while Roary can also provide pangenome profile curve when run with R (i.e. specified during parameter inputs) and its outputs can be viewed in Phandango (Hadfield et al., 2017).

Annotation

As early as 1997, comparisons of complete genomes were becoming indispensable to the understanding of different biological phenomena and, with the expected exponential increase in the number of sequenced genomes that would become available, Tatusov et al. developed the Clusters of orthologous groups (COG) system (Tatusov et al., 1997). This system, which originally consisted of 720 COGs, allows the automatic functional and phylogenetic annotation of genes and gene sets. As orthologs are defined as genes in different species that evolved from a common ancestral gene and retain the same function (Fitch, 1970), this allows the transfer of functional information from one member to an entire COG. The COG system, therefore, encompasses a framework for both functional and evolutionary genome analyses (Tatusov et al., 1997). A number of tools for functional annotations of the COGs of proteins have since been developed and used in genomics studies. Among these tools, the NCBI COG database (Galperin

Table 1 Bioinformatics tools and databases commonly used for comparative genomics and phylogenomics of LAB

Tools/database	Website/documentation	Authors
Genome alignment		
BLAST	https://blast.ncbi.nlm.nih.gov/Blast.cgi	Altschul et al. (1990)
Artemis Comparison Tool (ACT)	https://github.com/sanger-pathogens/Artemis	Rutherford et al. (2000)
Mauve	http://darlinglab.org/mauve/user-guide/progressivemauve.html	Darling et al. (2004)
Muscle	http://www.drive5.com/muscle/	Edgar (2004)
MAFFT	https://mafft.cbrc.jp/alignment/software/	Katoh and Standley (2013)
Pangenome analysis		
ERGO	http://ergo.integratedgenomics.com/ERGO/	Overbeek et al. (2003)
EDGAR	http://edgar.computational.bio/	Blom et al. (2009)
PGAP	https://sourceforge.net/projects/pgap/	Zhao et al. (2012)
Roary	https://sanger-pathogens.github.io/Roary/	Page et al. (2015)
Annotation		
RAST	https://rast.nmpdr.org/	Aziz et al. (2008); Overbeek et al., (2013); Brettin et al. (2015)
Prokka	https://github.com/tseemann/prokka	Seemann (2014)
NCBI PGAP	https://github.com/ncbi/pgap	Tatusova et al. (2016)
COG-database	http://www.ncbi.nlm.nih.gov/COG/	Tatusov et al. (1997); Galperin et al. (2015)
Ortho-MCL database	https://orthomcl.org/orthomcl/app	Li et al. (2003); Chen et al. (2006)
EggNog-Mapper	http://eggnog-mapper.embl.de/	Huerta-Cepas et al. (2019); Cantalapiedra et al. (2021)
Proteins/Carbohydrate metabolism/Metabolic pathways		
KEGG	https://www.kegg.jp/	Ogata et al. (1999)
TIGRFAMs	https://www.jcvi.org/research/tigrfams	Haft et al. (2003)
SignalP	https://services.healthtech.dtu.dk/service.php?SignalP-5.0	Almagro Armenteros et al.(2019)
TMHMM	https://services.healthtech.dtu.dk/service.php?TMHMM-2.0	Krogh et al. (2001)
LipoP	https://services.healthtech.dtu.dk/service.php?LipoP-1.0	Juncker et al. (2003)
Pfam	http://pfam.xfam.org/	Mistry et al. (2021)
Uniprot/SwissProt	https://www.uniprot.org/	Bateman et al. (2021)
CAZy	http://www.cazy.org/	Cantarel et al. (2009)
Phylogenetic analysis		
MEGA	https://www.megasoftware.net/	Hall (2013)
RA×ML	https://github.com/stamatak/standard-RAxML	Stamatakis (2014)

et al., 2015), Ortho-MCL (i.e. applying Markov Cluster algorithm) (Li et al., 2003) and Ortho-MCL database (Chen et al., 2006), and eggNOG-mapper (Cantalapiedra et al., 2021; Huerta-Cepas et al., 2019) have been used in studies of LAB. For more information about the different orthology resources, readers are directed to the paper of de Boissier and Habermann (2016) presenting a practical guide on the different orthology resources. Meanwhile, the rapid annotation of microbial genomes using subsystems technology (RAST) (Aziz et al., 2008; Brettin et al., 2015; Overbeek et al., 2013), Prokka (Seemann, 2014), and the NCBI PGAP (Tatusova et al., 2016), have also been widely used in genome annotation of LAB. RAST which functionally annotates genomes based on subsystems—sets of functional roles that together implement a specific biological complex or structural complex, is implemented and freely accessible on a web server (Table 1). Genomes assemblies are uploaded one at a time and annotated, and the identified subsystems are displayed in a pie chart. Prokka, on the other hand, is a standalone tool and can be run on a local computer. Genomes can be annotated in batches, and outputs of Prokka can be used as inputs for Roary for pangenome analysis, and Artemis (Rutherford et al., 2000) for genome visualization. However familiarity with the Linux environment and script (e.g. shell script) writing (for batch annotation) is needed. NCBI PGAP

can be accessed both online and standalone and has been instrumental in the identification of pseudogenes.

Metabolism

The determination of metabolic capabilities of LAB that allow them to inhabit a wide variety of habitats is of central importance in understanding their adaptation and evolution. There have been many studies using these tools and databases to identify genes and proteins associated with metabolic activities as well as metabolic pathway construction in LAB. Uniprot is a database housing a huge collection protein sequence (Bateman et al., 2021) while Pfam (Mistry et al., 2021) is a tool that annotates protein families and domains and uses the pfamseq as well as Uniprot databases. TIGRFAMs (Haft et al., 2013) is another tool that annotates proteins families and was designed to complement Pfam. TIGRFAMs usually avoid constructing model that duplicates what was previously covered by Pfam but constructs model for domains and repeats of a novel region. SignalP is specific for identification of signal peptides which in prokaryotes target proteins for translocation across inner plasma membrane (Almagro Armenteros et al., 2019). These tools have been used extensively for protein identification and functional annotation of LAB strains.

The Carbohydrate-Active EnZymes (CAZy) database (CAZyDB) (Cantarel et al., 2009) is a database specialized for enzymes that build and breakdown complex carbohydrates and glycoconjugates. Genomes can be annotated by CAZyDB however an automated system is lacking and researchers need to contact the CAZy development team to have their genomes of interest be annotated by the CAZy team (Yin et al., 2012). This limitation, among others, was addressed by dbCAN (Yin et al., 2012), a web server based on the classification scheme of CAZyDB which provided an automated annotation of genomes for CAZy. CAZyDB and dbCAN are frequently used for the identification of enzymes involved in carbohydrate metabolism of LAB. Meanwhile, the Kyoto encyclopedia of genes and genomes (KEGG) (Ogata et al., 1999) which has been used for metabolic pathway construction, from its inception in 1999 has upgraded and became a comprehensive and integrated database consisting of different databases serving four main purposes: for system information, genomic information, chemical information, and health information (Kanehisa et al., 2017). Each category consists of different databases that collect and provide the necessary data and information according to the purpose. A detailed description of each category and the different databases of KEGG are in the published paper by Kanehisa et al. (2017).

Adaptation and evolution of LAB

Among the major genera of LAB, the genome characteristics of members of the genus *Lactobacillus* vary considerably. The genomes of *Lactobacillus* species range in size from 1.23 Mb (*Lactobacillus sanfranciscensis*; present name, *Fructilactobacillus sanfranciscensis*) to 4.91 Mb (*Lactobacillus parakefiri*; present name, *Lentilactobacillus parakefiri*), with guanine-cytosine (GC) contents ranging from 31.93% to 52.07% (Sun et al., 2015b). It was reported that although *Lactobacillus* is traditionally defined as a genus, its genetic diversity is larger than that of a typical family (Sun et al., 2015b). Very recently, this matter was addressed when the reclassification of the genus *Lactobacillus* into 25 genera of which 23 are novel, was proposed (Zheng et al., 2020). Comparative genomics and phylogenomics allowed *Lactobacillus* to be sorted into three lifestyle categories: free-living, those encompassing plant and environmental isolates; host-adapted, those associated with vertebrate or invertebrate hosts (Duar et al., 2017); and nomadic, those that effectively migrate to different habitats and are characterized by having dynamic and flexible lifestyles (Martino et al., 2016). As *Lactobacillus* occupied various ecological niches, the hosts and environments exerted pressures that gradually modify the genome to the point that some eventually became specialized to such habitats (Martino et al., 2016). Niche specialization has been observed and documented in several host-adapted *Lactobacillus* species (Frese et al., 2011; Macklaim et al., 2011; Oh et al., 2010; van de Guchte et al., 2006). Specialization to a particular environment has driven the loss and gain of functional genes, as indicated by the presence of pseudogenes and genes acquired via horizontal gene transfer (HGT), respectively, and genome reduction (Douillard and de Vos, 2014; Duar et al., 2017; van de Guchte et al., 2006). Meanwhile, species with a nomadic lifestyle seemed to evolve to become generalists allowing them to migrate and occupy different niches (Martino et al., 2016). Such nomadic lifestyles are also evident in the characteristics of their genomes in that they have larger genomes than host-adapted and free-living lactobacilli, allowing a larger complement of functional genes, thus providing adaptive advantages in various habitats (Duar et al., 2017; Martino et al., 2016). Bioinformatics tools, especially those used in comparative genomics that allow the identification of niche-specific or specialized genes, as well as tools for phylogenomics, have been instrumental to our understanding of the lifestyle adaptation of *Lactobacillus* species (Duar et al., 2017; Martino et al., 2016; Siezen et al., 2004). Here, the information from comparative genomics and phylogenomics of LAB, mainly that of *Lactobacillus* species, is presented and the role of bioinformatics in understanding the adaption and evolution of LAB is discussed.

Niche-specific adaptation and specialization

Food

Lactobacillus delbrueckii subsp. *bulgaricus* (*L. bulgaricus*) is one of the best examples of a *Lactobacillus* species that has undergone niche specialization. van de Guchte et al. reported the complete genome of *L. bulgaricus* strain ATCC11842, isolated from Bulgarian yogurt, and revealed evidence that this species has undergone extensive reductive evolution driven by specialization (van de Guchte et al., 2006). The *L. bulgaricus* genome was shown to have a significantly higher GC content (~ 49.7%) than those of related species, such as *Lactobacillus acidophilus* and *Lactobacillus johnsonii*, both of which have GC content of ~ 34%. This deviation was mainly associated to the difference in GC in codon position 3 (GC3): 65.0% GC in *L. bulgaricus* while 25.0% and 24.4% in *L. acidophilus* and *L. johnsonii*, respectively (van de Guchte et al., 2006). Correlation analysis showed that the GC3 values of *L. acidophilus* and *L. johnsonii* strongly fit the correlation between GC3 and overall GC content in bacteria while the GC3 value of *L. bulgaricus* (65%) strongly deviates from the expected (54%). As the evolution in the third codon is generally faster than in the first and second codons, it is hypothesized that the genome of *L. bulgaricus* may be in active phase of evolution toward a higher GC content or it may be that the correlation between GC3 and overall GC content has been altered or damaged in *L. bulgaricus* (van de Guchte et al., 2006). Next, a 47.5-kb inverted repeat conserved in *L. bulgaricus* strains was hypothesized to represent a transient stage in the genome evolution of this species. Furthermore, the large number of pseudogenes suggests that the genome is in an active state of gene elimination and reduction, supported by the large numbers of rRNA and tRNA genes, which are 20–30% higher than the highest values observed to date for its genome size of ~ 1.8 Mb, indicating a recent phase of size reduction. The smaller number of genes encoding transcriptional regulators and the complete absence of a large number of enzymes for the biosynthesis of amino acids in *L. bulgaricus* reflect adaptation to the stable and nutritionally rich milk environment.

O'Sullivan et al. (2009) hypothesized that a group of genes could be determined and define the niche of an organism. They initially studied *Lactobacillus helveticus* DPC4571, a dairy LAB isolate, and *L. acidophilus* NCFM, a gut LAB isolate. While *L. helveticus* and *L. acidophilus* occupy two markedly different niches, genome analysis showed that they have high sequence similarity, even greater than the degrees of sequence similarity between *L. acidophilus* and other gut and multi-niche LAB. Expanding the analysis to 11 dairy, gut, and multi-niche LAB samples, the authors identified nine niche-specific genes forming a "barcode": six dairy-specific genes with functions related to the proteolytic system (i.e. two carboxypeptidases) and restriction/modification system (i.e. three type I and one type III restriction modification enzymes); and three gut-specific genes related to bile salt hydrolysis (two bile salt hydrolases) and sugar metabolism (i.e. maltose-6-phosphate glucosidase). As *L. helveticus* DPC4571 diverged from *L. acidophilus* NCFM, it lost some sugar metabolism capacity as well the bile salt hydrolase function, as evidenced by the presence of the bile salt hydrolase (*bsh*) gene in *L. helveticus* DPC4571 but with a frameshift mutation that likely renders it dysfunctional.

Lactobacillus sakei (present name, *Latilactobacillus sakei* subsp. *sakei*), a LAB that is found naturally in fresh meat and fish, has been widely used as a starter culture for sausage fermentation (Chaillou et al., 2005; Eisenbach et al., 2019; Nyquist et al., 2011). The complete genome of *L. sakei* 23K, which was originally isolated from a French sausage, was sequenced in 2005. Genome analysis of this strain identified genetic features that allow it to compete successfully on raw meat products (Chaillou et al., 2005). Chaillou et al. (2005) identified pertinent genes in the genome of *L. sakei* predicted to have roles in addressing the challenges faced by microorganisms in the meat environment, such as high protein and low carbohydrate supply (i.e. genes for purine salvage pathway, methylglyoxal synthase, and arginine deaminase pathway), and changing redox and oxygen levels (i.e. NADH oxidase and manganese-dependent superoxide dismutase), high salt concentration (i.e. Na^+/H^+ antiporters and ATP-dependent Na^+ efflux pump), and low temperatures during processing (i.e. putative cold shock proteins), in addition to competition among microorganisms for colonization and dominance (i.e. genes for biofilm formation, co-aggregation, and putative bacteriocin resistance). These genes for adaptation in the meat environment were found to be conserved in *L. sakei* strains isolated from meat, as shown in the study of Nyquist et al. where 18 strains of *L. sakei* mainly isolated from processed meat were compared with meat-borne *L. sakei* 23K using microarray comparative genome hybridization (CGH) (Nyquist et al., 2011).

