



Article

Reanalysis of *Lactobacillus paracasei* Lbs2 Strain and Large-Scale Comparative Genomics Places Many Strains into Their Correct Taxonomic Position

Samrat Ghosh ^{1,2}, Aditya Narayan Sarangi ¹, Mayuri Mukherjee ^{1,2}, Swati Bhowmick ¹ and Sucheta Tripathy ^{1,2,*}

- ¹ Structural Biology and Bioinformatics Division, CSIR-Indian Institute of Chemical Biology, 4 Raja S.C. Mullick Road, Kolkata 700032, India; samrat.ghosh2010@gmail.com (S.G.); a.n.sarangi@gmail.com (A.N.S.); mayuri1006@gmail.com (M.M.); bhowmickswati@gmail.com (S.B.)
² Academy of Scientific and Innovative Research (AcSIR), Ghaziabad 201002, India
* Correspondence: tsucheta@gmail.com or tsucheta@iicb.res.in; Tel.: +91-33-24995894

Received: 10 August 2019; Accepted: 14 October 2019; Published: 25 October 2019



Abstract: *Lactobacillus paracasei* are diverse Gram-positive bacteria that are very closely related to *Lactobacillus casei*, belonging to the *Lactobacillus casei* group. Due to extreme genome similarities between *L. casei* and *L. paracasei*, many strains have been cross placed in the other group. We had earlier sequenced and analyzed the genome of *Lactobacillus paracasei* Lbs2, but mistakenly identified it as *L. casei*. We re-analyzed Lbs2 reads into a 2.5 MB genome that is 91.28% complete with 0.8% contamination, which is now suitably placed under *L. paracasei* based on Average Nucleotide Identity and Average Amino Acid Identity. We took 74 sequenced genomes of *L. paracasei* from GenBank with assembly sizes ranging from 2.3 to 3.3 MB and genome completeness between 88% and 100% for comparison. The pan-genome of 75 *L. paracasei* strains hold 15,945 gene families (21,5232 genes), while the core genome contained about 8.4% of the total genes (243 gene families with 18,225 genes) of pan-genome. Phylogenomic analysis based on core gene families revealed that the Lbs2 strain has a closer relationship with *L. paracasei* subsp. *tolerans* DSM20258. Finally, the in-silico analysis of the *L. paracasei* Lbs2 genome revealed an important pathway that could underpin the production of thiamin, which may contribute to the host energy metabolism.

Keywords: *Lactobacillus paracasei*; horizontal gene transfer; carbohydrate active enzyme; thiamin; pan/core-genome; COG; host-specific gene

1. Introduction

Among the lactic acid bacteria, *Lactobacillus* is the prominent genus, which presently embraces more than 200 species [1]. These species are commonly found in a wide variety of niches (e.g., the gastrointestinal tract, fruit, vegetables, wine, milk, and meat), including the complex microbial community. These bacteria are widely used in many biotechnological applications, e.g., as a vaccine carrier, in bioplastic production, as probiotics, or as starter cultures, indicating their high commercial value [1].

Furthermore, *L. paracasei* is remarkable in adapting to diverse habitats, especially to the gastrointestinal tract. Like other lactic acid bacteria, strains of *L. paracasei* are also extensively used as a starter culture for dairy products, in the food industry, and as probiotics [2]. The taxonomic classification of *L. paracasei* and *L. casei* has been a matter of extensive debate [2]. In the NCBI database (GenBank), several strains of *L. casei* species are misclassified, and these strains needed to be placed under the *L. paracasei* species [3–5]. Thus, in this study, we did extensive analysis to place the species in their respective corrected taxa.

Besides this, comparative genomic analysis among several *L. paracasei* strains reveal crucial features, such as probiotic properties and niche adaptation [2]. Notably, the function of the sortase enzyme is allied with probiotic traits, such as immune signaling and adhesion to the intestinal mucosa [6]. These enzymes are involved in the modification of cell surface proteins containing the LPXTG motif among bacteria. The number of proteins containing LPXTG motifs vary from one to more than forty per genome in many species [7]. Some reports also emphasized the role of sortase in gut bacteria where cell envelope proteins are crucial for the establishment of interaction between host and probiotic strains [8]. Interestingly, it is also noticeable that *Lactobacillus paracasei* produces bacteriocins in the gastrointestinal tract [9]. These are natural antimicrobial substances (biologically active peptides) that act against other bacteria having a specific immune mechanism [10]. In some cases, bacteriocins are viewed as an essential property for the identification of a probiotic strain [11].

The metabolism of oligosaccharides is essential for the ecological fitness of Lactobacilli strains, though there is very little knowledge of the capability of Lactobacilli to exploit oligosaccharides as a carbon source [12]. In *Lactobacillus casei*, fermentation is a crucial source of energy, which can ferment glucose, fructose, mannose, galactose, mannitol, *N*-acetylglucosamine, and tagatose [13]. It has also been reported that commensal bacteria can contribute significantly to human metabolism through vitamin B12 production. They could also synthesize vitamins B and K [14]. *Lactobacillus rhamnosus* GG, isolated from the healthy human gut is capable of incorporating vitamins like B1, B2, and B9 in the culture medium. This is the sole probiotic strain in which thiamin production has been reported [15].

Here we present a detailed analysis of genomes of the *Lactobacillus casei* group and propose corrections of 38 strains earlier identified as *L. casei* into the *L. paracasei* species after extensive genomics analysis. Our earlier assembly of the Lbs2 strain was carried out using Allpaths-LG-49856, where the GC filtration method was used to remove contaminating reads, and the genome was described as *Lactobacillus casei* Lbs2 [16]. However, analysis with CheckM revealed that the genome was only 60% complete with 1.6% contamination. We, therefore, re-assembled the genome using SPAdes assembler, followed by contamination and completeness detection using CheckM, which provided a better result (completeness: 91.28%; contamination: 0.87%) than the previous assembly. Average Nucleotide Identity (ANI) and Average Amino Acid Identity (AAI) analysis of the re-assembled genome (Lbs2) was done with 123 publicly available *L. paracasei* ($n = 74$) and *L. casei* ($n = 49$) genome assemblies, which classified our strain as a member belonging to *L. paracasei* species. Next, we performed pan-genome analysis among the *L. paracasei* strains ($n = 75$) to understand the aspects of evolution, metabolism, physiology and technological properties. Finally, an in-depth analysis revealed the technological properties of *L. paracasei* Lbs2, which may help in improving host energy metabolism.

2. Materials and Methods

2.1. Brief Description of the Key Software Used in This Study

NxTrim (Chesterford Research Park, Little Chesterford, UK): This software removes Nextera mate-pair junction adaptors from the raw reads.

SPAdes (Algorithmic Biology Laboratory, St. Petersburg, Russia): An Eulerian de Bruijn graph-based assembler, designed for the assembly of single-cell and multi-cells bacterial data sets.

CheckM (Australian Centre for Ecogenomics, Queensland, Australia): A set of tools for assessing the quality of genomes recovered from isolates, single cells, or metagenomes. It estimates genome completeness and contamination by using collocated sets of genes that are ubiquitous and single-copy within a phylogenetic lineage.

BUSCO (Swiss Institute of Bioinformatics, Geneva, Switzerland): BUSCO (Benchmarking Universal Single-Copy Orthologs) comprises of lineage-specific core Orthologs. Based on the similarity search, the quality of genome assembly and the presence of gene content is assessed.

Prokka (Monash University, Clayton, Australia): A rapid prokaryotic genome annotation pipeline.

Roary (The Wellcome Trust Sanger Institute, Hinxton, Cambridge): A high-speed standalone pan-genome pipeline, which takes annotated assemblies in GFF3 format as input (produced by Prokka) and calculates the pan-genome.

FastTree (Lawrence Berkeley National Lab, Berkeley, CA, USA): Infers approximately-maximum-likelihood phylogenetic trees from alignments of nucleotide or protein sequences.

2.2. Processing of Reads, Re-Assembly, and Acquisition of Publicly Available Assemblies

Illumina paired-end (length 151 bp, coverage 500×) and mate-pair (length 101bp, coverage 150×) libraries of *L. casei* Lbs2 were previously generated using the Illumina MiSeq platform [16], and were contamination-checked and re-assembled to fix the assembly issues. For this, the reads from the paired-end library ($n = 4,784,475 \times 2$) were quality checked using FastQC (<https://github.com/s-andrews/FastQC>), which revealed all reads were of fairly good quality, without adapters and with an average length of 150 nt. Contaminant reads from the mate-pair library ($n = 2,180,266 \times 2$) were purged using NxTrim [17]. Finally, cleaned paired-end along with mate-pair reads were re-assembled using SPAdes v3.11.1 genome assembler [18] with k-mers 25 to 97 with a step size of 6 and coverage cutoff as auto.

We retrieved all the genome assemblies classified as *L. paracasei* ($n = 74$) and *L. casei* ($n = 49$), (a total of 123 genomes) from the NCBI genome database (Accessed on 8 July 2018).

2.3. Quality Assessment of Genomes

Contigs obtained using SPAdes v3.11.1 were analyzed with CheckM v1.0.11 (Australian Centre for Ecogenomics, Queensland, Australia) [19] to evaluate the completeness and contamination using the lineage_wf workflow. Plausible contaminants were detected iteratively with reference to tetra-nucleotide distributions using the ‘outliers’ command followed by contamination detection, to get a final set of contigs with maximum completeness and minimum contamination. Completeness was also cross verified using Benchmarking Universal Single-Copy Orthologs (BUSCO) [20]. The above process of contamination and completeness detection was done for all the genomes downloaded from the NCBI database as *L. paracasei* and *L. casei* as well.

2.4. Average Nucleotide Identity (ANI) and Average Amino Acid Identity (AAI) Calculation

To check the genetic relatedness, we calculated ANI and AAI with all *L. paracasei* and *L. casei* genomes with the *L. paracasei* Lbs2 strain as a reference based on the “pyani” (<https://github.com/widowquinn/pyani>) and “CompareM” (<https://github.com/dparks1134/CompareM>) package, respectively. All the results were visualized and plotted with R package “pheatmap” in R studio (<https://www.rstudio.com>).

