Comparative Genomic and Functional Characterization of *Lactobacillus casei* Group (LCG) Probiotic Strains Isolated from Traditional Yogurts by Next-Generation Sequencing

Mati Ullah[1]*, Muhammad Rizwan[2], Ali Raza[3], Xianlin Zhao[4], Yanling Sun[4], Sarah Gul[5], Muhammad Ihtesham Waheed[6], Muhammad Nadeem Khan[7], Alvina Gul[8], Sami Ullah Jan[9]* and Chao Huang[4]*

[1]*Department of Microbiology, Faculty of Veterinary & Animal Sciences, The University of Agriculture, Dera Ismail Khan, Pakistan*

[2]*College of Animal Sciences and Technology, Anhui Agricultural University, Hefei, Anhui, China*

[3]*Hefei National Laboratory for Physical Sciences at Microscale, School of Life Sciences, University of Science and Technology of China, Hefei, Anhui, China*

[4]*Biotechnology Research Center, Shandong Academy of Agricultural Sciences, Jinan, China*

[5]*Department of Biological Sciences, Faculty of Basic and Applied Sciences, International Islamic University, Islamabad, Pakistan*

[6]*Department of Plant Breeding and Genetics, University of Agriculture, Peshawar*

[7]*Department of Microbiology, Faculty of Biological Sciences, Quaid-I-Azam University, Islamabad, Pakistan*

[8]*Atta-Ur-Rahman School of Applied Biosciences, National University of Sciences and Technology, H-12 Campus, Islamabad, Pakistan*

[9]*Department of Bioinformatics and Biosciences, Faculty of Health & Life Sciences, Capital University of Science and Technology, Islamabad, Pakistan*

ABSTRACT

Lactobacillus casei group which is consists of *Lactobacillus rhamnosus*, *Lactobacillus casei*, and *Lactobacillus paracaseii* an important group of probiotic lactic acid bacteria with numerous commercial and applied health applications. In the present study, we comparatively analysed the genomes of three strains (*Lactobacillus rhamnosus* ATCC 53103, *Lactobacillus casei* L9, and *Lactobacillus paracasei* W56) of LCG isolated from traditional yogurt samples using the next-generation sequencing approach. The GC contents of the characterized strains were in the range of 46.03 to 46.52% while the protein-coding sequences (CDS) were in the range of 2544 to 2888. The pan-genome analysis of the three strains identified a total of 5449 genes with 377 as core genes. The functional characterization by subsystem category distribution analysis revealed "Carbohydrates" is the largest subsystem followed by "protein metabolism" while the biological pathway analysis by the Kyoto encyclopedia of genes and genomes (KEGG) database identified the "metabolic and biosynthesis of secondary metabolites pathway" as the dominant one. Similarly, the functional characterization by eggNOG database revealed that the COGs associated with "Carbohydrate transport and metabolism" were dominant. These find-ings provide a broader understanding of the genome features as well as their functional characterization of the famous probiotic bacterial group.

*   Corresponding author: samiullahjan@gmail.com, samiullahjan@gmail.com, 67731020@qq.com

INTRODUCTION

*L*actobacillus is the largest genus of family Lactobacillaceae currently containing more than 295 species and 54 subspecies (http://www.bacterio.net/lactobacillus.html) (Lee, 1996). *Lactobacillus* species are members of normal microflora in humans and animals colonizing their intestinal and urogenital tracts. Beside this, these are also present in diverse food products

ranging from fruits and vegetables to variety of naturally fermented products, where they act as natural conservators for texture, aroma, flavor and nutritional values of the products (Atrih *et al*., 2001; Messens and De, 2002; Mohd Adnan and Tan, 2007). The *Lactobacillus* spp. have been investigated widely as starter cultures for fermentation and probiotic as well (Gaggìa *et al*., 2010; Vinderola *et al*., 2019). Their extended use in fermented products resulted in their acknowledgment as GRAS (generally recognized as safe) by the US Food and Drug Authority (FDA), and won a place on the QPS (qualified presumption of safety) list collected by the European Food Safety Authority (EFSA, 2018).

Among the *Lactobacillus* species, *Lactobacillus rhamnosus*, *Lactobacillus casei*, and *Lactobacillus paracasei* are closely interrelated, collectively termed as *Lactobacillus casei* group (LCG). The species of LCG group are most studied ones because of their commercial, industrial and applied health applications. Commercially, they are used for fermentation to stabilize dairy products, often preparing food with better taste and texture. They are reported to produce numerous bioactive metabolites which upon consumption confer benefits to the host defining those probiotics. As such, many LCG strains are reported probiotics. One member, *L. rhamnosus* GG (LGG), is perhaps one of the most studied bacterial strains in relation to health applications (Segers and Lebeer, 2014). It has been reported that *L. rhamnosus* GG can shorten the duration of infectious diarrhea, reduce antibiotic-associated symptoms, and alleviate food allergy and atopic dermatitis in children (Doron *et al*., 2005). The other two species *L. casei and L. paracasei* have also been isolated from a variety of environmental habitats, including raw and fermented dairy (especially cheese) and plant materials (e.g., wine, pickle, silage, and kimchi). They are used as acid-producing starter cultures in milk fermentation, as adjunct cultures for intensification as well as for acceleration of flavor development in bacterial-ripened cheeses. Generally, these non-starter species share a dominant portion in the LAB community of ripening cheese.