Lactobacillus brevis (present name, *Levilactobacillus brevis*) has been identified as one of the common causes of beer spoilage (Feyereisen et al., 2019). Genome analyses showed that the genomes of beer-associated strains of *L. brevis* are larger than those of non-beer strains and possess larger numbers of coding sequences (CDSs) in their genomes. The unique genes in the *L. brevis* beer strains studied by Feyereisen et al. (2019) have been shown to encode proteins related to oxido-reduction reactions, transcription, membrane, and cell surface and membrane transport. These

genes were predicted to aid the organism in adapting to the beer environment with relatively low pH and iso-α-acids from hop compounds. Some of the genes identified in this study, such as genes predicted to encode an oligogalacturonide transporter, a short-chain dehydrogenase, and an RNA polymerase sigma factor ECF subfamily (Feyereisen et al., 2019), were also identified previously as beer spoilage diagnostic marker genes (Bergsveinson and Ziola, 2017).

Gastrointestinal tract (GIT)

The presence of the *bsh* gene has been linked with adaptation to the gut environment (O'Sullivan et al., 2009). In addition to the *bsh* gene, other genes that may contribute to the survival of *Lactobacillus kefiranofaciens* ZW3 in the gut environment include genes involved in acid tolerance, such as H^+-ATPase, ornithine decarboxylase, and amino acid permease (Xing et al., 2017). In *L. helveticus* MTCC 5463, Senan et al. (2014) identified a higher copy number of genes contributing to heat, cold, osmotic, and oxidative stress tolerance, which may support survival in the gut. Feces-derived *Lactobacillus crispatus* was also shown to harbor genes encoding the clustered regularly interspaced short palindromic repeats (CRISPR)-Cas system, glycoside hydrolase (GH) family members, and tetracycline/lincomycin resistance, which may be significant in adaptation to the complex intestinal environment (Zhang et al., 2020).

Lactobacillus reuteri (present name, *Limosilactobacillus reuteri*) is a known natural inhabitant (autochthonous) of the GITs of humans (Duar et al., 2017) and selected animals, such as pigs, rodents, and chickens, where it has been identified as one of the dominant species in the GIT (Frese et al., 2011; Oh et al., 2010). *L. reuteri* is also known to have undergone host specialization and has been commonly used as a model organism in studies of the evolutionary strategy of a vertebrate gut symbiont (Frese et al., 2011; Oh et al., 2010). Phylogenetic analysis using amplified-fragment length polymorphism (AFLP) and multi-locus sequence analysis (MLSA) data from *L. reuteri* isolated from six different host species (human, mouse, rat, pig, chicken, and turkey) revealed distinct monophyletic clades reflecting host origin (Oh et al., 2010). Comparative genomics of *L. reuteri* 100–23 (rodent strain) and *L. reuteri* F25 (human strain) revealed host-specific genes in their genomes (Frese et al., 2011). An auxiliary protein secretion system (SecA2 cluster), which is thought to have been acquired by *L. reuteri* via HGT and is hypothesized to play a significant role in colonization, and a urease gene cluster that is believed to contribute to the survival of *L. reuteri* in the forestomach of rodents were unique to the 100–23 strain, while strain F275 harbored the *pdu-cbi-cob-hem* cluster, which is absent in 100–23. The *pdu-cbi-cob-hem* cluster, which provides cobalamin (vitamin B_{12}) biosynthesis, glycerol utilization, propanediol fermentation, and reuterin production, is believed to contribute to colonization, survival, and fitness in the human gut (Frese et al., 2011). Comparative genomics is widely employed in studies to understand *L. reuteri* strains isolated from different hosts (Greppi et al., 2020; Son et al., 2020; Xu et al., 2020).

Vaginal tract

The known colonizers of the vaginal tract include *Lactobacillus crispatus*, *Lactobacillus gasseri*, *Lactobacillus jensenii*, and *Lactobacillus iners* (Petrova et al., 2015). These *Lactobacillus* species have been proposed as biomarkers for vaginal health (Petrova et al., 2015), such that the presence and concentration and dominance of these *Lactobacillus* species, particularly *L. crispatus*, in the vaginal tract are associated with vaginal health or dysbiosis (Petrova et al., 2015; van der Veer et al., 2019). *L. crispatus* and *L. gasseri* have been shown to be associated with healthy vaginal microbiota, while *L. iners* has been isolated from both healthy and dysbiotic vaginal microbiota (Pan et al., 2020; Petrova et al., 2015). van der Veer et al. recently compared *L. crispatus* isolated from healthy vaginal microbiota (*Lactobacillus*-dominated) with *L. crispatus* isolated from dysbiotic vaginal microbiota (non-*Lactobacillus*-dominated) (van der Veer et al., 2019). Their analysis revealed the presence of a 3-fragmented glycosyltransferase (GT) gene that was shown to be more abundant in the genomes of *L. crispatus* isolated from dysbiotic vaginal microbiota, suggesting a fitness advantage under dysbiotic conditions where there is an increased immune response. The first and longest fragment is hypothesized to harbor the active site for GT, while the other two fragments are hypothesized to play roles in a process known as phase variation, which allows bacteria to adapt rapidly and diversify their surface glycans, rendering an evolutionary advantage. *L. crispatus* is hypothesized to exploit this genetic variation to persist at low levels in the dysbiotic vaginal microbiota and escape the immune system (van der Veer et al., 2019).

L. crispatus and *L. gasseri* are also generally isolated and identified from the GIT (Pan et al., 2020). The gut and vaginal environments are very different in terms of epithelial cell types, nutrient availability, pH, and microbiome composition (Pan et al., 2020). Zhang et al. showed that the niche has exerted important impacts in the evolution of *L. crispatus.* They noted that *L. crispatus* samples isolated from the vagina and gut clustered separately in a phylogenetic tree constructed using the core genome of the *L. crispatus* samples. Genes related to acid tolerance, redox reactions, pullulanase, and carbohydrate-binding modules were identified from *L. crispatus* isolated from the vagina and were hypothesized to play roles in adaptation to the acidic environment

of the vagina, nutrient acquisition, and dominance over competing vaginal colonizers (Zhang et al., 2020).

L. iners and *L. jensenii* are two *Lactobacillus* species that have specialized to inhabit the vaginal tract and are found only in this niche (Duar et al., 2017). Both *L. iners* and *L. jensenii* are found in the healthy and dysbiotic vaginal tract (Petrova et al., 2015; Putonti et al., 2020). *L. iners* AB-1 has one of the smallest genomes among the LAB with a genome size of 1.3 Mb (Macklaim et al., 2011). As noted by Macklaim et al., the genome of *L. iners* AB-1 seems to have undergone rapid evolution events that led to large-scale gene loss and acquisition of genes for survival in the vaginal environment through HGT. The genome of *L. iners* AB-1 was shown to contain 95 possible pseudogenes and 65 predicted genes likely to have been acquired from organisms outside its genus (Macklaim et al., 2011). Gene clusters for an iron-sulfur cluster assembly system that may serve as an environmental sensor and regulate a response, several unique sigma factors that may regulate gene transcription, and cholesterol-binding lysin that may contribute to host cell adhesion or play a role in defense against other microbes in the genome of *L. iners* AB-1 were suggested to be important to its adaptation and survival in the vaginal tract. *L. iners* shows the highest degree of specialization among *Lactobacillus* species and has been suggested to have evolved to almost an obligate symbiotic lifestyle (Duar et al., 2017).

Nomadic lifestyle

The adaptation of most *Lactobacillus* species to a particular niche has been linked to large changes in the genome content reflecting the constraints imposed by the habitat (Martino et al., 2016). Such adaptation eventually leads to specialization where genes that are vital for survival or offer a competitive advantage in a specific habitat are retained, while other genes are lost. Apparent genome reduction has been observed such that some strains of *Lactobacillus* that are evolving toward specialization to a specific environment or host, such as in the case of *L. bulgaricus* in dairy, *L. reuteri* in the GIT, and *L. iners* in the vagina, have reduced genome sizes compared to other members of the genus. Adaptation and eventual specialization of *Lactobacillus* species are trends observed in LAB, supported by the results of comparative genomics and phylogenomics studies (Duar et al., 2017; Frese et al., 2011; Martino et al., 2016; Oh et al., 2010). However, particular *Lactobacillus* species seem to deviate from this norm, such as those exhibiting the nomadic lifestyle (Martino et al., 2016). Comparative genomics and phylogenomics studies of specialized *Lactobacillus* species, such as *L. reuteri*, have shown clustering of strains reflecting their host or origin of isolation (Oh et al., 2010). The same clustering according to the isolation source, however, was not observed in *Lactobacillus plantarum* (present name, *Lactiplantibacillus plantarum* subsp. *plantarum*), which has been used as a model organism for studying the nomadic lifestyle (Duar et al., 2017; Martino et al., 2016).

Martino et al. performed comparative genomics analysis of 54 strains of *L. plantarum* isolated from different habitats and examined different metrics, such as gene content, phylogeny, and the variome (variable regions), to explore potential links between intraspecies genetic variability and the potential origin of *L. plantarum* (Martino et al., 2016). The different analyses that Martino et al. performed, however, did not identify genomic signatures reflecting environmental adaptation. They concluded that the genomic adaptation of *L. plantarum* may have been driven by alternative selective pressures other than specific environmental adaptation. Their study also suggested that *L. plantarum* has potential capability to migrate frequently across different environments. Duar et al. (2017) also performed phylogenetic analysis based on the same 54 strains of *L. plantarum*, which also showed no obvious clustering by origin.

Specific evolutionary signatures related to adaptation, as well as distinct clustering reflecting their origin, were also not observed in *Lactobacillus rhamnosus* (present name, *Lacticaseibacillus rhamnosus*), *Lactobacillus casei* (present name, *Lacticaseibacillus casei*), and *Lactobacillus paracasei* (present name, *Lacticaseibacillus paracasei* subsp. *paracasei*). It was noted, however, that although *L. plantarum*, *L. rhamnosus*, *L. casei*, and *L. paracasei* do not form stable populations in hosts, they possess genomic features that allow them to persist and survive in host-associated niches (Duar et al., 2017; Martino et al., 2016).

Bioinformatics in initial screening of potential probiotic LAB

The great deal of attention given to LAB is undeniably due to their application in the food industry, as starter cultures, and as probiotics. Many producers that market LAB as probiotics have used genomic analysis to characterize the strains as, in addition to rapid strain characterization with the available genome-mining tools and databases, genomics is also instrumental in strain mining and prompt selection of specific properties (Douillard and de Vos, 2014). As probiotic features and associated health properties are strain-specific and cannot be generalized for the entire genus (Douillard et al., 2013), genome characterization using bioinformatics has played a vital role in initial screening for potential probiotic LAB strains. This screening for probiotic potential encompasses the identification of genes that may contribute to the fitness and survival of a strain in the highly challenging and dynamic GIT, actual benefit that the strain may provide to the host, and quality and safety of the LAB strain.

The common bioinformatics tools and databases that have been instrumental in the initial screening of the probiotic potential of LAB strains are presented in Table 2. A summary of the genetic features that have been shown to contribute to the probiotic properties of the LAB strains presented below is presented in Table 3.

Screening for genes involved in survival in the gastrointestinal environment

Most studies on probiotic LAB initially evaluated the probiotic potential of a LAB strain with the following criteria on survival and colonization in the GIT (Bagon et al., 2021; Douillard et al., 2013; Quilodrán-Vega et al., 2020; Valeriano et al., 2019; Xing et al., 2017): ability to resist low pH and high bile salt concentration; have antimicrobial activity or the ability to inhibit the growth of pathogens; and ability to adhere to epithelial cells and colonize the gut.

Genes for bile and low pH resistance

The ability of a potential probiotic LAB strain to overcome and survive the harsh conditions during transit in the GIT, in addition to the very low pH in the stomach and the high concentration of bile salts released in the intestine (Lebeer et al., 2008), is a prime consideration in the selection of a probiotic LAB (Arnold et al., 2018; Bagon et al., 2021; Valeriano et al., 2019). Bioinformatics tools for annotation, such as RAST (Aziz et al., 2008; Brettin et al., 2015; Overbeek et al., 2013), Prokka (Seemann, 2014), and NCBI-PGAP (Tatusova et al., 2016), have been very useful in initial screening for genes that may be involved in stress resistance. O'Sullivan et al. identified the *bsh* gene (also designated as choloylglycine hydrolase) as one of the gut-specific genes included in the nine barcode genes that may be used for identification of the niches of certain LAB (O'Sullivan et al., 2009). Whole-genome analysis was instrumental in the initial identification of *bsh* genes in several studies characterizing LAB strains with probiotic potential (Quilodrán-Vega et al., 2020; Xu et al., 2020; Yoo et al., 2017; Zhang et al., 2019a). Meanwhile, several genes for adaptation to acid stress, such as F_0F_1-ATP synthase, Na^+/H^+ antiporter (Senan et al., 2014; Xu et al., 2020), S-adenosylmethionine synthetase, and GTP pyrophosphokinase (Sun et al. 2015a), have also been mapped in the genome using various annotation and alignment (BLAST) tools.