2.5. Gene Prediction and Annotation

Coding sequences (CDSs) and all types of RNA genes of the Lbs2 strain and 74 *L. paracasei* strains were predicted and annotated with Prokka v1.13 (Monash University, Clayton, Australia) [21]. cluster of orthologous groups (COG) and pathway annotations were carried out with STRING [22] and KEGG databases [23] using BLASTP [24] keeping an e-value cut-off of $<10^{-5}$, respectively. Furthermore, for cataloging, subsystem categories of the *L. paracasei* Lbs2 strain were annotated using the Rapid Annotation using Subsystem Technology (RAST) webserver [25].

2.6. Construction of Pan/Core-Genome Families and Unique Genes

GFF formatted files derived from Prokka (v1.13) were analyzed using Roary (Rapid standalone pan genomic pipeline) [26] to elucidate the Pan/core-genome and unique genes. In Roary, Markov cluster (MCL) algorithm [27] was used to cluster orthologous genes. Additionally, core-genome and

singleton analysis, among *L. paracasei* Lbs2, *L. paracasei* subsp. *tolerans* DSM20258, and *L. paracasei* Lpc-37 strains were carried out with OrthoVenn (a web platform for orthologous gene clustering) [28].

2.7. Phylogenomic Analysis and Ka/Ks Calculation of Core Genomes

A phylogenomic tree based on core genes among the *L. paracasei* strains was built with the help of Roary and FastTree software (Lawrence Berkeley National Lab, Berkeley, CA, USA) [29]. Core gene sets were aligned using Roary based mafft [30], alignment was concatenated, and the final alignment was taken as input in the FastTree package for the construction of the maximum likelihood tree. Finally, the tree was visualized and edited using iTOL v4 [31].

The Ka/Ks value for each core gene pair was computed by the KaKs_Calculator [32] and ParaAT software (Beijing Institute of Genomics, Beijing, China) [33].

2.8. Comparative Genomic Analysis

Full-length genomes of *L. paracasei* Lbs2, *L. paracasei* subsp. *tolerans* DSM20258, and *L. paracasei* Lpc-37 were aligned by progressive MAUVE [34]. The circular map of *Lactobacillus paracasei* Lbs2 genome was created by the DNAPlotter software [35]. Comparison of important core-gene clusters and thiamin biosynthetic gene clusters among *L. paracasei* Lbs2, *L. paracasei* subsp. *tolerans* DSM20258, and *L. paracasei* Lpc-37 was done with the Easyfig comparison tool [36]. Stress linked genes and carbohydrate transporters were determined using the GO database (<http://archive.geneontology.org>). The carbohydrate activating enzyme families were detected by using the stand-alone dbCAN [37] against CAZy database [37]. In addition, sortases and proteins carrying LPXTG motifs were predicted with the stand-alone LOCP tool [38]. Furthermore, bacteriocins, CRISPRs, prophages, ISs, and GIs were identified using the BAGEL4 web-based resource [39], CRISPR finder stand-alone tool [40], PHASTER web server [41], ISSaga web platform [42,43], and IslandViewer4 [44], respectively.

2.9. Data Availability

Raw reads of the *Lactobacillus paracasei* Lbs2 strain are available under Bioproject PRJNA255080, and the re-assembled genome sequence of the same is present at GenBank with the accession number JPKN00000000.3.

3. Results and Discussion

3.1. Genomic Properties of the *L. paracasei* Strains

Pair-wise genome comparison metrics, such as Average Nucleotide Identity (ANI) and Average Amino Acid Identity (AAI) with >95% threshold as the cut-off, are frequently used as an operative method for species boundary demarcation [45,46]. In addition, it is also evidenced that the confidence of classification or probability of misclassification of a newly assembled genome depends on only ANI and AAI values from multiple genome (established species genome) comparison. In the present analysis, we calculated the ANI and AAI values for the reassembled Lbs2 genome with *L. paracasei* ($n = 74$) strains and *L. casei* ($n = 49$) strains separately. The ANI values are found to be around 98% and AAI values varied from 97% to 99% with *L. paracasei* ($n = 74$) strains (Supplementary Figure S1 and Supplementary Table S1). With *L. casei* ($n = 49$), the ANI and AAI varied from 77% to 98%, and 85% to 98%, respectively (Supplementary Figure S2 and Supplementary Table S2). Therefore, we inferred that our reassembled genome, Lbs2, should be placed under the *L. paracasei* species. Moreover, as per ANI and AAI analysis, we proposed that 38 publicly available strains of *L. casei* need to be moved out and placed under the *L. paracasei* species (Table 1).

To evaluate the degree of completeness among the *L. paracasei* genomes ($n = 75$), we used CheckM. The completeness of the reassembled *L. paracasei* Lbs2 genome is around 91.28% with 0.81% contamination compared to the earlier reported analysis with 62% completeness and 1.6% contamination (Supplementary Table S3A) [16]. Among the strains studied, the genome of the *L. paracasei* Lpc-37

strain is found to be at a higher level of completeness (100%) and without any contamination (0%), as per the CheckM output (Supplementary Table S3B). Because of this, the *L. paracasei* Lpc-37 strain is considered for further analysis. Genome assembly completeness based on BUSCO analysis among the *L. paracasei* strains varied from 42.7% to 99.6% (Supplementary Table S3C).

Table 1. Proposed name of *L. casei* strains based on the Average Nucleotide Identity (ANI) and Average Amino Acid Identity (AAI) calculation.

Existing Name in NCBI Database (GenBank)	Proposed Name
<i>Lactobacillus casei</i> 12A (Complete genome)	<i>Lactobacillus paracasei</i> 12A
<i>Lactobacillus casei</i> 12A (Draft)	<i>Lactobacillus paracasei</i> 12A
<i>Lactobacillus casei</i> 1316.rep1_LPAR (Scaf no.170)	<i>Lactobacillus paracasei</i> 1316.rep1_LPAR
<i>Lactobacillus casei</i> 1316.rep2_LPAR (Scaf no.264)	<i>Lactobacillus paracasei</i> 1316.rep2_LPAR
<i>Lactobacillus casei</i> 21/1	<i>Lactobacillus paracasei</i> 21/1
<i>Lactobacillus casei</i> 32G	<i>Lactobacillus paracasei</i> 32G
<i>Lactobacillus casei</i> 5b	<i>Lactobacillus paracasei</i> 5b
<i>Lactobacillus casei</i> 844_LCAS	<i>Lactobacillus paracasei</i> 844_LCAS
<i>Lactobacillus casei</i> A2-362 (Scaffold no. 162)	<i>Lactobacillus paracasei</i> A2-362
<i>Lactobacillus casei</i> A2-362 (Contig no. 167)	<i>Lactobacillus paracasei</i> A2-362
<i>Lactobacillus casei</i> BD II	<i>Lactobacillus paracasei</i> BD II
<i>Lactobacillus casei</i> BL23	<i>Lactobacillus paracasei</i> BL23
<i>Lactobacillus casei</i> BM-LC14617	<i>Lactobacillus paracasei</i> BM-LC14617
<i>Lactobacillus casei</i> CRF28	<i>Lactobacillus paracasei</i> CRF28
<i>Lactobacillus casei</i> DPC6800	<i>Lactobacillus paracasei</i> DPC6800
<i>Lactobacillus casei</i> DSM 20011	<i>Lactobacillus paracasei</i> DSM 20011
<i>Lactobacillus casei</i> GCRL163	<i>Lactobacillus paracasei</i> GCRL163
<i>Lactobacillus casei</i> HDS-01	<i>Lactobacillus paracasei</i> HDS-01
<i>Lactobacillus casei</i> HZ-1	<i>Lactobacillus paracasei</i> HZ-1
<i>Lactobacillus casei</i> KL1-Liu	<i>Lactobacillus paracasei</i> KL1-Liu
<i>Lactobacillus casei</i> LC2W	<i>Lactobacillus paracasei</i> LC2W
<i>Lactobacillus casei</i> LOCK919	<i>Lactobacillus paracasei</i> LOCK919
<i>Lactobacillus casei</i> Lc-10	<i>Lactobacillus paracasei</i> Lc-10
<i>Lactobacillus casei</i> Lc1542	<i>Lactobacillus paracasei</i> Lc1542
<i>Lactobacillus casei</i> LcY	<i>Lactobacillus paracasei</i> LcY
<i>Lactobacillus casei</i> Lpc-37 (Contig no.150)	<i>Lactobacillus paracasei</i> Lpc-37
<i>Lactobacillus casei</i> M36	<i>Lactobacillus paracasei</i> M36
<i>Lactobacillus casei</i> MJA12	<i>Lactobacillus paracasei</i> MJA12
<i>Lactobacillus casei</i> T71499	<i>Lactobacillus paracasei</i> T71499
<i>Lactobacillus casei</i> UCD174	<i>Lactobacillus paracasei</i> UCD174
<i>Lactobacillus casei</i> UW1	<i>Lactobacillus paracasei</i> UW1
<i>Lactobacillus casei</i> UW4 (Contig no. 122)	<i>Lactobacillus paracasei</i> UW4
<i>Lactobacillus casei</i> UW4 (Contig no. 144)	<i>Lactobacillus paracasei</i> UW4
<i>Lactobacillus casei</i> W14	<i>Lactobacillus paracasei</i> W14
<i>Lactobacillus casei</i> W16	<i>Lactobacillus paracasei</i> W16
<i>Lactobacillus casei</i> W56	<i>Lactobacillus paracasei</i> W56
<i>Lactobacillus casei</i> Z11	<i>Lactobacillus paracasei</i> Z11
<i>Lactobacillus casei</i> Zhang	<i>Lactobacillus paracasei</i> Zhang

The Genome sizes of *L. paracasei* strains ($n = 75$) ranged between 2.36 and 3.25 Mb, with an average value of 2.97 Mb. The GC contents varied from 46.05% to 46.97%, with a mean value of 46.30%. Additionally, N50 (base pairs) and L50 (number) values are from 2658 to 3,112,081 and 1 to 292, respectively (Supplementary Table S4). Besides these, other genomic feature details and source of the organisms are listed in Supplementary Table S4.

The reassembled genome of the *L. paracasei* Lbs2 strain was re-annotated with Prokka, resulting in a total of 2380 genes, including 2308 protein-coding, 20 t-RNA, 3 r-RNA, and 49 misc-RNA genes (Figure 1).