Antimicrobial properties of LAB

LAB are known to inhibit the growth of pathogens through the production of antimicrobials, including organic acids (mainly lactic acid and acetic acid), hydrogen peroxide, and bacteriocin (Abdelhamid et al., 2019; Evanovich et al., 2019; van der Veer et al., 2019). Bacteriocins are ribosomally synthesized antimicrobial peptides produced by bacteria that are effective against closely related (narrow-spectrum) or more distant species (broad-spectrum), to which the producing organism has a specific immune mechanism (Cotter et al., 2013; Walsh et al., 2015). These antimicrobial peptides have attracted a great deal of attention in recent years due to its potential use as an alternative to antibiotics (Cotter et al., 2013) and as a food preservative (Jin Ng et al., 2020; Johnson et al., 2018). Bacteriocins are considered potential alternatives to antibiotics due to their potency, low toxicity, availability of both narrow- and broad-spectrum agents, the

Table 2 Bioinformatics tools and databases frequently used in the screening of genes/features in assessing the probiotic potential of LAB

Gene/feature	Tools/database	Authors
Bacteriocin	BAGEL/BAGEL4	de Jong et al. (2006); Van Heel et al. (2018)
CRISPR-Cas system	CRISPRFinder	Grissa et al. (2007a)
	CRISPRdb	Grissa et al. (2007b)
Restriction modification system	REBASE	Roberts et al. (2009)
AMR genes	ARDB	Liu and Pop (2008)
	ResFinder	Zankari et al. (2012)
	CARD	McArthur et al. (2013)
Virulence factors	VFDB	Chen et al. (2005)
Mobile genetic elements		
Genomic island	IslandViewer	Langille and Brinkman (2009)
Plasmid	Plasmidspades	Antipov et al. (2016)
	PLSBD	Galata et al. (2018)
Phage	PHAST	Zhou et al. (2011)
	PHASTER	Arndt et al. (2016)
Insertion sequence	ISFinder	Siguier et al. (2006)

Table 3 Some of the known functional genes related to survival in the GIT tract and providing benefits to hosts identified in LAB strains

Probiotic property	Gene/gene product	Function	Representative LAB strain	Reference
Survival in GI tract				
Genes for pH and bile resistance	Bile salt hydrolase gene/choloylglycine hydrolase *(bsh)*	Bile resistance	*L. acidophilus* NCFM,	O'Sullivan et al. (2009)
			L. salivarius Ren	Sun et al. (2015a)
	F_0F_1-ATP synthase	Acid resistance	*L. helveticus* DPC 4571	Senan et al. (2014)
	Na + /H + antiporter		*L. helveticus* MTCC 5463	Senan et al. (2014)
			L. reuteri YSJL-12	Xu et al. (2020)
	S-adenosylmethionine synthetase		*L. salivarius* Ren	Sun et al. (2015a)
	GTPpyrophosphokinase		*L. salivarius* Ren	Sun et al. (2015a)
Genes for colonization and adhesion	Pilus-encoding gene cluster (*spaCBA* gene)	Adhesion	*L. rhamnosus* GG,	Kankainen et al. (2009)
			L. paracasei subsp. *paracasei* IBB3423	Koryszewska-Bagińska et al. (2019)
	cwaA gene		*L. plantarum* NL42	Zhang et al. (2015)
Conferring benefits to the hosts				
Production of Vitamins	*cbi-cob-hem* gene cluster	Cobalamin (Vit. 12) production	*L. reuteri* JCM 1112 ^{T}	Morita et al. (2008)
	GTP-cyclohydrolase II (*ribA*), Riboflavin synthase β-subunit(*ribH*), Riboflavin synthase α-subunit (*ribB*), Riboflavin specific deaminase/reductase (*ribG*)	Riboflavin (Vit. B2) Production	*L. fermentum* CECT 5716, *L. plantarum* Lp790, Lp813, and Lp998	Cárdenas et al. (2015) Zago et al. (2017)
Production of vitamins	Dihydroneopterin aldolase (*folb*), Dihydropteridine pyrophosphokinase (*folk*), Dihydropteroate synthase (*folP*), Formylthf-polyglutamate synthase (*folC1*), Folylpolyglutamate synthase (*folC2*), Dihydrofolate reductase (*dfrA*), GTP-cyclohydrolase I (*folE*), Anthranilate synthase (*trpE*), Anthranilate synthase (*trpG*)	Production of Folate (Vit. B9)	*L. fermentum* CECT 5716 *L. plantarum* Lp790, Lp813, and Lp998	Cárdenas et al. (2015) Zago et al. (2017)

Table 3 (continued)

Probiotic property	Gene/gene product	Function	Representative LAB strain	Reference
Immunomodulation		*cps* gene cluster (*cps1A,- cps1G*, and *cps1J*) (*cps2A-cp2K; cps3A -cps3J; cps4A-cps4J*)	*Synthesis of Cell wall polysaccharide moiety	*L. casei* Shirota *L. mucosae* LM1, *L. plantarum* SK151, *L. fermentum* SK152, *L. johnsonii* PF01	Yasuda et al. (2008) Salvador et al. (2021)
		LTA synthase *(ltaS)*, TA transporter subunits *(tagG, tagH)*, TA glycosylation proteins *(gtcA1, gtcA2, gtcA3)*, TA D-alanylation*(dltX)*, TA synthesis proteins *(tagF1,tagF2)*, *LTA exporters (MPE1,MPE2,MPE3, MPE4)*	*Teichoic acid (TA)- related genes	*L. plantarum* PS128 *L. plantarum* J26	Liu et al. (2015) Zhang et al. (2019b)
Immunomodulation		Bactericin immunity protein(*plnI*), Bacteriocins precursor peptide(*plnF, plnE*), Bacteriocins ABC transporter ATP binding protein *(plnG)*, Bacteriocins ABC transporter(*plnH*), Membrane protein(*plnU, plnV, plnW*) (Bile salt hydrolase gene/ choloylglycine hydrolase) *bsh1*	*Interleukin (IL)-secretion genes	*L. plantarum* Lp790, Lp813, and Lp998 *L. mucosae* LM1, *L. plantarum* SK151, *L. fermentum* SK152, *L. johnsonii* PF01	Zago et al. (2017) Salvador et al. (2021)
		COG1028 (short-chain alcohol dehydrogenase family according)-related genes	Synthesis of conjugated linoleicacid which alleviates Irritable Bowel Syndrome	*L. plantarum* CCFM8610	Y. Liu et al. (2021b)
Pathogen inhibition		*Lactobacillus* epithelium adhesion (LEA)	Competitive exclusion of *G. vaginalis*	*L. crispatus*	Ojala et al. (2014)
Pathogen inhibition		Pore-forming Amphipathic Helical Peptide (HP2-20) Family	Exhibits antifungal and antibacterial properties by pore formation	*L. brevis, L. bulgaricus, L. crispatus, L. gasseri, L. reuteri, L. ruminis,* *L. acidophilus,* *L. paracasei,* *L. planatarum,* *L. rhamnosus*	Zafar and Saier (2020)
		Exopolysaccharide gene cluster (EPS) (*epsA-E), wzy* (polymerase), *wzx* (flippase), glycosyltransferases	Reduce viral replication of rotavirus in porcine intestinal epithelial (PIE) cells	*L. debrueckii* TUA4408L	Kanmani et al. (2018)
Anti-tumor/Anti-cancer property		Pyruvate oxidase, NADH oxidase, Alpha-glycerophosphate Oxidase	Production of H_2O_2 which is believed to degrade 4-HAQO which can induce oral cancer	*L. salivarius* Ren	Sun et al. (2015a)
Cholesterol-lowering property		*bsh* gene	Reducing blood cholesterol level	*L. planatrum* EM	Kim et al. (2020)

Table 3 (continued)

Probiotic property	Gene/gene product	Function	Representative LAB strain	Reference
Anti-hyperammonemia	Carbamate kinase	Assimilation of ammonia, prevents hyperammonemia	*L. amylovorus* JBD401	Singh et al. (2018)
	Ornithine carbamoyl transferase			
	N-acyl-L-amino acid amidohydrolase, N-acetyl-L, L-diaminopimelate amino-transferase	For biosynthesis of L-glutamate from L- ornithine		
	Glutamine synthetase type I, Aspartate aminotransferase, N-acyl-L-amino acid ami-dohydrolase, N-acetyl-L,L-diaminopimelate ami-notransferase	Conversion of L- glutamate to various amino acids,		
	Ornithine decarboxylase	Decarboxylation of L-ornithine to bio-synthesize putrescine., entry to polyamine metabolic pathway		

*Complete gene sets not necessarily present in presented LAB strains

possibility of in situ production by probiotics, and the ability to be bioengineered (Cotter et al., 2013). Cotter et al. provided a comprehensive review on these properties of bacteriocins suggesting their potential as alternatives to antibiotics (Cotter et al., 2013). Meanwhile, the use of bacteriocin from LAB has also attracted attention in studies involving biopreservation of food, as LAB are recognized as safe (Johnson et al., 2018). Bacteriocin improves the safety and shelf-life of products by inhibiting the growth of organisms associated with spoilage (Ng et al., 2020; Johnson et al., 2018).

Bacteriocins are usually encoded by small genes that are commonly omitted during the annotation process, a limitation to identification that BAGEL (de Jong et al., 2006), a web server created to identify putative bacteriocin open reading frames (ORFs) in DNA sequences, has come to alleviate (de Jong et al., 2006). Since its inception in 2006, BAGEL, presently updated to BAGEL4 (Van Heel et al., 2018), has been pivotal in bacteriocin identification in LAB (Eisenbach et al., 2018; Eisenbach et al., 2019; Goel et al., 2020; Quilodrán-Vega et al., 2020; Walsh et al., 2015; Wang et al., 2020; Yoo et al., 2017; Zhou et al., 2020). In fact, in 2015, Walsh et al. identified bacteriocin genes in silico from the genomes of the GIT subset of the Human Microbiome Project's reference genome database, which included members of Firmicutes, mainly using BAGEL (specifically BAGEL3) (Van Heel et al., 2013; Walsh et al., 2015). Meanwhile, other researchers have established a bacteriocin database for use in their research (Drissi et al., 2014).

Genes for adhesion and colonization

To exert its probiotic or beneficial effects on the host, the probiotic LAB must be able to adhere and persist in the gut (Valeriano et al., 2014; Zhang et al., 2015). Good adherence capacity of a probiotic strain promotes gut residence time, pathogen exclusion, and interaction with the host for the protection of epithelial cells or immunomodulation (Lebeer et al., 2008). The adhesion-associated genes identified to date in LAB include LPxTG protein-encoding genes, mucus-binding pilin (e.g. *spaC* gene), mucus-binding proteins, surface layer (S-layer) proteins, cell-surface collagen-binding proteins, and mannose-specific adhesin (Douillard and de Vos, 2014; Zhang et al., 2015). Comparative genomics tools for annotation and alignment, such as RAST and BLAST, respectively, as well as tools for protein prediction, such as SignalP (Almagro Armenteros et al., 2019), TMHMM (Krogh et al., 2001), LipoP (Juncker et al., 2003), and Pfam (Mistry et al., 2021), are just a few of the bioinformatics tools that have played vital roles in the identification of genes associated with adhesion and colonization of LAB (Abdelhamid et al., 2019; Kant et al., 2014; Quilodrán-Vega et al., 2020; Valeriano et al., 2014; Valeriano et al., 2019). In

2009, Kankainen et al. identified a cluster of pilus-encoding genes (*spaCBA*) in *L. rhamnosus* GG, which was found to be responsible for the higher adhesion ability of GG than *L. rhamnosus* LC705, an industrial stain routinely used as a starter culture, which lacks the *spaCBA* gene (Kankainen et al., 2009). *L. paracasei* subsp. *paracasei* IBB3423, isolated recently from raw cow's milk, was found to possess as many as 59 genes with putative adhesion domains, including the chromosome- and plasmid-encoded *spaCBA* gene clusters that are hypothesized to be responsible for the even stronger adhesion of strain IBB3423 compared to *L. rhamnosus* GG (Koryszewska-Bagińska et al., 2019). Meanwhile, 42 adhesion-associated proteins, including the novel *cwaA* gene encoding a protein containing multiple domains (i.e. five cell wall surface anchor repeat domains and an LPxTG-like cell wall anchor motif) were identified in *L. plantarum* NL42 (Zhang et al., 2015). The *cwaA* gene is believed to improve the adhesion ability of other probiotic strains as in *Lactococcus lactis* (Zhang et al., 2015).

Screening for genes that may provide direct health benefits to the host

The FAO/WHO defines probiotics as "live microorganisms that when administered confer health benefits to the host" (Tarrah et al., 2020). Probiotic LAB have been shown to provide a number of documented health benefits to the host, including the production of vitamins (particularly B vitamins), immunomodulation, antagonistic activity against pathogens, antitumor and anticancer potentials, alleviation of irritable bowel syndrome (IBS), combating obesity, serum cholesterol reduction, antihyperammonemia, etc. Genome analysis has also been used in the identification of genes related to these functions.

Production of vitamins

LAB are known to be fastidious microorganisms, and are auxotrophic for several vitamins and rely on food sources for vitamins, amino acids, nucleic acid derivatives, and sugar (Lebeer et al., 2008; Leblanc et al., 2011; Vogel et al., 2011; Walter et al., 2011; Zafar and Saier, 2020). However, some strains of LAB have been shown to synthesize vitamins, particularly the B-group vitamins, which include folates, riboflavin, and cobalamin, among others (Cárdenas et al., 2015; Leblanc et al., 2011; Morita et al., 2008; Zago et al., 2017). In 2008, genome analysis of *L. reuteri* JCM1112^T identified a unique cluster of 58 genes that includes genes involved in the biosynthesis of reuterin and cobalamin (*pdu-cbi-cob-hem* gene cluster) (Morita et al., 2008). This gene cluster was mapped within a genomic island that seemed to have been inserted into the genome of JMC1112^T via HGT. Meanwhile, homology search using blastn identified genes for the biosynthesis of folate and riboflavin in the genome of *Lactobacillus fermentum* (present name, *Limosilactobacillus fermentum*) CECT 571 (Cárdenas et al., 2015) and three strains of *L. plantarum* (Lp790, Lp813, and Lp998) (Zago et al., 2017). It has been shown that vitamins produced by gut commensals are absorbed in the large intestine and incorporated into the liver and kidneys (Asrar and O'Connor, 2005). As humans are incapable of synthesizing most vitamins and rely on exogenous sources, production of vitamins by probiotic LAB and starter cultures has been an area of active research in the food industry and for mitigation of clinical and subclinical vitamin deficiencies. Vitamin-producing LAB could be a cost-effective source of vitamins that are less likely to produce undesirable side effects in comparison to chemically synthesized pseudovitamins (Leblanc et al., 2011; LeBlanc et al., 2013).