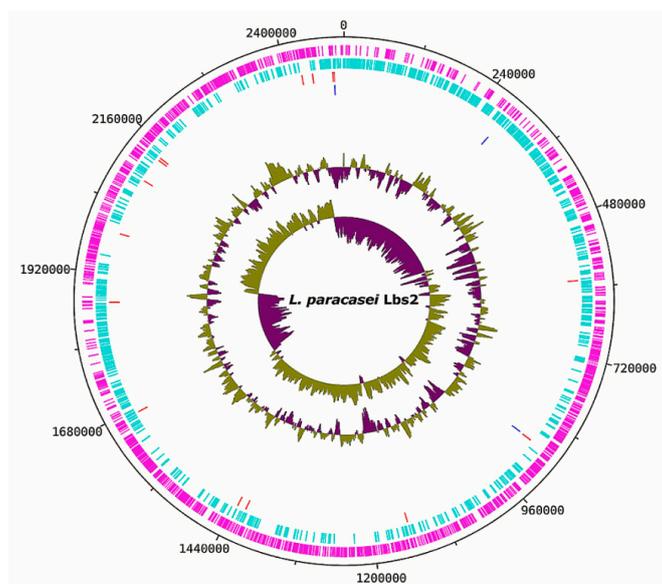


Figure 1. Circular map of the *L. paracasei* Lbs2 genome. Labeling from outside to inside of the circle, each ring carries information of the genome: coding sequences (CDSs) on the forward strand (magenta); CDSs on the reverse strand (cyan); tRNA genes (red); rRNA genes (blue); GC content; GC skew.

Further, annotation by the RAST server showed that a major proportion of protein coding groups are “Carbohydrates” (26.71%; $n = 105$), “DNA Metabolism” (14.50%; $n = 57$), and “Cofactors, Vitamins, Prosthetic Groups, Pigments” (12.97%; $n = 51$) (Supplementary Figure S3 and Supplementary Table S5). Notably, undigestible carbohydrates are the prime source of energy for the gut microbes. These undigested carbohydrates are originating from the plant sources that are defiant to enzymatic degradation and are not absorbed in the upper part of the intestinal tract. Such dietary compounds reach the large intestine, where they are get hydrolyzed by a limited range of organisms [47]. This could be one of the reasons for the presence of a higher percentage of carbohydrate metabolism-related genes in gut isolate *L. paracasei* Lbs2. Besides this, many reports suggested that carbohydrate metabolism in Lactobacilli is also crucial for niche adaptation or survival [48]. The distribution of functional COG categories across the *L. paracasei* strains ($n = 75$) are illustrated in a heatmap (Figure 2), and the pathway annotation of genes specific to *L. paracasei* strains are presented in Supplementary Table S6. Importantly, the presence of the higher number of ABC transporter gene family members among the *L. paracasei* strains ($n = 75$) may be responsible for the regulation of gene expression and interaction with the environment (Supplementary Table S6).

3.2. Pan/Core-Genome and Unique Genes Analysis

The pan/core-genome of *Lactobacillus paracasei* strains ($n = 75$) was calculated by Roary software. The pan-genome of 75 *Lactobacillus paracasei* strains hold 15,945 gene families (215,232 genes), while the core genome contained 243 gene families ($75 \times 243 = 18,225$ genes) (Supplementary Table S7 and Figure 3A). Roary pipeline uses an in-built CD-HIT clustering algorithm, which clustered the protein sequences with a sequence identity of 100% and a matching length of 100%. If one sequence is identical along its entire length to other orthologous counterparts in other species/strains, then it is said to be a core gene. The core genes include mostly ribosomal proteins and several housekeeping genes. In-addition, the COG functional annotation of these core genes showed that the majority of them belong to “translation, ribosomal structure and biogenesis (J)” and “transcription (K)” (Supplementary Table S7 and Figure 3B).

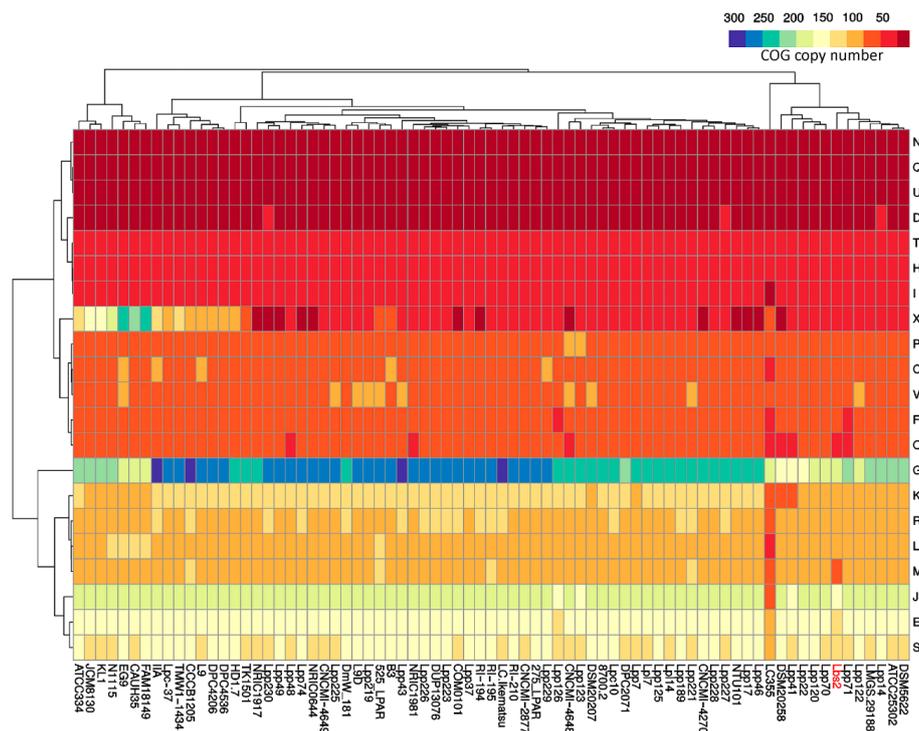


Figure 2. Cluster of Orthologous Groups (COG) frequency heatmap based on hierarchical clustering. The horizontal axis depicts functional COG categories, and the vertical axis represents 75 *L. paracasei* strains. Genome of interest ‘Lbs2’ is marked as red.

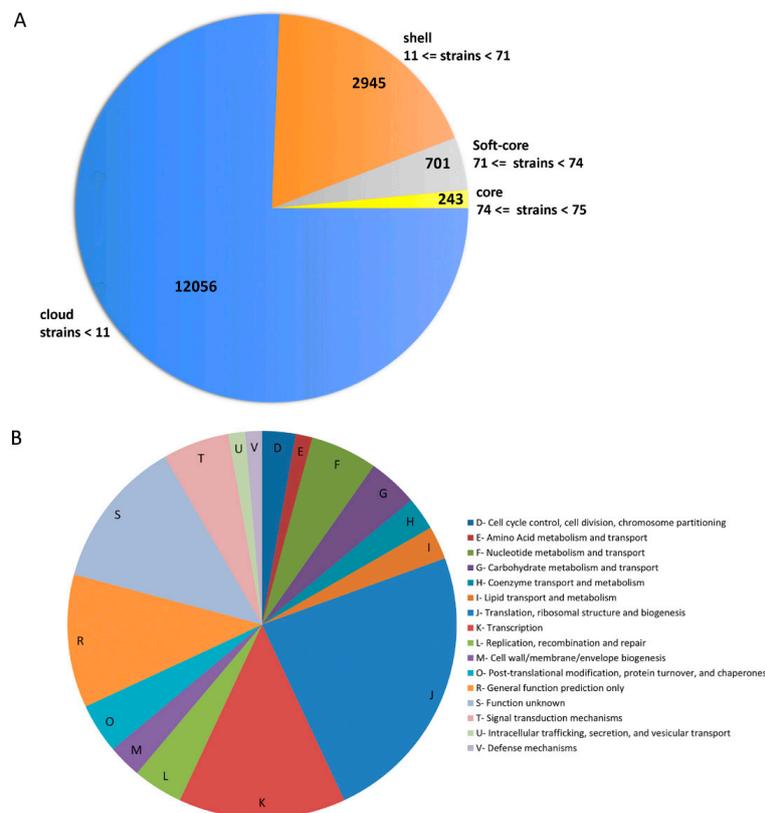


Figure 3. Pan-genome analysis of 75 *L. paracasei* strains. (A) Pie-chart representing core and accessory genes distribution. (B) Pie-chart representing the distribution of COG categories of 243 core-gene families.

Moreover, Ka/Ks ratios of all the core genes are found to be less than 1 (Figure 4 and Supplementary Table S8), suggesting strong purifying selective pressure (negative selection), which may reduce the genomic decay process.

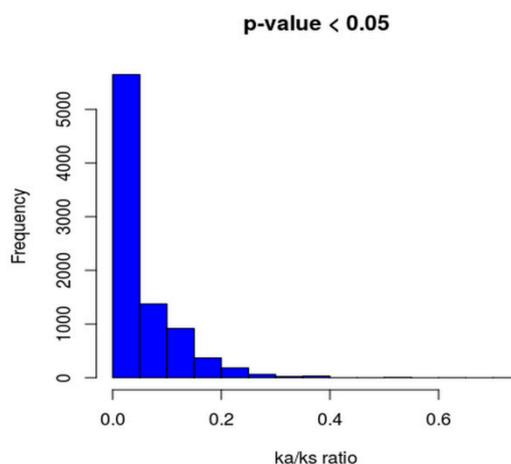


Figure 4. Histogram illustrating Ka/Ks ratios of each core gene.

Depending on the niches, we further analyzed the unique genes across the 75 strains of *L. paracasei*. The results showed that highest number ($n = 251$) of unique genes are associated with the Lpp126 strain while lowest number ($n = 4$) is found in the NRIC1917 strain. Furthermore, the *L. paracasei* Lpc-37 strain contained 13 unique genes. The reassembled genome, i.e., *L. paracasei* Lbs2, carries 54 genes. Interestingly, genes belonging to thiamin biosynthetic pathways (*thiE_2*, *thiM_2*) and the internalin gene family (*inlJ_1*) are part of the unique genes ($n = 54$) present in the Lbs2 strain (Supplementary Table S7).

3.3. Whole Genome Phylogenetic Analysis Reveals Closeness between *L. paracasei* Lbs2 and *L. paracasei* subsp. *Tolerans* DSM20258 Strain

Considering the conserved nature of the 16S rRNA sequence, it is broadly used over other functional genes for the identification of the *Lactobacillus* species [49]. The conventional mode of phylogenetic tree construction from the 16S rRNA gene is usually unstable and needs the inclusion of functional genes to enhance resolution at the strain level [50]. Thus, we have built a highly robust maximum-likelihood phylogenetic tree of the *L. paracasei* strains ($n = 75$) based on 243 conserved single-copy marker genes. Our analysis revealed that *L. paracasei* Lbs2 (reassembled genome), *L. paracasei* subsp. *tolerans* DSM20258, and *L. paracasei* Lpc-37 (higher degree of genome completeness) strains are originated from a common node, but only the former two share close proximities (Figure 5).

For this reason, strain DSM20258 is employed for further analysis along with the reassembled genome (*L. paracasei* Lbs2). Interestingly, hierarchical clustering of COG categories frequency revealed that both the Lbs2 and DSM20258 strains are in the same cluster (Figure 2).

3.4. Horizontal Gene Transfer Analysis Indicates *L. paracasei* Lbs2 Strain Acquired Important Niche-Specific Genes

Horizontal gene transfer (HGT) can be interpreted as the gaining of genetic material from other organisms without being its offspring. This event is an important force for the bacterial genome evolution. In the Lbs2 strain, we identified five genomic islands (GIs) using the IslandViewer4 tool (Supplementary Figure S4). These GIs hold a total of 205 (193 without duplicate) genes and most of them are putative (Supplementary Table S9). The lengths of these GIs range from 4671 to 163,188 bp. Interestingly, thiamin biosynthetic genes (*thiE_2*, *thiM_2*) are found in the largest GI (GI 4), which are also unique to the Lbs2 strain. GI 2 and GI 5 constituted only hypothetical genes, whereas GI

1 contained important genes like *inlJ_1*, *dps*, and GI 3 with *recX* gene (Supplementary Table S9). Further, we carried out a blast of horizontally transferred genes against the AMR (antimicrobial resistance) database [51] and found that none of the genes codes for an antimicrobial resistance gene. But blasting of the remaining genes ($n = 2187$) against the AMR database indicates that only three genes (e.g., two beta sub-unit of RNA polymerase and one Elongating factor Tu) are potentially antimicrobial resistant (Supplementary Table S7). This indicates that the antimicrobial resistances genes present in *Lbs2* are not acquired by horizontal transfer.

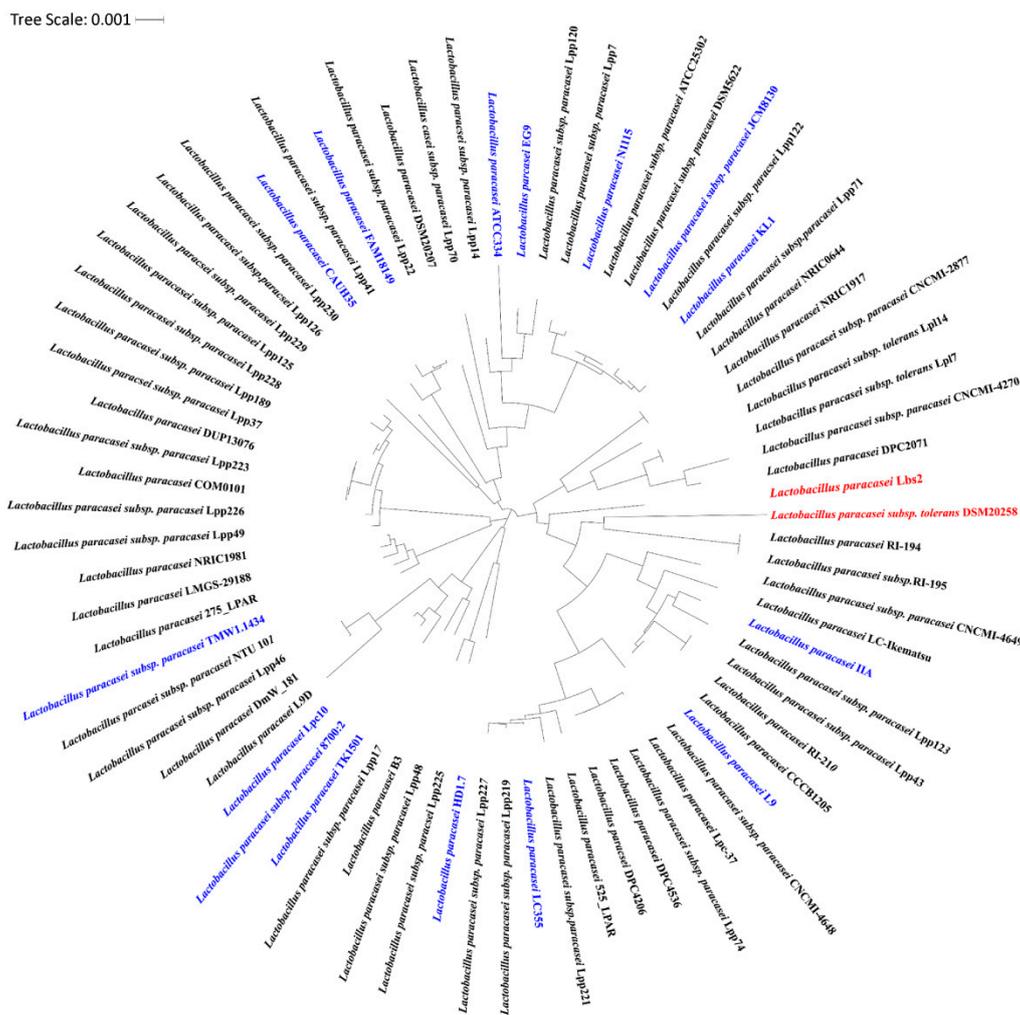


Figure 5. Whole genome phylogenetic tree of *L. paracasei* strains was inferred using the maximum likelihood method. The tree was built on the basis of the core-genome and it is presented as a cladogram. Complete genomes are marked as blue, while the genome of interest and its closest one are marked as red.

The pan-genome of *L. paracasei* strains ($n = 75$) contains several multidrug resistance (MDR) and antimicrobial resistance (AMR) genes (Supplementary Table S7), which may be intrinsic or acquired via a horizontal gene transfer mechanism. Intrinsic resistance is a chromosomally encoded inherent feature in bacteria and is not movable. The best example of intrinsic resistance in Lactobacilli is resistance to vancomycin [52]. Vancomycin blocks bacterial growth by affecting peptidoglycan synthesis, which is an essential component for the cell wall of gram-positive bacteria. Interestingly, pathway analysis throughout *L. paracasei* strains ($n = 75$) revealed the abundance of vancomycin resistance genes (Supplementary Table S6). Also, a study on *L. rhamnosus* suggests that the bacteria with vancomycin resistance did not transfer vancomycin resistance genes to recipients [52]. Very early reports on

the presence of vancomycin resistance genes in *Lactobacillus rhamnosus* suggests that the genes have diverged completely from the vancomycin resistance genes of Enterococcal strains [53]. There are already a large number of studies pointing towards the intrinsic nature of vancomycin resistance to *Lactobacillus* species. Generally, this antibiotic binds to the D-alanine residue of the peptidoglycan residue and blocks the peptidoglycan biosynthesis. In the case of Lactobacilli, the mechanism of resistance involves the substitution of the D-alanine residue with either a D-lactate or D-serine, preventing the binding of the antibiotic [54]. Vancomycin is generally used to treat a number of gram positive bacterial infections in the gut. Resistance to vancomycin may have been an ancient event and may have originated in Lactobacilli itself as a parallel event rather than an acquired event from the environment.

On the contrary, acquired resistance due to horizontal gene transfer poses a threat to nonpathogenic bacteria. The exchange of virulence genes from commensals to Lactobacilli and resistance genes from Lactobacilli to intestinal commensals inside the colon can totally change the genotypic profile of commensals and Lactobacilli [52].

3.5. Extracellular Properties of the *L. paracasei* Strains

In Lactobacilli, the LPXTG motif-containing extracellular proteins are anchored to the cell wall by sortase enzymes [55]. These proteins play important roles in adhesion and colonization. Several experiments have suggested that the act of sortases and their substrates is vital for deciphering various probiotic modes of action [6]. In the pan-genome ($n = 75$) of *L. paracasei*, we identified a total of 415 LPXTG motif-containing proteins and 141 sortase enzymes (Supplementary Figure S5). The number of the LPXTG motif and sortase enzymes varies greatly among the *L. paracasei* strains (0–12 for LPXTG and 0–4 for sortase). The highest numbers of LPXTG proteins ($n = 12$) are found in the 525_LPAR strain. Interestingly, both CNCMI-4648 (genome completeness = 92.45% as per CheckM analysis) and Lpp70 (genome completeness = 99.27% as per CheckM analysis) strains lacking LPXTG motif-containing proteins also lacked sortase protein, indicating a possible gene loss event. Additionally, LPXTG motifs and sortase enzymes are also identified in strains Lpc-37 (9 vs. 3), Lbs2 (6 vs. 2), and DSM20258 (2 vs. 1), respectively.

Identification of these proteins in Lactobacilli provides information about their key roles in resolving nutrient uptake through proteinase P and with positive probiotic traits, such as mucus barrier function, adhesion, and immune signaling [56]. Moreover, sortase expression signals in Lactobacilli have been utilized to develop gastrointestinal tract targeted oral vaccines [56].