Immunomodulation

Probiotic factors are those that directly contribute to the health-promoting effects of probiotics, and involve mechanisms such as maintaining microbial balance, epithelial protection, and immunomodulation (Lebeer et al., 2008). Lebeer et al. presented a comprehensive review of genes and molecules of lactobacilli supporting probiotic action that include the immunomodulatory interactions of lactobacilli involving different cell types, receptors, and cell structures (Lebeer et al., 2008). In 2008, Yasuda et al. surveyed the genome of *L. casei* Shirota, one of the pioneer probiotic strains, for candidate genes involved in the biosynthesis of polysaccharide moieties of the cell wall (extracellular polysaccharide (EPS) or capsular polysaccharide (CPS)) that are believed to have functions in immunomodulation (Yasuda et al., 2008). Genome analysis identified a cluster of 10 genes (*cps*1A to *cps*1J). The genes were mutated (knockout) and eight of which (*cps*1A to *cps*1G, and *cps*1J) were revealed to have roles in immunomodulation after immunological assays. Yasuda et al. employed genome analysis together with wet lab experiments and identified and verified genes involved in immunomodulation of *L. casei* Shirota that may function to reduce excessive immune reaction during activation of macrophages (Yasuda et al., 2008).

Some strains of *L. plantarum* are also known to exhibit immunomodulatory properties, and teichoic acids (TA) are recognized as key immunomodulatory molecules involved in regulation of the host immune response (Zhang et al., 2019b). Liu et al. (2015) conducted genome sequencing and comparative genomics of *L. plantarum* PS128, a probiotic strain with potential immunomodulatory activity, against seven other *L. plantarum* strains and found that PS128 contains a greater number of lipoteichoic acid exporter genes. However, further studies of these TA-related genes are

needed to determine their contributions to the antiinflammatory activity of PS128. Genome analysis of *L. plantarum* J26 also revealed TA-related genes that may be associated with its immunomodulatory effects (Zhang et al., 2019b). Meanwhile, genes for bacterial metabolites (i.e. related to the synthesis of bacteriocin and membrane proteins) or postbiotics in *L. plantarum* strains were also shown to be associated with positive immunomodulatory effects on dendritic cells (Zago et al., 2017). In a very recent study, Salvador et al. also used bioinformatics to mine for genes with functions related to bacteriocin, CPS, EPS, among others, for potential immunomodulatory properties of four probiotic strains isolated from plant and animal hosts (Salvador et al., 2021).

Several LAB strains are also known to be useful in treatment of IBS (Lebeer et al., 2008; Liu et al., 2015), a chronic intestinal disorder accompanied by low-grade inflammation, visceral hypersensitivity, and gut microbiota dysbiosis (Liu et al., 2021b). In a recent study, Liu et al. (2021b) investigated the abilities of strains of *L. casei* and *L. plantarum* to alleviate the symptoms of low-grade inflammation, visceral hypersensitivity, and gut microbiota dysbiosis in mice, together with comparative genomics, and identified *L. plantarum* CCFM8610 as a strain with significant ability to mediate the symptoms of IBS. Genome analysis and correlation analysis identified the ability to synthesize conjugated linoleic acid and COG1028-related genes as associated with the IBS-alleviating effects of *L. plantarum* CCFM8610.

Pathogen inhibition (antifungal and antiviral properties)

Pathogens are inhibited by LAB through the action of their metabolites such as organic acids, hydrogen peroxide, and bacteriocins, as well as by competitive exclusion brought about by the effective adherence of probiotics to the cells of the host (Lebeer et al., 2008; Quilodrán-Vega et al., 2020). Strains of *L. plantarum* and *L. crispatus* have been shown to inhibit fungal pathogens, such as *Candida albicans* and *Gardnerella vaginalis* (Beck et al., 2019; Ojala et al., 2014). Comparative genomics of 10 *L. crispatus* strains showed that *Lactobacillus* epithelial adhesin (LEA), among other proteins encoded in the core genome of *L. crispatus*, could mediate the competitive exclusion of *G. vaginalis* from epithelial cells, which is commonly associated with bacterial vaginosis (BV) (Ojala et al., 2014). Zafar and Saier (2020) identified a hemolysin that is a member of the pore-forming amphipathic helical peptide (HP2-20) family in 10 *Lactobacillus* spp. Members of the HP2-20 family are known to exhibit broad antifungal and antibacterial activities by creating pores in the cell membrane of target fungi or bacteria, with little or no lysis of mammalian cells (Zafar and Saier, 2020). Meanwhile, whole-genome characterization of *L. delbrueckii* TUA4408L identified a typical EPS gene cluster (Kanmani et al., 2018). Kanmani et al. demonstrated that *L. delbrueckii* TUA4408L and its EPS improved the resistance of porcine intestinal cells against rotavirus infection by reducing viral replication and regulation of inflammatory response (Kanmani et al., 2018). Viral titers of Newcastle disease (ND) and infectious bursal disease (IBD) were also reduced by probiotic *Lactobacillus* and *Bifidobacterium* strains to which the antiviral metabolites produced by the strains, such as lactic acid and bacteriocins, as well as the EPS have been attributed (Abdelhamid et al., 2019).

Antitumor and anticancer potentials of LAB

Several strains of LAB have been shown to have potential activities against tumors and cancer. As cancer-related mobility and morbidity are increasing worldwide, there have been a number of searches for alternative cancer prevention strategies, including the use of probiotics (Zhang et al., 2013). Some food and food components have been recognized as important adjuvants for cancer therapy like the fermented milk, a product of biological activities of LAB, that has been confirmed to be protective against colorectal cancer (Wu et al., 2021). Several species and strains of LAB have also been reported to have antitumor and anticancer potentials such as *L. brevis* KU15176 (Hwang et al., 2022), *L. plantarum* NCU116 (Zhou et al., 2017), and *L. salivarius* Ren (Zhang et al., 2013). Large numbers of studies involving in vitro (employing immune and cancer cell lines) and in vivo (cancer induced in mice and rats) evaluation of these potential anticancer and antitumor activities of LAB have been reported (Azam et al., 2014; Chondrou et al., 2018; El Ghany et al., 2015; Hwang et al., 2022; Nami et al., 2014; Saxami et al., 2016; Tukenmez et al., 2019; Tuo et al., 2015; Zhou et al., 2017). Very recently, Liu et al. (2021a) provided a mini-review of the anticancer substances from LAB, including the EPS, peptidoglycan, nucleic acid, bacteriocin, and S-layer proteins; while Wu et al. (2021) focused on the anticancer effects and mechanism of EPS from LAB. As anticancer and antitumor activities and potentials of LAB are commonly evaluated by wet lab experiments, genome analysis is usually complementary to these studies by assessing the quality and safety of the strains (Tegopoulos et al., 2021). Sun et al. (2015a) however, identified genes in *Lactobacillus salivarius* (present name, *Ligilactobacillus salivarius*) Ren that may be associated with its antitumor activity. *L. salivarius* Ren was previously shown to inhibit oral cancer induced by 4-nitroquioline 1-oxide (4NQO) in rats (Zhang et al., 2013). In vivo evidence indicated that protection of DNA against oxidative damage is one of the mechanisms by which *L. salivarius* Ren or its secretions inhibited the 4NQO-induced oral cancer (Zhang et al., 2013). Genes related to hydrogen peroxide production, including pyruvate oxidase, NADH oxidase, and α-glycerophosphate oxidase,

were identified in the genome of *L. salivarius* Ren (Sun et al. 2015a).

Others

Bioinformatics was also used to compare and understand strains of *Lactobacillus* that may have potential in promoting weight loss or weight gain and may have applications in combating obesity or malnutrition. Drissi et al. (2014) compared the gene sets of two groups of *Lactobacillus* species associated with weight gain and weight protection, respectively and found out that based on the genes present and absent in each group, defense mechanisms for enhanced glycolysis and defense against oxidative stress are developed in weight protection-associated *Lactobacillus* species, while a limited ability to breakdown fructose or glucose is associated with the weight gain associated *Lactobacillus*. In the study, genes encoding proteins involved in transcription and carbohydrate transport and metabolism were primarily identified in weight protection-associated *Lactobacillus*, whereas genes encoding proteins involved in replication, recombination, and repair were primarily identified in weight gain-associated *Lactobacillus* spp. Furthermore, weight protection-associated *Lactobacillus* spp. encodes significant copies of glucose permease, while thiolase, a lipid metabolism-associated enzyme was found only in weight-gain associated *Lactobacillus* spp.

Meanwhile, LAB were also shown to remove cholesterol. Growing, resting, and even dead cells of *L. plantarum* EM isolated from kimchi were demonstrated to have the ability to remove cholesterol and may have the potential to reduce serum cholesterol level (Choi and Chang, 2015). Recent complete genome analysis of *L. plantarum* EM identified five *bsh* genes related to the serum cholesterol-lowering effect of this strain (Kim et al., 2020). Bioinformatics analysis also contributed to the screening and identification of strains that can ameliorate hyperammonemia (Singh et al., 2018). Hyperammonemia occurs when there is increased production of ammonia and when detoxification is inhibited, or generally because of a disturbance in ammonia metabolism (Häberle, 2013; Singh et al., 2018). As ammonia is a known neurotoxin, elevated ammonia in the blood may result in encephalopathy, which is characterized by vomiting, hypotonia, lethargy, seizures, and coma, and may result in irreversible brain damage if not treated early and thoroughly (Häberle, 2013; Singh et al., 2018). Several thousand strains of probiotic bacteria were screened for the ability to reduce blood ammonia levels in mice with hyperammonemia, among which *Lactobacillus amylovorus* JBD401 showed the greatest ammonia-reducing activity (Singh et al., 2018). Genomic analysis of JBD401 identified genes involved in the assimilation of ammonia to synthesize amino acids and precursors for the polyamine metabolic pathway (Singh et al., 2018). The use of probiotics in the reduction of blood ammonia level is of potential clinical importance as high blood ammonia levels have been suggested to be associated with Alzheimer's disease.

Screening for quality and safety

Safety is another important characteristic of a potential probiotic LAB strain in addition to its ability to persist and colonize the GIT and provide benefits to the host. Therefore, the potential probiotic LAB must be assessed for its quality and safety. Many studies have used genomics to assess the quality and safety of potential and commercial probiotics (Beck et al., 2019; Evanovich et al., 2019; Tarrah et al., 2020; Wang et al., 2021). Whole-genome analysis is a promising technique for safety evaluation of probiotics (Wang et al., 2021). The presence or absence of the following features: CRISPR-Cas system, restriction and modification (RM) system – which reduce the incidence of phage invasion and foreign DNA integration; virulence factors (VF), antimicrobial resistance (AMR) genes, and toxic metabolites—features that are ideally absent in potential probiotic LAB; mobile genetic elements (MGEs), such as genomic islands, transposons, insertion sequences, plasmids, and phages—which could mobilize the AMR and virulence genes, as well as toxic metabolites; have been frequently investigated in studies characterizing potential probiotic LAB (Beck et al., 2019; Tarrah et al., 2020; Wang et al., 2021).

CRISPR-Cas and RM systems

Bacteria have evolved several mechanisms to counter the attack by bacteriophages and continue proliferation, such as adsorption inhibition, prevention of phage DNA penetration, host-controlled RM, and abortive infection (Sorek et al., 2008; Su et al., 1999). RM systems are the most common phage resistance mechanisms found in bacteria (Su et al., 1999). The first RM system in LAB and gram-positive bacteria, *LlaFI*, was identified in *L. lactis* in 1998 (Su et al., 1999). The function of the CRISPR-Cas system was verified in 2007 in another LAB, *Streptococcus thermophilus*, in which the CRISPR-Cas system was shown to mediate immunity against infection by phage (Barrangou et al., 2007; Sorek et al., 2008). The CRISPR-Cas system also limits the likelihood of transfer of AMR genes and virulence factors as it attacks foreign genetic material entering the cell (Douillard and de Vos, 2014; Wang et al., 2021). Several studies have scanned the genomes of potential probiotic LAB for the presence of RM and/or CRISPR-Cas systems (Evanovich et al., 2019; Goel et al., 2020; Jia et al., 2020; Kanmani et al., 2018; Valeriano et al., 2019; Wang et al., 2021; Xing et al., 2017; Xu et al., 2020; Zhang et al., 2019a)

using bioinformatics tools and databases that have been made available for prediction and analysis of these systems, such as REBASE (Roberts et al., 2009) and CRISPRFinder (Grissa et al., 2007a), and CRISPRdb (Grissa et al., 2007b). It is noteworthy, however, that several additional bioinformatics tools and databases have also been developed and made available for prediction and analysis of the CRISPR-Cas system (Alvarenga et al., 2018).

AMR genes

The emergence and spread of antimicrobial-resistant bacteria are major public health concerns. Bacteria develop resistance to antimicrobials through mutation or acquisition of resistance genes via HGT (Liu and Pop, 2008). One of the important characteristics of a probiotic LAB is the ability to inhibit the growth of pathogenic microorganisms. With the continuous spread of AMR genes and the increase in pathogenic microorganisms that are resistant to antimicrobials, probiotic LAB have been widely studied in the search for possible alternatives to conventional antimicrobial drugs (Abdelhamid et al., 2019). However, potential probiotic LAB must also be screened first for any transferable resistance genes in its genome that may contribute to the potential spread of AMR genes in the GIT. Ideally, a probiotic LAB must be susceptible to at least two antibiotics and must not show intrinsic antimicrobial resistance (Borriello et al., 2003; Wang et al., 2021). It should be noted that *Lactobacillus* species were found to be generally resistant to vancomycin (Borriello et al., 2003; Greppi et al., 2020), but the resistance gene appeared to be chromosomally located and therefore not easily transferable to other bacteria (Borriello et al., 2003). Other resistance genes against aminoglycosides and tetracycline have also been reported in *Lactobacillus* spp. (Greppi et al., 2020; O'Donnell et al., 2015). Bioinformatics databases and tools, such as the antibiotic resistance genes database (ARDB) (Liu and Pop, 2008), the comprehensive antibiotic resistance database (CARD) (McArthur et al., 2013), and ResFinder (Zankari et al., 2012), have been widely used in screening potential probiotic LAB for AMR genes.

Virulence factors and toxic metabolites

Probiotics are expected to provide benefits to the host and must not possess genes encoding virulence factors or for the production of toxic metabolites that may harm the host. Therefore, the genomes of potential probiotics are also scanned for the presence of genes encoding virulence factors and the production of toxic metabolites. The virulence factor database (VFDB) (Chen et al., 2005) has been widely used for this purpose, although RAST and Islandviewer4 (Bertelli et al., 2017) are also available to obtain information on the presence of virulence factors (Tarrah et al., 2020). Some common markers for safety evaluation of probiotic LAB include genes for the formation of biogenic amines, nitroreductase, amino acid decarboxylase, and azoreductases, β-glucuronidase, hemolysin, fibrinogen-binding proteins, α-chymotrypsin, and N-acetyl-β-glucosaminidase (Borriello et al., 2003; Son et al., 2017; Wang et al., 2021). Biogenic amines, which include tyramine, histamine, putrescine, spermidine, and cadaverine, are nitrogenous compounds that are formed during the decarboxylation of amino acids by bacteria (Cárdenas et al., 2015; Evanovich et al., 2019). Accumulations of these compounds have implications in food processing and storage, and may cause human health problems (Evanovich et al., 2019). The presence of β-glucuronidase is also monitored in LAB, as it is known to be carcinogenic and may re-dissociate carcinogens bound in glucuronic acid in the liver by detoxification in the colon, which can lead to their re-absorption in the body (Han et al., 2020; Yang et al., 2020). The genomes of *Lactobacillus* species were scanned for genes responsible for fibrinogen binding, which has been considered a potential pathogenicity trait, as it can lead to platelet aggregation, serious medical complications and, in some instances, death (Collins et al., 2012). LAB have been implicated in several cases where they were responsible for bacteremia, endocarditis, peritonitis, abscesses, and meningitis, often in immunocompromised individuals (Cannon et al., 2005; Zafar and Saier, 2020). Hence, the safety of LAB must be thoroughly examined and genetic analysis through bioinformatics plays a key role in this process, as potential pathogenicity traits may go undetected without full genetic analysis of strains (Collins et al., 2012).

Mobile genetic elements

The presence of mobile genetic elements (MGEs) in the genome of potential probiotic strains is commonly explored, as these are segments of DNA that move within and between bacteria and can therefore disseminate genes for AMR, virulence, and even toxic metabolite production (Evanovich et al., 2019; Wang et al., 2021). Several dedicated bioinformatics tools and databases have been developed for the identification of MGEs, such as plasmids (PSLDB; Galata et al., 2018), phages (PHAST and PHASTER; Arndt et al., 2016; Zhou et al., 2011), genomic islands (Islandviewer; Langille and Brinkman, 2009), and insertion sequences (ISFinder; Siguier et al., 2006). Numerous bioinformatics tools have been developed for prediction of MGEs (Alvarenga et al., 2018), as HGT events involving the transfer of AMR genes have both economic and public health consequences. Potential probiotic LAB have also been commonly analyzed for identification of MGEs in their genomes (Evanovich et al., 2019; Goel et al., 2020; Tarrah et al., 2020; Wang et al., 2021; Xu et al., 2020).

Adaptive evolution: de novo mutation of ingested transient probiotic LAB and its impact to the resident microbiota

As discussed previously, potential probiotic LAB is screened for the presence of genes that will allow it to survive the harsh conditions and pressures exerted by the host upon transit to the GIT. Recent studies have demonstrated that during the transit of probiotics toward the GIT, they also undergo adaptive evolution wherein de novo mutations (e.g. acquisition of SNPs) occur that modulate the expression of relevant genes (i.e. genes for carbon utilization, acid tolerance) and allow them to adapt to the challenging environment within the host (Huang et al., 2021). Highly consistent single nucleotide mutations (SNPs) were observed to be acquired by *L. plantarum* HNU082 (Lp082) as it colonized and adapted in the gut of humans, mice, and zebrafish, and such mutations were absent in in vitro cultivated Lp082 in MRS agar within 28 days which suggest that the mutation only occurred during the transit of Lp082 (Huang et al., 2021). Furthermore, mutant strains of Lp082 which harbored the SNPs in gene 1257 (annotated as bacterial alpha-L-rhamnosidase 6 hairpin glycosidase) and Gene 1717 (inner membrane protein response to acidic pH) exhibited higher capability of utilizing rhamnose and higher survival rate when grown in low pH condition in vitro, respectively than the original Lp082 strain.

The application of probiotics has been used as one approach to modulate the gut microbiota which has been recognized as a critical determinant of human health by directly contributing to pathologies, influencing the host's predisposition to disease, and providing cues to maintain metabolic and immunological functions (Walter et al., 2018). When the ingested probiotics reached the gut, however, not only does it encounter pathogenic microorganisms but a community of highly diverse microbial populations of different organisms that may consist of bacteria, archaea, and eukaryotes, essentially an open ecosystem, living in an intimate relationship with the host that is housed within the GIT (Derrien and van Hylckama Vlieg, 2015). Several studies have investigated the impact of ingested probiotics which, in ecological perspective, are essentially considered transient microorganisms attempting to invade an ecosystem of resident or autochthonous GIT community (Huang et al., 2021; Maldonado-Gómez et al., 2016; McNulty et al., 2011; Pinto et al., 2020). Walter et al. (2018) provided an ecological framework on the use of live microbes in the modulation of the gut microbiome. As most studies on the impact of probiotics on the resident microbiota involve the application of metagenomics pipelines, we will only briefly discuss the general observations in these studies with special attention to those that employed tools and databases mentioned in this review. For detailed review of computational tools for metagenomic pipelines readers may refer to the recent publication of Yang et al. (2021).

The ingestion of *L. plantarum* HNU082 (Lp082) induced minimal ecological changes (i.e. compositional and functional changes) in the human gut in the study of Huang et al., (2021) and this was in congruence with an earlier studies of Maldonado-Gómez et al. (2016) with *B. longum* AH1206, and McNulty et al. (2011) with consortium of fermented milk strains (FMS). However it is noteworthy that ingestion of Lp082 induced a dramatic change in the mutation frequency (one to two orders of magnitude greater in competitors than in related species in mouse model) of the resident gut microbes in human and mouse model (Huang et al., 2021). Meanwhile, introduction of fermented milk products (FMP) and FMS in the human and mouse model, respectively in the study of McNulty et al. (2011) resulted in significant changes in the expression of microbiome-encoded enzymes specifically those related to plant polysaccharide metabolism. *Bifidobacterium animalis* subsp. *lactis* upregulated a locus in vivo that is involved in the catabolism of xylooligosaccharides (McNulty et al., 2011). In these studies, BLAST tools, MEGA-X, the KEGG and CAZy databases were used in identification of mobile elements, phylogenetic analyses, and functional annotation, respectively.

Advantages, limitations, and challenges of bioinformatics

Bioinformatics provides easy access and an overview of the pool of information coded in the genome through available genome-mining tools and databases (Abdelhamid et al., 2019). Genome analyses through bioinformatics give detailed information on the potential metabolic capabilities of an organism and provide a map that opens up the potential for manipulation or bioengineering for strain improvement for biotherapeutics (Klaenhammer et al., 2005). In the study of adaptation and evolution, bioinformatics has been instrumental in the identification of novel genes that allow LAB to adapt and thrive in specific or diverse environments (Duar et al., 2017; Frese et al., 2011; Macklaim et al., 2011; Martino et al., 2016; Oh et al., 2010; van de Guchte et al., 2006). In the identification of potential LAB for probiotic application, bioinformatics has played key roles in initial screening and provides an efficient method to assess initially the safety of a probiotic strain, as genome analyses allow the detailed examination of genes, especially those involved in virulence or production of toxic metabolites, which may go unnoticed or unidentified by conventional wet lab experiments (Abdelhamid et al., 2019; Beck et al., 2019; Collins et al., 2012; Tarrah et al., 2020; Wang et al., 2021).

However, like any other method, bioinformatics analysis has its limitations and challenges. Most bioinformatics pipelines require familiarity with the Linux environment and programming languages, such as Python and Perl (Gauthier et al., 2018), which may be overwhelming to some classical researchers who are more adept with, and used to, performing conventional wet lab experiments. In addition, adequate funding is needed to provide the computers that support the storage and memory requirements of high-throughput analyses (Gauthier et al., 2018). Bioinformatics analysis also relies on good-quality assembly to provide comprehensive and reliable information, because incomplete (draft) or highly fragmented assemblies, which are more readily available than complete genomes, may introduce bias and yield inconclusive results (Quilodrán-Vega et al., 2020; Schmid et al., 2018). In addition, although bioinformatics analysis may be sufficient for some studies, e.g. phylogenomics using complete genome assemblies, in studies for clinical and industrial application of probiotic LAB, whole-genome analysis should always be supported by in vitro and in vivo assays to complement the initial analysis or findings, as some genes or determinants identified in the initial genome analysis are not always expressed in vitro and in vivo (Cárdenas et al., 2015; Zago et al., 2017).

In conclusion, bioinformatics has revolutionized the way by which we study organisms. In the era of modern-day molecular biology when genome sequencing has become relatively affordable and accessible, genome analysis through bioinformatics has become the gold standard for research. In the case of LAB, bioinformatics tools and databases for comparative genomics and phylogenomics have been widely used and have been shown to be helpful in understanding adaptation and eventual evolution of LAB in specific and/or diverse environments through the identification of specific genes for survival in certain habitats, up to phylogenetic analyses leading to the understanding of their specific lifestyles. Moreover, bioinformatics has made screening for potential probiotic LAB strains more efficient with easier and faster identification of specific desirable genes with regard to survival in the GIT as well as genes that may provide benefits to the host. Furthermore, bioinformatics has also facilitated more exhaustive safety evaluation of probiotic LAB, as genes encoding virulence factors, toxic metabolites, and antibiotic resistance can be easily predicted from the available tools and databases. These features however still need further validation through wet lab assays and expression experiments.

This review has presented some of the bioinformatics tools and databases that have been frequently employed in genome analysis of LAB. It should be noted, however, that the tools and databases presented here are only a subset of the vast collection of bioinformatics tools and databases available for genome analysis. This mini-review also discussed the role of bioinformatics in understanding the adaptation and evolution of LAB, and how bioinformatics has improved the efficiency and thoroughness of screening, and the safety evaluation of probiotic strains.

Acknowledgements This work was supported by the Korea Institute of Planning and Evaluation for Technology in Food, Agriculture, and Forestry (IPET) through the High Value-added Food Technology Development Program, funded by the Ministry of Agriculture, Food and Rural Affairs (MAFRA) (grant no. 321035052HD040). It was also supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2020R1A2C1010406). Figure 1 was partly generated using Servier Medical Art, provided by Servier, licensed under a Creative Commons Attribution 3.0 unported license accessed at https://smart.servier.com/category/cellular-biology/nucleic-acids/ and Vecteezy accessed at https://www.vecteezy.com/

Declarations

Conflict of interest The authors declare no conflict of interest.

References

Abdelhamid AG, El-Masry SS, El-Dougdoug NK. Probiotic *Lactobacillus* and *Bifidobacterium* strains possess safety characteristics, antiviral activities and host adherence factors revealed by genome mining. EPMA Journal. 10: 337-350 (2019)

Almagro Armenteros JJ, Tsirigos KD, Sønderby CK, Petersen TN, Winther O, Brunak S, von Heijne G, Nielsen H. SignalP 5.0 improves signal peptide predictions using deep neural networks. Nature Biotechnology. 37:420–423 (2019)

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. Journal of Molecular Biology. 215: 403-410 (1990)

Alvarenga DO, Moreira LM, Chandler M, Varani AM. Chapter 7 genetic elements in prokaryotic genomes. Methods in Molecular Biology. 1704 (2018)

Antipov D, Hartwick N, Shen M, Raiko M, Lapidus A, Pevzner PA. plasmidSPAdes: assembling plasmids from whole genome sequencing data. Bioinformatics. 32: 3380-3387 (2016)

Arndt D, Grant JR, Marcu A, Sajed T, Pon A, Liang Y, Wishart DS. PHASTER: a better, faster version of the PHAST phage search tool. Nucleic Acids Research. 44: W16 (2016)

Arnold JW, Simpson JB, Roach J, Kwintkiewicz J, Azcarate-Peril MA. Intra-species genomic and physiological variability impact stress resistance in strains of probiotic potential. Frontiers in Microbiology. 9: 1-14 (2018)

Asrar FM, O'Connor DL. Bacterially synthesized folate and supplemental folic acid are absorbed across the large intestine of piglets. Journal of Nutritional Biochemistry. 16: 587-593 (2005)

Azam R, Ghafouri-Fard S, Tabrizi M, Modarressi MH, Ebrahimzadeh-Vesal R, Daneshvar M, Mobasheri MB, Motevaseli E. *Lactobacillus acidophilus* and *Lactobacillus crispatus* culture supernatants downregulate expression of cancer-testis genes in the MDA-MB-231 cell line. Asian Pacific Journal of Cancer Prevention. 15: 4255-4259 (2014)