3.6. Various Stress Factors Characterized Across the *L. paracasei* Strains

Lactobacilli have faced several environmental stress factors, such as low pH, bile salts, and oxidative and osmotic stress, during their transit through the diverse habitat. Our analysis has revealed that the number of genes associated with oxidative stress, osmotic stress, and salt stress varies greatly across the *L. paracasei* strains ($n = 75$), while the rest of the stress-related gene numbers remains fairly constant irrespective of their environment (Supplementary Figure S6 and Supplementary Table S10). In this study, we have also found the highest number of genes linked with oxidative stress ($n = 11,416$) compared to other stress linked genes, such as osmotic ($n = 6359$), salt ($n = 3721$), nitrosative ($n = 617$), heat ($n = 155$), DNA damage ($n = 773$), cold ($n = 0$), acid ($n = 0$), and bile ($n = 76$) stresses across the all strains of *L. paracasei*. A total of 330 stress linked genes are identified in the Lpc-37 strain alone, whereas the Lbs2 and DSM20258 strain contained 260 and 257 stress associated genes each. These genes can change the activity of the general metabolism, membrane components, and transporter systems of the cell in a hostile environment [57]. Glycine betaine/proline transport system (ATP-binding protein) and glycine betaine/carnitine transport (ATP-binding protein) encoding genes are found to be predominant among the osmotic stress related genes, which are thought to be involved in the accumulation of glycine, betaine, and carnitine in response to increased external osmolarity [58]. Interestingly, the predominant oxidative stress encoding genes are annotated as calcium transporting

ATPase followed by a hydrogen peroxide-inducible gene activator. It is proposed that the hydrogen peroxide-inducible genes activator has a crucial role in hydrogen peroxide scavenging for the repair of oxidative damage [59]. In addition, an important enzyme, bile salt hydrolase related to adaptation to the gut environment is identified among the *L. paracasei* strains except for the Lbs2 strain that is isolated from the gut (Supplementary Table S10). Notably, the Lbs2 strain was cultivated for many generations outside of its native environment and may cause the loss of the gene encoding the bile salt hydrolase (*bsh*) enzyme. Apart from that, as the Lbs2 draft assembly covered 91.28% of the genome, this could also be the reason for not capturing the *bsh* gene. Importantly, the presence of this enzyme in plant isolates such as Lpl14, Lpp189, Lpp46, Lpp49, and CNCMI-2877 indicates that these strains can be good probiotic candidates (Supplementary Table S4). Moreover, the heatmap of stress linked genes has shown a closer relationship between Lbs2 and DSM20258 strains, as they are part of the same cluster (Supplementary Figure S6).

3.7. Bacteriocins Identified Among the *Lactobacillus paracasei* Strains

Bacteriocins are small cationic peptides with a key function like quorum sensing. In the genomes of *L. paracasei* strains ($n = 75$), a total of 191 bacteriocins are predicted with BAGEL4 (Supplementary Figure S7A and Supplementary Table S10). The highest numbers of bacteriocins are found to be present in 275_LPAR, FAM 18149, Lpp 122, and Lpp 219 strains, but absent in DSM20258, Lpp48, and Lpp70 strains. Moreover, a lone copy of bacteriocin is found in Lbs2, Lpc-37, CAUH 35, Lpp 189, Lpp 228, Lpp 49, and TMW1.1434 strains (Supplementary Figure S7A). Based on their structural properties, bacteriocins from Lactobacilli are divided into three major classes. Class I bacteriocins, the lantibiotics, are small peptides that undergo extensive posttranslational modifications. Class II bacteriocins are unmodified, heat-stable peptides. Class III bacteriocins are the least characterized to date [60]. Bacteriocins of the *L. paracasei* Lbs2 strain, identified as Enterolysin_A, belongs to Class III types of bacteriocins (Supplementary Figure S7B). The sequence analysis of Enterolysin A suggested that this bacteriocin consists of two separate domains, an N-terminal catalytic domain and a C-terminal substrate recognition domain [58]. This Enterolysin_A played a crucial role in cell-wall degradation of pathogens [60]. Among other strains of Lactobacilli, the class II bacteriocins are pretty common. Examples of this class of bacteriocins predicted in other members of Lactobacilli are LSEI_2386, LSEI_2163, Thermophilin_A, and Carnocin_CP52 (Supplementary Table S10). Already, there are several reports on the structural characterization of different bacteriocins [61]. It has also been reported that Class II bacteriocins might contain peptides with a double-glycine leader and hypothesized that the existence of a disulfide bridge among these peptides plays an important role in antibacterial activity [62]. In recent times, the majority of studies have focused on bacteriocins producing probiotics, which can inhibit the growth of gut pathogens. It is thought that the production of bacteriocins could provide probiotic functionality in three different ways. Initially, it may act as colonizing peptides, building the dominance of a producer into an already occupied habitat [63]. Secondly, bacteriocin may play the role of killing peptides, where it is directly inhibiting competing strains or pathogens [64]. Finally, bacteriocins may serve as signaling peptides, either signaling other bacteria via quorum sensing and signaling cells of the host immune system [65]. It has also been reported that bacteriocins can be deployed as anticancer agents, either through their impact on cancerous cells or the suppression of bacteria associated with the initiation of disease [66].

3.8. A Wide Distribution of Mobile-Genetic Elements and CRISPR-Cas Systems

Using PHASTER, a total of 276 prophages were detected among the 75 *L. paracasei* strains. The maximum number of prophage regions ($n = 10$) were observed in the strain EG9, whereas only one prophage region was detected in 13 strains (275_LPAR, B3, CNCMI-2877, CNCMI-4648, Lpp46, Lpp49, Lpp70, Lpp123, Lpp126, Lpp189, Lpp221, Lpp227, Lpp228, and Lbs2). In Lbs2, the observed prophage region is highly related to *Lactobacillus* phages and encodes seven proteins, but was found to be incomplete (5.4 Kb). In the strain DSM20258, two incomplete prophage regions of length 16.4

and 18.6 Kb, respectively, coding 10 proteins each, were observed, which are also highly related to *Lactobacillus* phages. Moreover, strain Lpc-37 has an incomplete prophage (14 Kb) and two putative prophages (25.8 and 41.6 Kb) regions (Supplementary Table S11). It is believed that the presence of prophages in Lactobacilli genomes may protect them from superinfection by other phages or plasmid [67].

Prokaryotic genomes possess a large number of CRISPR loci, which play a vital role in controlling horizontal gene transfer [68]. It is a well-known fact that some bacteria have gained the CRISPR-cas system as a defense system against phage invasion. A total of 238 CRISPR loci across the *L. paracasei* genomes ($n = 75$) were predicted using a stand-alone tool, CRISPR finder (Supplementary Figure S8A). CNCMI-2877, Lpp226, Lpp229, and Lpp41 strains are lacking CRISPR loci, evidently because these genomes are incomplete. The DmW_181 strain carries the maximum number of CRISPR loci despite having a very fragmented assembly with 127 scaffolds. In addition, the number of CRISPR loci predicted in Lbs2, DSM20258, and Lpc-37 strains are two, one, and three, respectively. In Supplementary Figure S8B, CRISPR loci of *L. paracasei* Lbs2 are depicted.

We have also identified IS elements using the ISSaga platform, which may contribute to bacterial genome evolution. From *L. paracasei* Lbs2, *L. paracasei* subsp. *tolerans* DSM20258, and *L. paracasei* Lpc-37 genomes, a total of 10, 5, and 98 IS elements are predicted (Supplementary Table S12). Compared to the 74 genomes of *L. paracasei*, FAM18149 exhibited a higher number of IS elements that may suggest a higher potential for genome plasticity.

3.9. Broad Range Carbohydrate-Active Enzymes (CAZymes) and Carbohydrate Transporters Identified in the Pan-Genome of *Lactobacillus paracasei*

Carbohydrates are the prime source of energy for all organisms. A major fraction of genes found in the pan-genome ($n = 75$) of *L. paracasei* are associated with carbohydrate metabolism and transport; thus, we compared CAZymes encoding genes across the strains. Currently, a total of 153 GH (glycoside hydrolases), 106 GT (glycosyl transferase), 27 PL (polysaccharide lyase), 83 CBM (Carbohydrate-Binding Module), 16 CE (carbohydrate esterase), 15 AA (auxiliary activities) families, and 42 GH13, 37 GH43 subfamilies are present in the CAZy database (<http://www.cazy.org>).

In this study, 32 GHs, 17 GTs, 2 PLs, 7 CBMs, 7 CEs, 4 AAs families, and 7 GH13, 1 GH43 subfamilies are identified in the *L. paracasei* strains (Supplementary Figure S9 and Supplementary Table S13). Our analysis also revealed that the *L. paracasei* strains carbohydrate-activating enzyme numbers ranged from 72 (present in the Lpp123 strain) to 115 (present in the Lpc-37 and NRIC0644 strains). Among the 115 enzymes of the Lpc-37 strain, 46 are identified as GHs, 35 as GTs, 16 as CEs, 11 as CBMs, 5 as AAs, and 2 as PLs. Compared to the 115 enzymes of the *L. paracasei* Lpc-37 strain, the reassembled genome (*L. paracasei* Lbs2) and its closest one (*L. paracasei* subsp. *tolerans* DSM20258) contained 86 (37 are identified as GHs, 22 as GTs, 14 as CEs, 6 as CBMs, 5 as AAs, and 2 as PLs.) and 73 (24 are identified as GHs, 24 as GTs, 15 as CEs, 6 as CBMs, 4 as AAs, and 0 as PLs.) enzymes, respectively. To transform undigested carbohydrates, present in the gastrointestinal tract or in the environment, several glycosyl hydrolase family (GH) enzymes are being used by Lactobacilli. In *L. paracasei* Lbs2 β -xylosidase and α -L-iduronidase (GH39), enzymes are unique with respect to Lpc-37 and DSM20258 strains, indicating that the Lbs2 strain probably needs these enzymes in its ecological niche.

Interestingly, cellulose synthase (GT2), a key enzyme for cellulose biosynthesis, is found to be predominant across the 75 strains of *L. paracasei*, which could hoard cellulose on the cell wall surface as an extracellular matrix for cell adhesion and biofilm formation to defend itself from the surrounding environment [69–71]. Noticeably, Lactobacilli uses glycogen as carbohydrate storage forms, and it has been reported that they are synthesizing glycogen to engage with more diverse habitats [72]. Glycogen synthase (GT5) and glycogen phosphorylase (GT35) enzymes are involved in glycogen synthesis. The current study showed that the GT5 and GT35 families encoding gene numbers remain the same

among the Lpc-37, Lbs2, and DSM20258 strains and could probably be required for adaptation in diverse habitats.

Furthermore, potential carbohydrate transporters are identified in *L. paracasei* strains using the GO database, which ranged from 99 (present in the Lpp126 strain) to 149 (present in the IIA strain). *L. paracasei* Lbs2, *L. paracasei* Lpc-37, and *L. paracasei* subsp. *tolerans* DSM20258 contain 114, 106, and 142 potential carbohydrate transporters, respectively (Supplementary Figure S10 and Supplementary Table S14).

3.10. In-Depth Comparative Analysis of *L. paracasei* Lbs2 Against *L. paracasei* subsp. *tolerans* DSM20258 and *L. paracasei* Lpc-37

To further examine the technological and probiotic properties of *L. paracasei* Lbs2, we carried out a comparative genomic analysis against *L. paracasei* subsp. *tolerans* DSM20258, and *L. paracasei* Lpc-37. *L. paracasei* Lbs2 isolated from a healthy human gut (north Indian) has a genome size of 2.50 Mb. The genome size of *L. paracasei* subsp. *tolerans* DSM20258 and *L. paracasei* Lpc-37 are 2.36 Mb and 3.16 Mb, respectively. The detail of genomic features among these three strains are elaborated in Table 2.

Table 2. Genome assembly statistics of the three *Lactobacillus paracasei* strains (Lpc-37, Lbs2, DSM20258).

Features	Strain		
	Lpc-37	Lbs2	DSM20258
Source	Microbial food product	Human Gut	Not available
Genome Status	Draft	Draft	Draft
Accession Number	NOKL00000000.1	JPKN00000000.3	AYYJ00000000.1
N50 (bp)	3,112,081	10,992	14,516
L50	1	68	49
Completeness (%)	100	91.28	97.19
Contamination (%)	0	0.87	0
Size (Mb)	3.16	2.50	2.36
GC%	46.33	46.97	46.44
Genes	3125	2380	2424
Proteins	3010	2308	2339
t-RNA	59	20	37
r-RNA	15	3	2
Other-RNA	41	49	46

Our analysis using OrthoVenn also unveiled that the number of core genes among these three strains is 1678 (Supplementary Figure S11A), higher than that computed for the pan-genome of *L. paracasei* (Figure 3A). The distribution of the core COG functional categories are depicted in Supplementary Figure S11B; the majority of them are associated with translation, ribosomal structure and biogenesis (J), amino acid transport and metabolism (E), carbohydrate transport and metabolism (G), and function unknown (S) (Supplementary Table S15). Additionally, singletons of the three strains allocated in all COG categories, most of them affiliated to carbohydrate transport and metabolism (G), mobilome: prophages and transposase (X) and defense mechanism (V) (Supplementary Figure S12 and Supplementary Table S15). It could be suggested that the high prevalence of proteins in the G (Carbohydrate transport and Metabolism) COG category have a direct effect on the niche diversity in which organisms can grow. On the contrary, mobilome: prophages and transposase can create genome plasticity.

Furthermore, whole genome alignments based on progressive MAUVE among the three strains lack of extensive synteny (Supplementary Figure S13) corroborates with the absence of thiamin biosynthetic gene cluster in the DSM20258 strain (Supplementary Figure S14). However, the glycyl/betain transporter gene cluster has shown perfect synteny among these strains (Supplementary Figure S15A).

3.11. Evaluating the Technological and Probiotic Traits of *L. paracasei* Lbs2

Probiotic properties of *L. paracasei* subsp. *tolerans* DSM20258 have been extensively studied in the past [1]. The genome of the DSM20258 strain contains important probiotic traits, like bile salt hydrolase. Surprisingly, this trait was absent from the gut isolate *L. paracasei* Lbs2, which may be due to genome incompleteness. In addition, gene encoding bacteriocin were not present in the DSM20258 genome, while the Lbs2 genome contained enterolysin_A like bacteriocin. Furthermore, sortase-dependent cell surface proteins in the DSM 20258 strain, which may interact with the host, are also found to be present in the Lbs2 strain.

We then examined the technological aspects of gut isolate *L. paracasei* Lbs2, keeping in mind that a human being cannot synthesize most of the vitamins and for this it needs to be outsourced. Probiotic bacteria, located in the human gut, such as Lactobacilli, can *de novo* synthesize and provide vitamins to the human body. Gut microbiota in humans, capable of synthesizing vitamin K, and most of the water soluble B vitamins, such as pyridoxine, folates, riboflavin, cobalamin, and thiamin. Among the water soluble B vitamins, thiamin (vitamin B1) as thiamine pyrophosphate (TPP), played a crucial role in host energy metabolism since it acts as a co-factor for major metabolic pathways, such as the pentose phosphate pathway, glycolysis, and Kreb's cycle (Figure 6A). The pentose phosphate pathway is needed for steroids, nucleic acids, fatty acids, and the aromatic amino acid biosynthesis. These products from the pentose phosphate pathway are used as precursors of different neurotransmitters and other bioactive compounds vital for brain function [73]. Interestingly, the thiamin biosynthetic gene cluster along with the TPP riboswitch is found to be present in *L. paracasei* Lbs2 and *L. paracasei* Lpc-37 genomes but is absent in DSM20258 (Supplementary Figure S14). In this study, we have shown the plausible role of these genes in thiamin biosynthesis (Figure 6B).

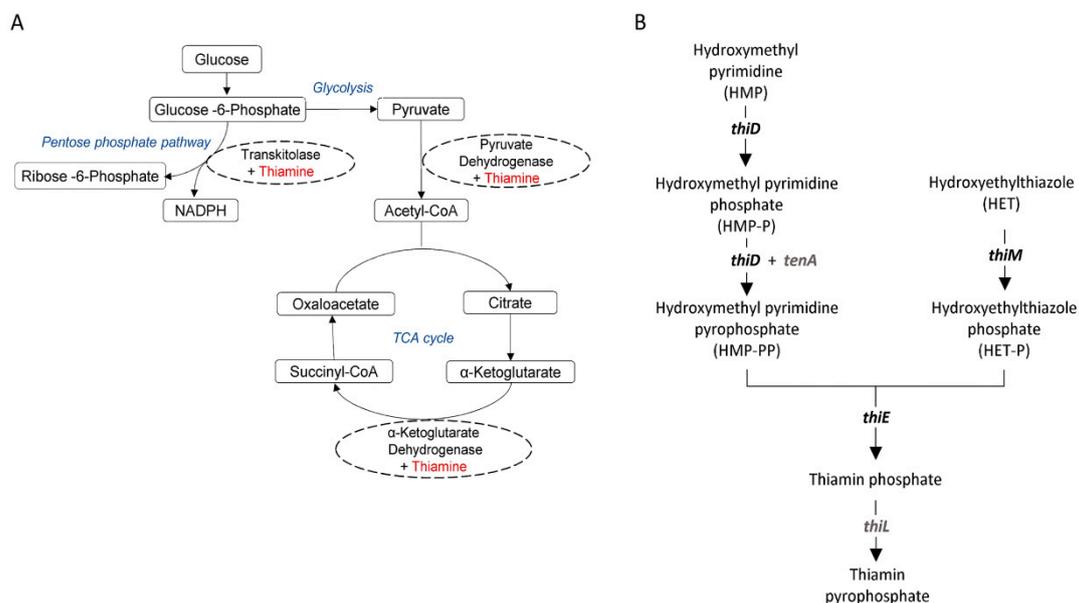


Figure 6. The schematic illustration of pathways. **(A)** The role of thiamin (marked as red) as a co-factor in major metabolic pathways (marked as blue). **(B)** Genes involved in thiamin biosynthesis.

Surprisingly, from the pan-genome analysis, it was found that two thiamin biosynthesis related genes (*thiE_2* and *thiM_2*) are unique to the *L. paracasei* Lbs2 genome (Supplementary Table S7). Hence, the presence of unique thiamin biosynthesis genes in the *L. paracasei* Lbs2 genome could be one of the reasons for probiotic adaptation leading to improved host energy metabolism.

4. Conclusions

In this study, we reassembled and analyzed gut isolated Lbs2 strains earlier identified as *L. casei* species and placed it under the *L. paracasei* species, after careful genomic analysis. Our analysis was based on pair wise genome distance (ANI and AAI) calculations, which also indicated that many strains of *L. casei* were classified wrongly in the NCBI genome repository, and this needed to be reclassified as *L. paracasei*.

Reclassification followed by the pan-genome analysis of *L. paracasei* strains revealed that the Lbs2 strain holds a small core-proteome; most of them are ribosomal proteins. Furthermore, phylogeny based on core-genome indicates that the *L. paracasei* subsp. *tolerans* DSM20258 strain is more closely related to *L. paracasei* Lbs2 as compared to other strains. Interestingly, probiotic features of the DSM20258 strain are established in earlier studies [1], which make it easier for us to find out the probiotic traits, such as sortase-dependent cell surface proteins, bacteriocins, etc., present in the Lbs2 strain. However, important probiotic traits, like bile salt hydrolase, are missing from the Lbs2 strain compared to DSM20258; this may be due to genome gaps. Finally, we have also identified a thiamin biosynthetic gene cluster in the Lbs2 strain that is thought to be involved in enhancing host energy metabolism. Surprisingly, *thiE_2* and *thiM_2* genes are found to be unique in the Lbs2 strain. These unique genes are also located in the genome islands on the Lbs2 strain.

In summary, our results indicate further study is required with other members of the *L. casei* group, like *L. rhamnosus* and *L. casei*, for potential reclassification among closely related members of the *L. casei* group and to better understand the more technical aspects lying behind the host adaptation.

Supplementary Materials: Supplementary materials can be found at <http://www.mdpi.com/2076-2607/7/11/487/s1>.

Author Contributions: S.T. conceived the project. S.G. and S.T. designed the project. S.G., A.N.S., M.M., and S.B. performed genome analysis. S.G. and S.T. wrote the manuscript.

Funding: S.T. would like to acknowledge the Department of Biotechnology (DBT) Govt. of India—Ramalingaswamy fellowship for supporting this work. S.G. and M.M. would like to acknowledge DBT and CSIR respectively for providing fellowship to pursue a Ph.D. program.

Acknowledgments: The author would like to acknowledge the Computational biology and Genomics Lab members of CSIR-IICB for valuable discussion.