Aziz RK, Bartels D, Best A, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O. The RAST Server: Rapid

annotations using subsystems technology. BMC Genomics. 9: 1-15 (2008)
Bagon BB, Diane V, Valeriano V, Kyoung Oh J, Alain E, Pajarillo B, Lee JY, Kang D-K. Exoproteome Perspective on the Bile Stress Response of *Lactobacillus johnsonii.* Proteomes. 9: 1-14 (2021)
Bansal AK. Bioinformatics in microbial biotechnology – a mini review. Microbial Cell Factories. 4: 19 (2005)
Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, Romero DA, Horvath P. CRISPR provides acquired resistance against viruses in prokaryotes. Science. 315: 1709-1712 (2007)
Bateman A, Martin MJ, Orchard S, Magrane M, Agivetova R, Ahmad S, Alpi E, Bowler-Barnett EH, Britto R, Bursteinas B, Bye-A-Jee H, Coetzee R, Cukura A, Silva A Da, Denny P, Dogan T, Ebenezer TG, Fan J, Castro LG, Garmiri P, Georghiou G, Gonzales L, Hatton-Ellis E, Hussein A, Ignatchenko A, Insana G, Ishtiaq R, Jokinen P, Joshi V, Jyothi D, Lock A, Lopez R, Luciani A, Luo J, Lussi Y, MacDougall A, Madeira F, Mahmoudy M, Menchi M, Mishra A, Moulang K, Nightingale A, Oliveira CS, Pundir S, Qi G, Raj S, Rice D, Lopez MR, Saidi R, Sampson J, Sawford T, Speretta E, Turner E, Tyagi N, Vasudev P, Volynkin V, Warner K, Watkins X, Zaru R, Zellner H, Bridge A, Poux S, Redaschi N, Aimo L, Argoud-Puy G, Auchincloss A, Axelsen K, Bansal P, Baratin D, Blatter MC, Bolleman J, Boutet E, Breuza L, Casals-Casas C, de Castro E, Echioukh KC, Coudert E, Cuche B, Doche M, Dornevil D, Estreicher A, Famiglietti ML, Feuermann M, Gasteiger E, Gehant S, Gerritsen V, Gos A, Gruaz-Gumowski N, Hinz U, Hulo C, Hyka-Nouspikel N, Jungo F, Keller G, Kerhornou A, Lara V, Le Mercier P, Lieberherr D, Lombardot T, Martin X, Masson P, Morgat A, Neto TB, Paesano S, Pedruzzi I, Pilbout S, Pourcel L, Pozzato M, Pruess M, Rivoire C, Sigrist C, Sonesson K, Stutz A, Sundaram S, Tognolli M, Verbregue L, Wu CH, Arighi CN, Arminski L, Chen C, Chen Y, Garavelli JS, Huang H, Laiho K, McGarvey P, Natale DA, Ross K, Vinayaka CR, Wang Q, Wang Y, Yeh LS, Zhang J. UniProt: The universal protein knowledgebase in 2021. Nucleic Acids Research. 49: D480-D489 (2021)
Beck BR, Park GS, Lee YH, Im S, Jeong DY, Kang J. Whole genome analysis of *Lactobacillus plantarum* strains isolated from kimchi and determination of probiotic properties to treat mucosal infections by *Candida albicans* and *Gardnerella vaginalis*. Frontiers in Microbiology. 10: 1-13 (2019)
Bergsveinson J, Ziola B. Comparative genomic and plasmid analysis of beer-spoiling and non-beer-spoiling *Lactobacillus brevis* isolates. Canadian Journal of Microbiology. 63: 970-983 (2017)
Bertelli C, Laird MR, Williams KP, Lau BY, Hoad G, Winsor GL, Sl Brinkman F. IslandViewer 4: expanded prediction of genomic islands for larger-scale datasets. Nucleic Acids Research. 45: W30 (2017)
Blom J, Albaum SP, Doppmeier D, Pühler A, Vorhölter FJ, Zakrzewski M, Goesmann A. EDGAR: A software framework for the comparative analysis of prokaryotic genomes. BMC Bioinformatics. 154: 1-14 (2009)
de Boissier P, Habermann BH. A practical guide to orthology resources. (2016)
Borriello SP, Hammes WP, Holzapfel W, Marteau P, Schrezenmeir J, Vaara M, Valtonen V. Safety of Probiotics That Contain Lactobacilli or Bifidobacteria. Clinical Infectious Diseases. 36: 775-780 (2003)
Brettin T, Davis JJ, Disz T, Edwards RA, Gerdes S, Olsen GJ, Olson R, Overbeek R, Parrello B, Pusch GD, Shukla M, Thomason Iii JA, Stevens R, Vonstein V, Wattam AR, Fangfang Xia &. RASTtk: A modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. Scientific Reports. 5: 1-6 (2015)
Cannon JP, Lee TA, Bolanos JT, Danziger LH. Pathogenic relevance of *Lactobacillus*: A retrospective review of over 200 cases. European Journal of Clinical Microbiology and Infectious Diseases. 24: 31-40 (2005)
Cantalapiedra CP, Hernández-Plaza A, Letunic I, Bork P, Huerta-Cepas J. eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. Molecular Biology and Evolution. 38: 5825 (2021)
Cantarel BI, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B. The Carbohydrate-Active EnZymes database (CAZy): An expert resource for glycogenomics. Nucleic Acids Research. 37: 233-238 (2009)
Cárdenas N, Laiño JE, Delgado S, Jiménez E, Juárez del Valle M, Savoy de Giori G, Sesma F, Mayo B, Fernández L, LeBlanc JG, Rodríguez JM. Relationships between the genome and some phenotypical properties of *Lactobacillus fermentum* CECT 5716, a probiotic strain isolated from human milk. Applied Microbiology and Biotechnology. 99: 4343-4353 (2015)
Carriço JA, Rossi M, Moran-Gilad J, Van Domselaar G, Ramirez M. A primer on microbial bioinformatics for nonbioinformaticians. Clinical Microbiology and Infection. 24: 342-349 (2018)
Chaillou S, Champomier-Vergès MC, Cornet M, Crutz-Le Coq AM, Dudez AM, Martin V, Beaufils S, Darbon-Rongère E, Bossy R, Loux V, Zagorec M. The complete genome sequence of the meat-borne lactic acid bacterium *Lactobacillus sakei* 23K. Nature Biotechnology. 23: 1527-1533 (2005)
Chen F, Mackey AJ, Stoeckert CJ, Roos DS. OrthoMCL-DB: querying a comprehensive multi-species collection of ortholog groups. Nucleic acids research. 34: 363-368 (2006)
Chen L, Yang J, Yu J, Yao Z, Sun L, Shen Y, Jin Q. VFDB: a reference database for bacterial virulence factors. Nucleic Acids Research. 33 (2005)
Choi EA, Chang HC. Cholesterol-lowering effects of a putative probiotic strain *Lactobacillus plantarum* EM isolated from kimchi. LWT - Food Science and Technology. 62: 210-217 (2015)
Chondrou P, Karapetsas A, Kiousi DE, Tsela D, Anestopoulos I, Kotsianidis I. *Lactobacillus paracasei* K5 displays adhesion , antiproliferative activity and apoptotic effects in human colon cancer cells. Beneficial Microbes. 1–9 (2018)
Collins J, van Pijkeren JP, Svensson L, Claesson MJ, Sturme M, Li Y, Cooney JC, van Sinderen D, Walker AW, Parkhill J, Shannon O, O'Toole PW. Fibrinogen-binding and platelet-aggregation activities of a *Lactobacillus salivarius* septicaemia isolate are mediated by a novel fibrinogen-binding protein. Molecular Microbiology. 85:862–877 (2012)
Cotter PD, Ross RP, Hill C. Bacteriocins-a viable alternative to antibiotics? Nature Reviews Microbiology. 11:95–105 (2013)
Darling ACE, Mau B, Blattner FR, Perna NT. Mauve: Multiple alignment of conserved genomic sequence with rearrangements. Genome Research. 14:1394–1403 (2004)
Derrien M, van Hylckama Vlieg JET. Fate, activity, and impact of ingested bacteria within the human gut microbiota. Trends in Microbiology. 23:354–366 (2015)
Dieckmann MA, Beyvers S, Christian Nkouamedjo-Fankep R, Harald P, Hanel G, Jelonek L, Blom J, Goesmann A. EDGAR3.0: comparative genomics and phylogenomics on a scalable infrastructure. Nucleic Acids Research. 49: W185-W192 (2021)
Dominguez V, Angel D, Hjerde E, Sterck L, Capella-Gutierrez S, Notredame C, Pettersson OV, Amselem J, Bouri L, Bocs S, Klopp C, Gibrat J-F, Vlasova A, Leskosek BL, Soler L, Binzer-Panchal M, Lantz H. Ten steps to get started in Genome Assembly and Annotation. F1000Research. 148: 1-24 (2018)
Douillard FP, Ribbera A, Järvinen HM, Kant R, Pietilä TE, Randazzo C, Paulin L, Laine PK, Caggia C, von Ossowski I, Reunanen J, Satokari R, Salminen S, Palva A, de Vosa WM. Comparative genomic and functional analysis of *Lactobacillus casei* and *Lactobacillus rhamnosus* strains marketed as probiotics. Applied and Environmental Microbiology. 79:1923–1933 (2013)

Douillard FP, de Vos WM. Functional genomics of lactic acid bacteria: From food to health. Microbial Cell Factories. 13 (2014)

Drissi F, Merhej V, Angelakis E, Kaoutari A El, Carrière F, Henrissat B, Raoult D. Comparative genomics analysis of *Lactobacillus* species associated with weight gain or weight protection. Nutrition and Diabetes. 4 (2014)

Duar RM, Lin XB, Zheng J, Martino ME, Grenier T, Pérez-Muñoz ME, Leulier F, Gänzle M, Walter J. Lifestyles in transition: evolution and natural history of the genus *Lactobacillus*. FEMS microbiology reviews. 41:S27–S48 (2017)

Edgar RC. MUSCLE: A multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics. 5:1–19 (2004)

Eisenbach L, Geissler AJ, Ehrmann MA, Vogel RF. Comparative genomics of *Lactobacillus sakei* supports the development of starter strain combinations. Microbiological Research. 221:1–9 (2019)

Eisenbach L, Janßen D, Ehrmann MA, Vogel RF. Comparative genomics of *Lactobacillus curvatus* enables prediction of traits relating to adaptation and strategies of assertiveness in sausage fermentation. International Journal of Food Microbiology. 286:37–47 (2018)

Evanovich E, De Souza Mendonça Mattos PJ, Guerreiro JF. Comparative genomic analysis of *Lactobacillus plantarum*: An overview. International Journal of Genomics. 2019 (2019)

Feyereisen M, Mahony J, Kelleher P, Roberts RJ, O'Sullivan T, Geertman JMA, Van Sinderen D. Comparative genome analysis of the *Lactobacillus brevis* species. BMC Genomics. 20:1–15 (2019)

Fitch WM. Distinguishing homologous from analogous proteins. Systematic Zoology. 19:99–113 (1970)

Frese SA, Benson AK, Tannock GW, Loach DM, Kim J, Zhang M, Oh PL, Heng NCK, Patil PB, Juge N, MacKenzie DA, Pearson BM, Lapidus A, Dalin E, Tice H, Goltsman E, Land M, Hauser L, Ivanova N, Kyrpides NC, Walter J. The evolution of host specialization in the vertebrate gut symbiont *Lactobacillus reuteri*. PLoS Genetics. 7 (2011)

Galata V, Fehlmann T, Backes C, Keller A. PLSDB: a resource of complete bacterial plasmids. Nucleic Acids Research. 47:195–202 (2018)

Galperin MY, Makarova KS, Wolf YI, Koonin E V. Expanded Microbial genome coverage and improved protein family annotation in the COG database. Nucleic Acids Research. 43:D261–D269 (2015)

Gauthier J, Vincent AT, Charette SJ, Derome N. A brief history of bioinformatics. Briefings in Bioinformatics. 1–16 (2018)

El Ghany KA, Hamouda R, Elhafez EA, Mahrous H, Salem-Bekhit M, Hamza HA. A potential role of *Lactobacillus acidophilus* LA1 and its exopolysaccharides on cancer cells in male albino mice. Biotechnology and Biotechnological Equipment. 29:977–983 (2015)

Goel A, Halami PM, Tamang JP. Genome Analysis of *Lactobacillus plantarum* Isolated From Some Indian Fermented Foods for Bacteriocin Production and Probiotic Marker Genes. Frontiers in Microbiology. 11:1–12 (2020)

Greppi A, Asare PT, Schwab C, Zemp N, Stephan R, Lacroix C. Isolation and Comparative Genomic Analysis of Reuterin-Producing *Lactobacillus reuteri* From the Chicken Gastrointestinal Tract. Frontiers in Microbiology. 1166:1-17 (2020)

Grissa I, Vergnaud G, Pourcel C. CRISPRFinder: A web tool to identify clustered regularly interspaced short palindromic repeats. Nucleic Acids Research. 35:52–57 (2007a)

Grissa I, Vergnaud G, Pourcel C. The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. BMC Bioinformatics. 8:1–10 (2007b)

van de Guchte M, Penaud S, Grimaldi C, Barbe V, Bryson K, Nicolas P, Robert C, Oztas S, Mangenot S, Couloux A, Loux V, Dervyn R, Bossy R, Bolotin A, Batto J-M, Walunas T, Gibrat J-F, Bessie` res P, Weissenbach J, Ehrlich SD, Maguin E. Open access bulgaricus : a chronicle of evolution in action. PNAS. 103:9274–9279 (2006)

Häberle J. Clinical and biochemical aspects of primary and secondary hyperammonemic disorders. Archives of Biochemistry and Biophysics. 536:101–108 (2013)

Hadfield J, Croucher NJ, Goater RJ, Abudahab K, Aanensen DM, Harris SR. Phandango: an interactive viewer for bacterial population genomics. Bioinformatics, Application Note. 34:292–293 (2017)

Haft DH, Selengut JD, Richter RA, Harkins D, Basu MK, Beck E. TIGRFAMs and Genome Properties in 2013. Nucleic acids research. 41:D387–D395 (2013)

Haft DH, Selengut JD, White O. The TIGRFAMs database of protein families. Nucleic Acids Research. 31:371–373 (2003)

Hall BG. Building phylogenetic trees from molecular data with MEGA. Molecular Biology and Evolution. 30:1229–1235 (2013)

Han KJ, Lee JE, Lee NK, Paik HD. Antioxidant and Anti-inflammatory effect of probiotic *Lactobacillus plantarum KU15149* derived from Korean homemade Diced-radish kimchi. Journal of Microbiology and Biotechnology. 30:591–598 (2020)

Van Heel AJ, De Jong A, Montalbá N-Ló Pez M, Kok J, Kuipers OP. BAGEL3: automated identification of genes encoding bacteriocins and (non-)bactericidal posttranslationally modified peptides. Nucleic acids research. 41:W448-453 (2013)

Van Heel AJ, De Jong A, Song C, Viel JH, Kok J, Kuipers OP. BAGEL4: A user-friendly web server to thoroughly mine RiPPs and bacteriocins. Nucleic Acids Research. 46:W278–W281 (2018)

Huang S, Jiang S, Huo D, Allaband C, Estaki M, Cantu V, Belda-Ferre P, Vázquez-Baeza Y, Zhu Q, Ma C, Li C, Zarrinpar A, Liu YY, Knight R, Zhang J. Candidate probiotic *Lactiplantibacillus plantarum* HNU082 rapidly and convergently evolves within human, mice, and zebrafish gut but differentially influences the resident microbiome. Microbiome. 151: 1-17 (2021)

Huerta-Cepas J, Szklarczyk D, Heller D, Hernández-Plaza A, Forslund SK, Cook H, Mende DR, Letunic I, Rattei T, Jensen LJ, Von Mering C, Bork P. EggNOG 5.0: A hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. Nucleic Acids Research. 47:D309–D314 (2019)

Hwang C, Lee N, Paik H. The anti-cancer potential of heat-killed *Lactobacillus brevis* KU15176 upon AGS cell lines through intrinsic apoptosis pathway. IJMS. 23:4073 (2022)

Jia Y, Yang B, Ross P, Stanton C, Zhang H, Zhao J, Chen W. Comparative genomics analysis of *Lactobacillus mucosae* from different niches. Genes. 95:1-20 (2020)

Jin Ng Z, Abu Zarin M, Keong Lee C, Shun Tan J. Application of bacteriocins in food preservation and infectious disease treatment for humans and livestock: a review. The Royal Society of Chemistry. 10:38937–38964 (2020)

Johnson EM, Jung DYG, Jin DYY, Jayabalan DR, Yang DSH, Suh JW. Bacteriocins as food preservatives: Challenges and emerging horizons. Critical Reviews in Food Science and Nutrition. 58:2743–2767 (2018)

de Jong A, van Hijum SAFT, Bijlsma JJE, Kok J, Kuipers OP. BAGEL: A web-based bacteriocin genome mining tool. Nucleic Acids Research. 34:273–279 (2006)

Juncker AS, Willenbrock H, Von Heijne G, Brunak S, Nielsen H, Krogh A. Prediction of lipoprotein signal peptides in Gram-negative bacteria. Protein Science. 12:1652–1662 (2003)

Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K. KEGG: New perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Research. 45:D353–D361 (2017)

Kankainen M, Paulin L, Tynkkynen S, Von Ossowski I, Reunanen J, Partanen P, Satokari R, Vesterlund S, Hendrickx APA, Lebeer

S, De Keersmaecker SCJ, 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. Comparative genomic analysis of *Lactobacillus rhamnosus* GG reveals pili containing a human-mucus binding protein. Proceedings of the National Academy of Sciences of the United States of America. 106:17193–17198 (2009)

Kanmani P, Albarracin L, Kobayashi H, Hebert EM, Saavedra L, Komatsu R, Gatica B, Miyazaki A, Ikeda-Ohtsubo W, Suda Y, Aso H, Egusa S, Mishima T, Salas-Burgos A, Takahashi H, Villena J, Kitazawa H. Genomic characterization of *Lactobacillus delbrueckii* TUA4408L and evaluation of the antiviral activities of its extracellular polysaccharides in porcine intestinal epithelial cells. Frontiers in Immunology. 9:1–16 (2018)

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

Karin Lagesen DWU and TMW. Genome update. Microbiology Genome update. 156:603–608 (2010)

Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Molecular Biology and Evolution. 30:772–780 (2013)

Kim E, Choon H, Hae C, Kim Y. Complete genome sequence of *Lactobacillus plantarum* EM, A putative probiotic strain with the cholesterol - lowering effect and antimicrobial activity. Current Microbiology. (2020)

Klaenhammer TR, Barrangou R, Buck BL, Azcarate-Peril MA, Altermann E. Genomic features of lactic acid bacteria effecting bioprocessing and health. FEMS Microbiology Reviews. 29:393–409 (2005)

Kleerebezem M, Boekhorst J, van Kranenburg R, Molenaar D, Kuipers OP, Leer R, Tarchini R, Peters SA, Sandbrink HM, E J Fiers MW, Stiekema W, Klein Lankhorst RM, Bron PA, Hoffer SM, Nierop Groot MN, Kerkhoven R, de Vries M, rn Ursing B, de Vos WM, Siezen RJ. Complete genome sequence of *Lactobacillus plantarum* WCFS1. PNAS. 100:1990–1995 (2003)

Koryszewska-Bagińska A, Gawor J, Nowak A, Grynberg M, Aleksandrzak-Piekarczyk T. Comparative genomics and functional analysis of a highly adhesive dairy *Lactobacillus paracasei* subsp. *paracasei* IBB3423 strain. Applied Microbiology and Biotechnology. 103:7617–7634 (2019)

Krogh A, Larsson B, Von Heijne G, Sonnhammer ELL. Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. Journal of Molecular Biology. 305:567–580 (2001)

Langille MGI, Brinkman FSL. IslandViewer: an integrated interface for computational identification and visualization of genomic islands. Bioinformatics. 25:664–665 (2009)

Lebeer S, Vanderleyden J, De Keersmaecker SCJ. Genes and molecules of *Lactobacilli* supporting probiotic action. Microbiology and Molecular Biology Reviews. 72:728–764 (2008)

Leblanc JG, Laiñ JE, Juarez Del Valle M, Vannini V, Van Sinderen D, Taranto MP, Font De Valdez G, Savoy De Giori G, Sesma F. B-Group vitamin production by lactic acid bacteria-current knowledge and potential applications. Journal of Applied Microbiology. 111:1297–1309 (2011)

LeBlanc JG, Milani C, de Giori GS, Sesma F, van Sinderen D, Ventura M. Bacteria as vitamin suppliers to their host: A gut microbiota perspective. Current Opinion in Biotechnology. 24:160–168 (2013)

Lee JY, Han GG, Kim EB, Choi YJ. Comparative genomics of *Lactobacillus salivarius* strains focusing on their host adaptation. Microbiological Research. 205:48–58 (2017)

Li L, Stoeckert CJJ, Roos DS. OrthoMCL: Identification of ortholog groups for Eukaryotic genomes. Genome Research. 13:2178–2189 (2003)

Liu B, Pop M. ARDB-Antibiotic resistance genes database. Nucleic Acids Research. 37:443–447 (2008)

Liu C, Zheng J, Ou X, Han Y. Anti-cancer substances and safety of lactic acid bacteria in clinical treatment. Frontiers in Microbiology. 12:1–11 (2021)

Liu WH, Yang CH, Lin CT, Li SW, Cheng WS, Jiang YP, Wu CC, Chang CH, Tsai YC. Genome architecture of *Lactobacillus plantarum* PS128, a probiotic strain with potential immunomodulatory activity. Gut Pathogens. 7:1–7 (2015)

Liu Y, Xiao W, Yu L, Tian F, Wang G, Lu W, Narbad A, Chen W, Zhai Q. Evidence from comparative genomic analyses indicating that: *Lactobacillus* -mediated irritable bowel syndrome alleviation is mediated by conjugated linoleic acid synthesis. Food and Function. 12:1121–1134 (2021)

Macklaim JM, Gloor GB, Anukam KC, Cribby S, Reid G. At the crossroads of vaginal health and disease, the genome sequence of *Lactobacillus iners* AB-1. Proceedings of the National Academy of Sciences of the United States of America. 108:4688–4695 (2011)

Maldonado-Gómez MX, Martínez I, Bottacini F, O'Callaghan A, Ventura M, van Sinderen D, Hillmann B, Vangay P, Knights D, Hutkins RW, Walter J. Stable engraftment of *Bifidobacterium longum* AH1206 in the human gut depends on individualized features of the resident microbiome. Cell Host and Microbe. 20:515–526 (2016)

Mao B, Yin R, Li X, Cui S, Zhang H, Zhao J, Chen W. Comparative genomic analysis of *Lactiplantibacillus plantarum* isolated from different niches. Genes. 12:1–13 (2021)

Martino ME, Bayjanov JR, Caffrey BE, Wels M, Joncour P, Hughes S, Gillet B, Kleerebezem M, van Hijum SAFT, Leulier F. Nomadic lifestyle of *Lactobacillus plantarum* revealed by comparative genomics of 54 strains isolated from different habitats. Environmental Microbiology. 18:4974–4989 (2016)

McArthur AG, Waglechner N, Nizam F, Yan A, Azad MA, Baylay AJ, Bhullar K, Canova MJ, De Pascale G, Ejim L, Kalan L, King AM, Koteva K, Morar M, Mulvey MR, O'Brien JS, Pawlowski AC, Piddock LJV, Spanogiannopoulos P, Sutherland AD, Tang I, Taylor PL, Thaker M, Wang W, Yan M, Yu T, Wright GD. The comprehensive antibiotic resistance database. Antimicrobial Agents and Chemotherapy. 57:3348–3357 (2013)

McNulty NP, Yatsunenko T, Hsiao A, Faith JJ, Muegge BD, Goodman AL, Henrissat B, Oozeer R, Cools-Portier S, Gobert G, Chervaux C, Knights D, Lozupone CA, Knight R, Duncan AE, Bain JR, Muehlbauer MJ, Newgard CB, Heath AC, Gordon JI. The impact of a consortium of fermented milk strains on the gut microbiome of gnotobiotic mice and monozygotic twins. Science Translational Medicine. 106:1-26 (2011)

Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, Sonnhammer ELL, Tosatto SCE, Paladin L, Raj S, Richardson LJ, Finn RD, Bateman A. Pfam: The protein families database in 2021. Nucleic Acids Research. 49:D412–D419 (2021)

Morita H, Toh H, Fukuda S, Horikawa H, Oshima K, Suzuki T, Murakami M, Hisamatsu S, Kato Y, Takizawa T, Fukuoka H, Yoshimura T, Itoh K, O'sullivan DJ, Mckay LL, Ohno H, Kikuchi J, Masaoka T, Hattori M. Comparative genome analysis of *Lactobacillus reuteri* and *Lactobacillus fermentum* reveal a genomic island for reuterin and cobalamin production. DNA Research. 15:151–161 (2008)

Nami Y, Haghshenas B, Radiah D, Rosli R. Assessment of probiotic potential and anticancer activity of newly isolated vaginal bacterium *Lactobacillus plantarum* 5BL. Microbiology and Immunology. 58:492–502 (2014)

Nyquist OL, McLeod A, Brede DA, Snipen L, Aakra Å, Nes IF. Comparative genomics of Lactobacillus sakei with emphasis

on strains from meat. Molecular Genetics and Genomics. 285:297–311 (2011)
O'Donnell MM, Harris HMB, Lynch DB, Ross RP, O'Toole PW. *Lactobacillus ruminis* strains cluster according to their mammalian gut source. BMC Microbiology. 80:1-20 (2015)
O'Sullivan O, O'Callaghan J, Sangrador-Vegas A, McAuliffe O, Slattery L, Kaleta P, Callanan M, Fitzgerald GF, Ross RP, Beresford T. Comparative genomics of lactic acid bacteria reveals a niche-specific gene set. BMC Microbiology. 9:1–9 (2009)
Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, Kanehisa M. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Research. 27:29–34 (1999)
Oh PL, Benson AK, Peterson DA, Patil PB, Moriyama EN, Roos S, Walter J. Diversification of the gut symbiont *Lactobacillus reuteri* as a result of host-driven evolution. ISME Journal. 4:377–387 (2010)
Ojala T, Kankainen M, Castro J, Cerca N, Edelman S, Westerlund-Wikström B, Paulin L, Holm L, Auvinen P. Comparative genomics of *Lactobacillus crispatus* suggests novel mechanisms for the competitive exclusion of *Gardnerella vaginalis*. BMC Genomics. 15:1–21 (2014)
Overbeek R, Larsen N, Walunas T, D' Souza M, Pusch G, Selkov E, Liolios K, Joukov V, Kaznadzey D, Anderson I, Bhattacharyya A, Burd H, Gardner W, Hanke P, Kapatral V, Mikhailova N, Vasieva O, Osterman A, Vonstein V, Fonstein M, Ivanova N, Kyrpides N. The ErgoTM genome analysis and discovery system. Nucleic Acids Research. 31:164–171 (2003)
Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, Edwards RA, Gerdes S, Parrello B, Shukla M, Vonstein V, Wattam AR, Xia F, Stevens R. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). Nucleic Acids Research. 42:D207–D214 (2013)
Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MTG, Fookes M, Falush D, Keane JA, Parkhill J. Roary: Rapid large-scale prokaryote pan genome analysis. Bioinformatics. 31:3691–3693 (2015)
Pan M, Hidalgo-Cantabrana C, Goh YJ, Sanozky-Dawes R, Barrangou R. Comparative Analysis of *Lactobacillus gasseri* and *Lactobacillus crispatus* Isolated From Human Urogenital and Gastrointestinal Tracts. Frontiers in Microbiology. 10:1–17 (2020)
Petrova MI, Lievens E, Malik S, Imholz N, Lebeer S. *Lactobacillus* species as biomarkers and agents that can promote various aspects of vaginal health. Frontiers in Physiology. 6:1–18 (2015)
Petrova MI, Macklaim JM, Wuyts S, Verhoeven T, Vanderleyden J, Gloor GB, Lebeer S, Reid G. Comparative genomic and phenotypic analysis of the vaginal probiotic *Lactobacillus rhamnosus* GR-1. Frontiers in Microbiology. 9:1–12 (2018)
Pinto UM, Sun Z, Dias FS, Pimentel N, Rosa A, Marincola FC, Cosentino S, Pisano MB, Putzu D, Cesare Marincola F, Mossa V, Viale S, Fadda ME. Influence of autochthonous putative probiotic cultures on microbiota, lipid components and metabolome of Caciotta cheese. Frontiers in Microbiology. 11: 1-19 (2020)
Putonti C, Shapiro JW, Ene A, Tsibere O, Wolfe AJ. Comparative Genomic Study of *Lactobacillus jensenii* and the Newly Defined *Lactobacillus mulieris* Species Identifies Species- Specific Functionality. mSphere. 5:1–5 (2020)
Quilodrán-Vega S, Albarracin L, Mansilla F, Arce L, Zhou B, Islam MA, Tomokiyo M, Al Kassaa I, Suda Y, Kitazawa H, Villena J. Functional and genomic characterization of *Ligilactobacillus salivarius* TUCO-L2 isolated from Lama glama milk: A promising immunobiotic strain to combat infections. Frontiers in Microbiology. 11:1–24 (2020)
Roberts RJ, Vincze T, Posfai J, Macelis D. REBASE-A database for DNA restriction and modification: Enzymes, genes and genomes. Nucleic Acids Research. 38:2009–2011 (2009)
Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Ad` M-A, Rajandream A, Barrell B. Artemis: sequence visualization and annotation. Bioinformatics Applications Note. 16:944–945 (2000)
Salvador PBU, Dalmacio LMM, Kim SH, Kang D-K, Balolong MP. Immunomodulatory potential of four candidate probiotic *Lactobacillus* strains from plant and animal origin using comparative genomic analysis. Access Microbiology. 3: 1-9 (2021)
Saxami G, Karapetsas A, Lamprianidou E. Two potential probiotic *Lactobacillus* strains isolated from olive microbiota exhibit adhesion and anti-proliferative effects in cancer cell lines. Journal of Functional Foods. 24:461–471 (2016)
Schmid M, Muri J, Melidis D, Varadarajan AR, Somerville V, Wicki A, Moser A, Bourqui M, Wenzel C, Eugster-Meier E, Frey JE, Irmler S, Ahrens CH. Comparative genomics of completely sequenced *Lactobacillus helveticus* genomes provides insights into strain-specific genes and resolves metagenomics data down to the strain level. Frontiers in Microbiology. 9:1–20 (2018)
Seemann T. Genome analysis Prokka: rapid prokaryotic genome annotation. Bioinformatics. 30:2068–2069 (2014)
Senan S, Prajapati JB, Joshi CG. Comparative genome-scale analysis of niche-based stress-responsive genes in *Lactobacillus helveticus* strains. Genome. 57:185–192 (2014)
Siezen RJ, Van Enckevort FHJ, Kleerebezem M, Teusink B. Genome data mining of lactic acid bacteria: The impact of bioinformatics. Current Opinion in Biotechnology. 15:105–115 (2004)
Siguier P, Perochon J, Lestrade L, Mahillon J, Chandler M. ISfinder: the reference centre for bacterial insertion sequences. Nucleic Acids Research. 34:D32–D36 (2006)
Singh P, Chung HJ, Lee IA, D'Souza R, Kim HJ, Hong ST. Elucidation of the anti-hyperammonemic mechanism of *Lactobacillus amylovorus* JBD401 by comparative genomic analysis. BMC genomics. 19:292 (2018)
Son S, Oh JD, Lee SH, Shin D, Kim Y. Comparative genomics of canine *Lactobacillus reuter*i reveals adaptation to a shared environment with humans. Genes and Genomics. 42:1107–1116 (2020)
Son SH, Jeon HL, Yang SJ, Lee NK, Paik HD. In vitro characterization of *Lactobacillus brevis* KU15006, an isolate from kimchi, reveals anti-adhesion activity against foodborne pathogens and antidiabetic properties. Microbial Pathogenesis. 112:135–141 (2017)
Sorek R, Kunin V, Hugenholtz P. CRISPR - A widespread system that provides acquired resistance against phages in bacteria and archaea. Nature Reviews Microbiology. 6:181–186 (2008)
Stamatakis A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 30:1312–1313 (2014)
Su P, Im H, Hsieh H, Kang'a S, Dunn NW. LlaFI, a type III restriction and modification system in *Lactococcus lactis*. Applied and Environmental Microbiology. 65:686–693 (1999)
Sun E, Ren F, Liu S, Ge S, Zhang M, Guo H, Jiang L, Zhang H, Zhao L. Complete genome sequence of *Lactobacillus salivarius* Ren, a probiotic strain with anti-tumor activity. Journal of Biotechnology. 210:57–58 (2015)
Sun Z, Harris HMB, 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. Expanding the biotechnology potential of lactobacilli through comparative genomics of 213 strains and associated genera. Nature Communications. 6:1–13 (2015)
Tarrah A, Pakroo S, Corich V, Giacomini A. Whole-genome sequence and comparative genome analysis of *Lactobacillus paracasei* DTA93, a promising probiotic lactic acid bacterium. Archives of Microbiology. 202:1997–2003 (2020)