Conflicts of Interest: The authors declare that they have no conflict of interests. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Sun, Z.; Harris, H.M.; McCann, A.; Guo, C.; Argimon, S.; Zhang, W.; Yang, X.; Jeffery, I.B.; Cooney, J.C.; Kagawa, T.F.; et al. Expanding the biotechnology potential of lactobacilli through comparative genomics of 213 strains and associated genera. *Nat. Commun.* **2015**, *6*, 8322. [[CrossRef](#)] [[PubMed](#)]
2. Smokvina, T.; Wels, M.; Polka, J.; Chervaux, C.; Brisse, S.; Boekhorst, J.; van HylckamaVlieg, J.E.; Siezen, R.J. Lactobacillus paracasei comparative genomics: Towards species pan-genome definition and exploitation of diversity. *PLoS ONE* **2013**, *8*, e68731. [[CrossRef](#)] [[PubMed](#)]
3. Wuyts, S.; Wittouck, S.; De Boeck, I.; Allonsius, C.N.; Pasolli, E.; Segata, N.; Lebeer, S. Large-Scale Phylogenomics of the Lactobacillus casei Group Highlights Taxonomic Inconsistencies and Reveals Novel Clade-Associated Features. *MSystems* **2017**, *2*. [[CrossRef](#)] [[PubMed](#)]
4. Salvetti, E.; Harris, H.M.B.; Felis, G.E.; O'Toole, P.W. Comparative Genomics of the Genus Lactobacillus Reveals Robust Phylogroups That Provide the Basis for Reclassification. *Appl. Env. Microbiol.* **2018**, *84*. [[CrossRef](#)]
5. Huang, C.H.; Li, S.W.; Huang, L.; Watanabe, K. Identification and Classification for the Lactobacillus casei Group. *Front. Microbiol.* **2018**, *9*, 1974. [[CrossRef](#)] [[PubMed](#)]
6. Call, E.K.; Goh, Y.J.; Selle, K.; Klaenhammer, T.R.; O'Flaherty, S. Sortase-deficient lactobacilli: Effect on immunomodulation and gut retention. *Microbiology* **2015**, *161*, 311–321. [[CrossRef](#)] [[PubMed](#)]

7. Boekhorst, J.; de Been, M.W.; Kleerebezem, M.; Siezen, R.J. Genome-wide detection and analysis of cell wall-bound proteins with LPxTG-like sorting motifs. *J. Bacteriol.* **2005**, *187*, 4928–4934. [[CrossRef](#)] [[PubMed](#)]
8. Munoz-Provencio, D.; Rodriguez-Diaz, J.; Collado, M.C.; Langella, P.; Bermudez-Humaran, L.G.; Monedero, V. Functional analysis of the *Lactobacillus casei* BL23 sortases. *Appl. Env. Microbiol.* **2012**, *78*, 8684–8693. [[CrossRef](#)]
9. Dicks, L.M.T.; Dreyer, L.; Smith, C.; van Staden, A.D. A Review: The Fate of Bacteriocins in the Human Gastro-Intestinal Tract: Do They Cross the Gut-Blood Barrier? *Front. Microbiol.* **2018**, *9*, 2297. [[CrossRef](#)]
10. Cotter, P.D.; Hill, C.; Ross, R.P. Bacteriocins: Developing innate immunity for food. *Nat. Rev. Microbiol.* **2005**, *3*, 777–788. [[CrossRef](#)]
11. Corr, S.C.; Li, Y.; Riedel, C.U.; O'Toole, P.W.; Hill, C.; Gahan, C.G. Bacteriocin production as a mechanism for the anti-infective activity of *Lactobacillus salivarius* UCC118. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 7617–7621. [[CrossRef](#)] [[PubMed](#)]
12. De Vos, W.M.; Vaughan, E.E. Genetics of lactose utilization in lactic acid bacteria. *FEMS Microbiol. Rev.* **1994**, *15*, 217–237. [[CrossRef](#)] [[PubMed](#)]
13. Cai, H.; Rodriguez, B.T.; Zhang, W.; Broadbent, J.R.; Steele, J.L. Genotypic and phenotypic characterization of *Lactobacillus casei* strains isolated from different ecological niches suggests frequent recombination and niche specificity. *Microbiology* **2007**, *153*, 2655–2665. [[CrossRef](#)] [[PubMed](#)]
14. LeBlanc, J.G.; Milani, C.; de Giori, G.S.; Sesma, F.; van Sinderen, D.; Ventura, M. Bacteria as vitamin suppliers to their host: A gut microbiota perspective. *Curr. Opin. Biotechnol.* **2013**, *24*, 160–168. [[CrossRef](#)] [[PubMed](#)]
15. LeBlanc, J.G.; Chain, F.; Martin, R.; Bermudez-Humaran, L.G.; Courau, S.; Langella, P. Beneficial effects on host energy metabolism of short-chain fatty acids and vitamins produced by commensal and probiotic bacteria. *Microb. Cell Fact.* **2017**, *16*, 79. [[CrossRef](#)]
16. Bhowmick, S.; Malar, M.; Das, A.; Kumar Thakur, B.; Saha, P.; Das, S.; Rashmi, H.M.; Batish, V.K.; Grover, S.; Tripathy, S. Draft Genome Sequence of *Lactobacillus casei* Lbs2. *Genome Announc.* **2014**, *2*. [[CrossRef](#)]
17. O'Connell, J.; Schulz-Trieglaff, O.; Carlson, E.; Hims, M.M.; Gormley, N.A.; Cox, A.J. NxTrim: Optimized trimming of Illumina mate pair reads. *Bioinformatics* **2015**, *31*, 2035–2037. [[CrossRef](#)]
18. Bankevich, A.; Nurk, S.; Antipov, D.; Gurevich, A.A.; Dvorkin, M.; Kulikov, A.S.; Lesin, V.M.; Nikolenko, S.I.; Pham, S.; Pribelski, A.D.; et al. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **2012**, *19*, 455–477. [[CrossRef](#)]
19. Parks, D.H.; Imelfort, M.; Skennerton, C.T.; Hugenholtz, P.; Tyson, G.W. CheckM: Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* **2015**, *25*, 1043–1055. [[CrossRef](#)]
20. Simao, F.A.; Waterhouse, R.M.; Ioannidis, P.; Kriventseva, E.V.; Zdobnov, E.M. BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **2015**, *31*, 3210–3212. [[CrossRef](#)]
21. Seemann, T. Prokka: Rapid prokaryotic genome annotation. *Bioinformatics* **2014**, *30*, 2068–2069. [[CrossRef](#)] [[PubMed](#)]
22. Szklarczyk, D.; Franceschini, A.; Wyder, S.; Forslund, K.; Heller, D.; Huerta-Cepas, J.; Simonovic, M.; Roth, A.; Santos, A.; Tsafou, K.P.; et al. STRING v10: Protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* **2015**, *43*, D447–D452. [[CrossRef](#)] [[PubMed](#)]
23. Moriya, Y.; Itoh, M.; Okuda, S.; Yoshizawa, A.C.; Kanehisa, M. KAAS: An automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res.* **2007**, *35*, W182–W185. [[CrossRef](#)] [[PubMed](#)]
24. Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; Lipman, D.J. Basic local alignment search tool. *J. Mol. Biol.* **1990**, *215*, 403–410. [[CrossRef](#)]
25. Overbeek, R.; Olson, R.; Pusch, G.D.; Olsen, G.J.; Davis, J.J.; Disz, T.; Edwards, R.A.; Gerdes, S.; Parrello, B.; Shukla, M.; et al. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Res.* **2014**, *42*, D206–D214. [[CrossRef](#)] [[PubMed](#)]
26. Page, A.J.; Cummins, C.A.; Hunt, M.; Wong, V.K.; Reuter, S.; Holden, M.T.; Fookes, M.; Falush, D.; Keane, J.A.; Parkhill, J. Roary: Rapid large-scale prokaryote pan genome analysis. *Bioinformatics* **2015**, *31*, 3691–3693. [[CrossRef](#)]
27. Enright, A.J.; Van Dongen, S.; Ouzounis, C.A. An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res.* **2002**, *30*, 1575–1584. [[CrossRef](#)]

28. Wang, Y.; Coleman-Derr, D.; Chen, G.; Gu, Y.Q. OrthoVenn: A web server for genome wide comparison and annotation of orthologous clusters across multiple species. *Nucleic Acids Res.* **2015**, *43*, W78–W84. [[CrossRef](#)]
29. Price, M.N.; Dehal, P.S.; Arkin, A.P. FastTree: Computing large minimum evolution trees with profiles instead of a distance matrix. *Mol. Biol. Evol.* **2009**, *26*, 1641–1650. [[CrossRef](#)]
30. Katoh, K.; Misawa, K.; Kuma, K.; Miyata, T. MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **2002**, *30*, 3059–3066. [[CrossRef](#)]
31. Letunic, I.; Bork, P. Interactive tree of life (iTOL) v3: An online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* **2016**, *44*, W242–W245. [[CrossRef](#)] [[PubMed](#)]
32. Wang, D.; Zhang, Y.; Zhang, Z.; Zhu, J.; Yu, J. KaKs_Calculator 2.0: A toolkit incorporating gamma-series methods and sliding window strategies. *Genom. Proteom. Bioinform.* **2010**, *8*, 77–80. [[CrossRef](#)]
33. Zhang, Z.; Xiao, J.; Wu, J.; Zhang, H.; Liu, G.; Wang, X.; Dai, L. ParaAT: A parallel tool for constructing multiple protein-coding DNA alignments. *Biochem. Biophys. Res. Commun.* **2012**, *419*, 779–781. [[CrossRef](#)] [[PubMed](#)]
34. Darling, A.E.; Mau, B.; Perna, N.T. progressiveMauve: Multiple genome alignment with gene gain, loss and rearrangement. *PLoS ONE* **2010**, *5*, e11147. [[CrossRef](#)] [[PubMed](#)]
35. Carver, T.; Thomson, N.; Bleasby, A.; Berriman, M.; Parkhill, J. DNAPlotter: Circular and linear interactive genome visualization. *Bioinformatics* **2009**, *25*, 119–120. [[CrossRef](#)]
36. Sullivan, M.J.; Petty, N.K.; Beatson, S.A. Easyfig: A genome comparison visualizer. *Bioinformatics* **2011**, *27*, 1009–1010. [[CrossRef](#)] [[PubMed](#)]
37. Yin, Y.; Mao, X.; Yang, J.; Chen, X.; Mao, F.; Xu, Y. dbCAN: A web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res.* **2012**, *40*, W445–W451. [[CrossRef](#)]
38. Plyusnin, I.; Holm, L.; Kankainen, M. LOCP—locating pilus operons in gram-positive bacteria. *Bioinformatics* **2009**, *25*, 1187–1188. [[CrossRef](#)]
39. Van Heel, A.J.; de Jong, A.; Song, C.; Viel, J.H.; Kok, J.; Kuipers, O.P. BAGEL4: A user-friendly web server to thoroughly mine RiPPs and bacteriocins. *Nucleic Acids Res.* **2018**, *46*, W278–W281. [[CrossRef](#)]
40. Grissa, I.; Vergnaud, G.; Pourcel, C. The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. *BMC Bioinform.* **2007**, *8*, 172. [[CrossRef](#)]
41. Arndt, D.; Grant, J.R.; Marcu, A.; Sajed, T.; Pon, A.; Liang, Y.; Wishart, D.S. PHASTER: A better, faster version of the PHAST phage search tool. *Nucleic Acids Res.* **2016**, *44*, W16–W21. [[CrossRef](#)] [[PubMed](#)]
42. Varani, A.M.; Siguier, P.; Gourbeyre, E.; Charneau, V.; Chandler, M. ISSaga is an ensemble of web-based methods for high throughput identification and semi-automatic annotation of insertion sequences in prokaryotic genomes. *Genome Biol.* **2011**, *12*, R30. [[CrossRef](#)] [[PubMed](#)]
43. Ekstrom, A.; Yin, Y. ORFanFinder: Automated identification of taxonomically restricted orphan genes. *Bioinformatics* **2016**, *32*, 2053–2055. [[CrossRef](#)] [[PubMed](#)]
44. Bertelli, C.; Laird, M.R.; Williams, K.P.; Simon Fraser University Research Computing Group; Lau, B.Y.; Hoad, G.; Winsor, G.L.; Brinkman, F.S.L. IslandViewer 4: Expanded prediction of genomic islands for larger-scale datasets. *Nucleic Acids Res.* **2017**, *45*, W30–W35. [[CrossRef](#)]
45. Konstantinidis, K.T.; Tiedje, J.M. Towards a genome-based taxonomy for prokaryotes. *J. Bacteriol.* **2005**, *187*, 6258–6264. [[CrossRef](#)]
46. Goris, J.; Konstantinidis, K.T.; Klappenbach, J.A.; Coenye, T.; Vandamme, P.; Tiedje, J.M. DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. *Int. J. Syst. Evol. Microbiol.* **2007**, *57*, 81–91. [[CrossRef](#)]
47. Barrangou, R.; Azcarate-Peril, M.A.; Duong, T.; Connors, S.B.; Kelly, R.M.; Klaenhammer, T.R. Global analysis of carbohydrate utilization by *Lactobacillus acidophilus* using cDNA microarrays. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 3816–3821. [[CrossRef](#)]
48. Watson, D.; O’Connell Motherway, M.; Schoterman, M.H.; van Neerven, R.J.; Nauta, A.; van Sinderen, D. Selective carbohydrate utilization by lactobacilli and bifidobacteria. *J. Appl. Microbiol.* **2013**, *114*, 1132–1146. [[CrossRef](#)]
49. Salvetti, E.; Torriani, S.; Felis, G.E. The Genus *Lactobacillus*: A Taxonomic Update. *Probiotics Antimicrob. Proteins* **2012**, *4*, 217–226. [[CrossRef](#)]
50. Claesson, M.J.; van Sinderen, D.; O’Toole, P.W. *Lactobacillus* phylogenomics—Towards a reclassification of the genus. *Int. J. Syst. Evol. Microbiol.* **2008**, *58*, 2945–2954. [[CrossRef](#)]

51. Lakin, S.M.; Dean, C.; Noyes, N.R.; Dettenwanger, A.; Ross, A.S.; Doster, E.; Rovira, P.; Abdo, Z.; Jones, K.L.; Ruiz, J.; et al. MEGARes: An antimicrobial resistance database for high throughput sequencing. *Nucleic Acids Res.* **2017**, *45*, D574–D580. [[CrossRef](#)] [[PubMed](#)]
52. Jose, N.M.; Bunt, C.R.; Hussain, M.A. Implications of Antibiotic Resistance in Probiotics. *Food Rev. Int.* **2015**, *31*, 52–62. [[CrossRef](#)]
53. Tynkkynen, S.; Singh, K.V.; Varmanen, P. Vancomycin resistance factor of *Lactobacillus rhamnosus* GG in relation to enterococcal vancomycin resistance (van) genes. *Int. J. Food Microbiol.* **1998**, *41*, 195–204. [[CrossRef](#)]
54. Gueimonde, M.; Sanchez, B.; de los Reyes-Gavilán, C.G.; Margolles, A. Antibiotic resistance in probiotic bacteria. *Front. Microbiol.* **2013**, *4*, 202. [[CrossRef](#)]
55. Sengupta, R.; Altermann, E.; Anderson, R.C.; McNabb, W.C.; Moughan, P.J.; Roy, N.C. The role of cell surface architecture of lactobacilli in host-microbe interactions in the gastrointestinal tract. *Mediat. Inflamm.* **2013**, *2013*, 237921. [[CrossRef](#)]
56. Call, E.K.; Klaenhammer, T.R. Relevance and application of sortase and sortase-dependent proteins in lactic acid bacteria. *Front. Microbiol.* **2013**, *4*, 73. [[CrossRef](#)]
57. Su, X.; Sun, F.; Wang, Y.; Hashmi, M.Z.; Guo, L.; Ding, L.; Shen, C. Identification, characterization and molecular analysis of the viable but nonculturable *Rhodococcus biphenylivorans*. *Sci. Rep.* **2015**, *5*, 18590. [[CrossRef](#)]
58. Gul, N.; Poolman, B. Functional reconstitution and osmoregulatory properties of the ProU ABC transporter from *Escherichia coli*. *Mol. Membr. Biol.* **2013**, *30*, 138–148. [[CrossRef](#)]
59. Liu, X.; Sun, M.; Cheng, Y.; Yang, R.; Wen, Y.; Chen, Z.; Li, J. OxyR Is a Key Regulator in Response to Oxidative Stress in *Streptomyces avermitilis*. *Microbiology* **2016**, *162*, 707–716. [[CrossRef](#)]
60. Nilsen, T.; Nes, I.F.; Holo, H. Enterolysin A, a cell wall-degrading bacteriocin from *Enterococcus faecalis* LMG 2333. *Appl. Env. Microbiol.* **2003**, *69*, 2975–2984. [[CrossRef](#)]
61. Perez, R.H.; Zendo, T.; Sonomoto, K. Novel bacteriocins from lactic acid bacteria (LAB): various structures and applications. *Microb. Cell. Fact.* **2014**, *13*, S3. [[CrossRef](#)] [[PubMed](#)]
62. Kuo, Y.C.; Liu, C.F.; Lin, J.F.; Li, A.C.; Lo, T.C.; Lin, T.H. Characterization of putative class II bacteriocins identified from a non-bacteriocin-producing strain *Lactobacillus casei* ATCC 334. *Appl. Microbiol. Biotechnol.* **2013**, *97*, 237–246. [[CrossRef](#)] [[PubMed](#)]
63. Riley, M.A.; Wertz, J.E. Bacteriocin diversity: Ecological and evolutionary perspectives. *Biochimie* **2002**, *84*, 357–364. [[CrossRef](#)]
64. Majeed, H.; Gillor, O.; Kerr, B.; Riley, M.A. Competitive interactions in *Escherichia coli* populations: The role of bacteriocins. *ISME J.* **2011**, *5*, 71–81. [[CrossRef](#)]
65. Meijerink, M.; van Hemert, S.; Taverne, N.; Wels, M.; de Vos, P.; Bron, P.A.; Savelkoul, H.F.; van Bilsen, J.; Kleerebezem, M.; Wells, J.M. Identification of genetic loci in *Lactobacillus plantarum* that modulate the immune response of dendritic cells using comparative genome hybridization. *PLoS ONE* **2010**, *5*, e10632. [[CrossRef](#)]
66. Kaur, S. Bacteriocins as Potential Anticancer Agents. *Front. Pharmacol.* **2015**, *6*, 272. [[CrossRef](#)]
67. Bondy-Denomy, J.; Qian, J.; Westra, E.R.; Buckling, A.; Guttman, D.S.; Davidson, A.R.; Maxwell, K.L. Prophages mediate defense against phage infection through diverse mechanisms. *ISME J.* **2016**, *10*, 2854–2866. [[CrossRef](#)]
68. Douillard, F.P.; Ribbera, A.; Kant, R.; Pietila, T.E.; Jarvinen, H.M.; Messing, M.; Randazzo, C.L.; Paulin, L.; Laine, P.; Ritari, J.; et al. Comparative genomic and functional analysis of 100 *Lactobacillus rhamnosus* strains and their comparison with strain GG. *PLoS Genet.* **2013**, *9*, e1003683. [[CrossRef](#)]
69. Ross, P.; Mayer, R.; Benziman, M. Cellulose biosynthesis and function in bacteria. *Microbiol. Rev.* **1991**, *55*, 35–58.
70. Lorca, G.L.; Font de Valdez, G.; Ljungh, A. Characterization of the protein-synthesis dependent adaptive acid tolerance response in *Lactobacillus acidophilus*. *J. Mol. Microbiol. Biotechnol.* **2002**, *4*, 525–532.
71. Jiao, Y.; Cody, G.D.; Harding, A.K.; Wilmes, P.; Schrenk, M.; Wheeler, K.E.; Banfield, J.F.; Thelen, M.P. Characterization of extracellular polymeric substances from acidophilic microbial biofilms. *Appl. Environ. Microbiol.* **2010**, *76*, 2916–2922. [[CrossRef](#)] [[PubMed](#)]

72. Wang, L.; Wise, M.J. Glycogen with short average chain length enhances bacterial durability. *Naturwissenschaften* **2011**, *98*, 719–729. [[CrossRef](#)] [[PubMed](#)]
73. Kerns, J.C.; Arundel, C.; Chawla, L.S. Thiamin deficiency in people with obesity. *Adv. Nutr.* **2015**, *6*, 147–153. [[CrossRef](#)] [[PubMed](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).