Tatusov RL, Koonin E V., Lipman DJ. A genomic perspective on protein families. Science. 278:631–637 (1997)

Tatusova T, Dicuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, Lomsadze A, Pruitt KD, Borodovsky M, Ostell J. NCBI prokaryotic genome annotation pipeline. Nucleic Acids Research. 44:6614–6624 (2016)

Tegopoulos K, Stergiou OS, Kiousi DE, Tsifintaris M, Koletsou E, Papageorgiou AC, Argyri AA, Chorianopoulos N, Galanis A, Kolovos P. Genomic and phylogenetic analysis of *Lactiplantibacillus plantarum* L125, and evaluation of its anti-proliferative and cytotoxic activity in cancer cells. Biomedicines. 9:1718 (2021)

Tettelin H, Riley D, Cattuto C, Medini D. Comparative genomics: the bacterial pan-genome. Current Opinion in Microbiology. 11:472–477 (2008)

Tukenmez U, Aktas B, Aslim B, Yavuz S. the relationship between the structural characteristics of lactobacilli-eps and its ability to induce apoptosis in colon cancer cells in vitro Ummugulsum. Nature Scientific Reports. 9:1–14 (2019)

Tuo Y, Jiang S, Qian F, Mu G, Liu P, Guo Y, Ma C. Short communication: Antiproliferative effect of 8 different *Lactobacillus* strains on K562 cells. Journal of Dairy Science. 98:106–110 (2015)

Valeriano VD, Parungao-Balolong M, Kang D-K. In vitro evaluation of the mucin-adhesion ability and probiotic potential of *Lactobacillus mucosae* LM1. Journal of Applied Microbiology. 117:485–497 (2014)

Valeriano VD V., Oh JK, Bagon BB, Kim H, Kang DK. Comparative genomic analysis of *Lactobacillus mucosae* LM1 identifies potential niche-specific genes and pathways for gastrointestinal adaptation. Genomics. 111:24–33 (2019)

van der Veer C, Hertzberger RY, Bruisten SM, Tytgat HLP, Swanenburg J, Angelino-Bart A de K, Schuren F, Molenaar D, Reid G, de Vries H, Kort R. Comparative genomics of human *Lactobacillus crispatus* isolates reveals genes for glycosylation and glycogen degradation: Implications for in vivo dominance of the vaginal microbiota. Microbiome. 7:1–14 (2019)

Verce M, De Vuyst L, Weckx S. Comparative genomics of *Lactobacillus fermentum* suggests a free-living lifestyle of this lactic acid bacterial species. Food Microbiology. 89:103448 (2020)

Vogel RF, Pavlovic M, Ehrmann MA, Wiezer A, Liesegang H, Offschanka S, Voget S, Angelov A, Böcker G, Liebl W. Genomic analysis reveals *Lactobacillus sanfranciscensis* as stable element in traditional sourdoughs. Microbial Cell Factories. 10:1–11 (2011)

Walsh CJ, Guinane CM, Hill C, Ross RP, O'Toole PW, Cotter PD. In silico identification of bacteriocin gene clusters in the gastrointestinal tract, based on the Human Microbiome Project's reference genome database. BMC Microbiology. 15:1–11 (2015)

Walter J, Britton RA, Roos S. Host-microbial symbiosis in the vertebrate gastrointestinal tract and the *Lactobacillus reuteri* paradigm. Proceedings of the National Academy of Sciences of the United States of America. 108:4645–4652 (2011)

Walter J, Maldonado-Gómez MX, Martínez I. To engraft or not to engraft: an ecological framework for gut microbiome modulation with live microbes. Current Opinion in Biotechnology. 49:129–139 (2018)

Wang S, Yang B, Ross P, Stanton C, Zhang H, Zhao J, Chen W. Comparative Genomics Analysis of *Lactobacillus ruminis* from Different Niches. Genes. 11:1–17 (2020)

Wang Y, Liang Q, Lu B, Shen H, Liu S, Shi Y, Leptihn S, Li H, Wei J, Liu Chengzhi, Xiao H, Zheng X, Liu Chao, Chen H. Whole-genome analysis of probiotic product isolates reveals the presence of genes related to antimicrobial resistance, virulence factors, and toxic metabolites, posing potential health risks. BMC Genomics. 22:1–12 (2021)

Wu J, Zhang Y, Ye L, Wang C. The anti-cancer effects and mechanisms of lactic acid bacteria exopolysaccharides in vitro: A review. Carbohydrate Polymers. 253:117308 (2021)

Xing Z, Geng W, Li C, Sun Y, Wang Y. Comparative genomics of *Lactobacillus kefiranofaciens* ZW3 and related members of *Lactobacillus*. spp reveal adaptations to dairy and gut environments. Scientific Reports. 7:1–12 (2017)

Xu S, Cheng J, Meng X, Xu Y, Mu Y. Complete genome and comparative genome analysis of *Lactobacillus reuteri* YSJL-12, a potential probiotics strain isolated from healthy sow fresh feces. Evolutionary Bioinformatics. 16:1–12 (2020)

Yang C, Chowdhury D, Zhang Z, Cheung WK, Lu A, Bian Z, Zhang L. A review of computational tools for generating metagenome-assembled genomes from metagenomic sequencing data. Computational and Structural Biotechnology Journal. 19:6301–6314 (2021)

Yang SJ, Kim KT, Kim TY, Paik HD. Probiotic properties and antioxidant activities of *Pediococcus pentosaceus* SC28 and *Levilactobacillus brevis* KU15151 in fermented black gamju. Foods. 9:1–12 (2020)

Yasuda E, Serata M, Sako T. Suppressive effect on activation of macrophages by *Lactobacillus casei* strain shirota genes determining the synthesis of cell wall-associated polysaccharides. Applied and Environmental Microbiology. 74:4746–4755 (2008)

Yin Y, Mao X, Yang J, Chen X, Mao F, Xu Y. DbCAN: A web resource for automated carbohydrate-active enzyme annotation. Nucleic Acids Research. 40:W445-W451 (2012)

Yoo DA, Bagon BB, Valeriano VD V., Oh JK, Kim H, Cho S, Kang DK. Complete genome analysis of *Lactobacillus fermentum* SK152 from kimchi reveals genes associated with its antimicrobial activity. FEMS microbiology letters. 364:1-10 (2017)

Zafar H, Saier MH. Comparative genomics of the transport proteins of ten *Lactobacillus* strains. Genes. 11:1–20 (2020)

Zago M, Scaltriti E, Bonvini B, Fornasari ME, Penna G, Massimiliano L, Carminati D, Rescigno M, Giraffa G. Genomic diversity and immunomodulatory activity of *Lactobacillus plantarum* isolated from dairy products. Beneficial Microbes. 8:597–604 (2017)

Zankari E, Hasman H, Cosentino S, Vestergaard M, Rasmussen S, Lund O, Aarestrup FM, Larsen MV. Identification of acquired antimicrobial resistance genes. Journal of Antimicrobial Chemotherapy. 67:2640–2644 (2012)

Zhang B, Zuo F, Yu R, Zeng Z, Ma H, Chen S. Comparative genome-based identification of a cell wall-anchored protein from Lactobacillus plantarum increases adhesion of *Lactococcus lactis* to human epithelial cells. Scientific Reports. 5:1–12 (2015)

Zhang M, Wang F, Jiang L, Liu R, Zhang L, Lei X, Li J, Jiang J, Guo H, Fang B, Zhao L, Ren F. *Lactobacillus Salivarius Ren* inhibits rat oral cancer induced by 4-nitroquioline 1-oxide. Cancer Prevention Research. 686–694 (2013)

Zhang Q, Zhang L, Ross P, Zhao J, Zhang H, Chen W. Comparative genomics of *Lactobacillus crispatus* from the gut and vagina reveals genetic diversity and lifestyle adaptation. Genes. 11:1–14 (2020)

Zhang W, Wang J, Zhang D, Liu H, Wang S, Wang Y, Ji H. Complete genome sequencing and comparative genome characterization of *Lactobacillus johnsonii* ZLJ010, a potential probiotic with health-promoting properties. Frontiers in Genetics. 10:1–13 (2019)

Zhang Z, Man C, Sun L, Yang X, Li M, Zhang W, Jiang Y. Short communication: Complete genome sequence of *Lactobacillus plantarum* J26, a probiotic strain with immunomodulatory activity. Journal of Dairy Science. 102:10838–10844 (2019)

Zhao Y, Sun C, Zhao D, Zhang Y, You Y, Jia X, Yang J, Wang L, Wang J, Fu H, Kang Y, Chen F, Yu J, Wu J, Xiao J. PGAP-X: Extension on pan-genome analysis pipeline. BMC Genomics. 19 (2018)

Zhao Y, Wu J, Yang J, Sun S, Xiao J, Yu J. PGAP: Pan-genomes analysis pipeline. Bioinformatics. 28:416–418 (2012)

Zheng J, Wittouck S, Salvetti E, Franz CMAP, Harris HMB, Mattarelli P, O'toole PW, Pot B, Vandamme P, Walter J, Watanabe K, Wuyts S, Felis GE, Gänzle MG, Lebeer S. A taxonomic note on the genus Lactobacillus: Description of 23 novel genera, emended description of the genus *Lactobacillus beijerinck* 1901, and union of Lactobacillaceae and Leuconostocaceae. International Journal of Systematic and Evolutionary Microbiology. 70:2782–2858 (2020)

Zheng J, Zhao X, Lin XB, Gänzle M. Comparative genomics *Lactobacillus reuteri* from sourdough reveals adaptation of an intestinal symbiont to food fermentations. Scientific Reports. 5:1–11 (2015)

Zhou X, Hong T, Yu Q, Nie S, Gong D, Xiong T, Xie M. Exopolysaccharides from *Lactobacillus plantarum* NCU116 induce c-Jun dependent Fas/Fasl-mediated apoptosis via TLR2 in mouse intestinal epithelial cancer cells. Scientific Reports. 1–13 (2017)

Zhou X, Yang B, Stanton C, Ross RP, Zhao J, Zhang H, Chen W. Comparative analysis of *Lactobacillus gasseri* from Chinese subjects reveals a new species-level taxa. BMC Genomics. 21:1–16 (2020)

Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS. PHAST: A fast phage search tool. Nucleic Acids Research. 39:W347–W352 (2011)