Yogurt, a dairy fermented product is essential part of daily diet around the globe and its uses history is as old as human itself. It has been fermented in a traditional way when human even did not know the basics of fermentation, modern technology, and equipment. From that time until today yogurt is not only safe to use but also acts as therapeutics to treat many illnesses (Fernández *et al*., 2015). Yogurt fermented in traditional way contains a diverse array of microorganisms, most of which are lactic acid bacteria. Unique characteristics and health benefits of yogurt are definitely due to intrinsic lactic acid bacteria (Joishy *et al*., 2019). Isolated bacteria from yogurt are also often used as probiotics or starter culture for fermentation (Nawaz *et al*., 2019). But to this day, characterization of these bacteria is being done with conventional culturing and biochemical testing methods (Patil *et al*., 2010). However, genetic as well as comparative studies of this important product and associated microbes are almost non-existent (Aryana and Olson, 2017).

Comparative genomics study of lactic acid bacteria is helpful to identify the core and of pan-genomes. Several studies have highlighted the importance of this approach to characterize lactic acid bacteria (Sun *et al*., 2015; Zheng *et al*., 2015). In the present study, we applied the comparative genomics approach using next-generation sequencing technology to investigate the genomic features and their functional annotation of *Lactobacillus casei* group isolated from traditional yogurts. To our knowledge, this study is the first of its kind that provide detail characterization about the three species of LCG group from naturally fermented traditional yogurt environment.

MATERIALS AND METHODS

Yogurt samples

Homemade yogurt samples prepared from domestic cow milk were collected from local vendors in the Karak district of Khyber Pakhtunkhwa, Pakistan. The samples were collected in sterile bottles along with ice pack and immediately sent to the laboratory under cold conditions for further analysis.

Isolation and identification of probiotic bacterial species

The collected samples were aseptically homogenized, and one milliliter of the sample was dissolved in 9 milliliters of phosphate-buffered saline (PBS) as an initial dilution. The samples were further serially diluted, and the different dilutions were plated on general Man–Rogosa–Sharpe (MRS) agar plates. Additionally, selective culture media plates were used for each species/strain. The selective media used was MRS-Bile for *Lactobacillus casei/ paracasei,* and MRS Vancomycin for *Lactobacillus rhamnosus,* respectively. Plates were incubated in triplicate under anaerobic conditions at 37, and 45°C for 24–72 hours. The Yuejin anaerobic incubator was used for probiotic bacterial growth (Shanghai Yuejin Medical Instruments Co., Ltd, Shanghai, China). The pure bacterial colonies for each species were isolated using subculture technique. The pure culture isolates were used for Gram staining, cell morphology, etc. The two closely related species of the *Lactobacillus casei* group (*Lactobacillus casei* and *Lactobacillus paracasei*) were distinguished through custom-designed specific PCR primer as described previously (Patro *et al*., 2016).

DNA isolation

The pure bacterial colonies from agar plate were grown overnight in MRS broth followed by DNA extraction through Tiangen bacterial DNA isolation kit (Tiangen, Beijing, China). Further, Qubit 2.0 fluorometer (Invitrogen, Carlsbad, CA, USA) was used to assessed quantity of freshly extracted DNA.

DNA library preparation and sequencing

DNA Libraries for high-throughput sequencing were built through commercial VazymeTruePrep™ DNA Library Prep Kit V2 for Illumina (Vazyme Biotech, Nanjing, China). For sequencing, Illumina HiSeq-2000 sequencing platform (BGI-Shenzhen, Shenzhen, China) was used.

Bioinformatics analysis

Trimming, quality control, and de novo assembly

The trimming of the sequencing reads was carried out via Trimmomatic version 0.38 (Bolger *et al*., 2014). The contaminated reads were filtered using BWA aligned to the reference sequences. The assembly of the high-quality sequencing reads was performed using shovill (version 1.1.0, https://github.com/tseemann/shovill) assembler. The de novo assembly results were evaluated using the Quast software (Gurevich *et al*., 2013). DNA Plotter (Carver *et al*., 2009), an interactive Java application was used for generation of circular DNA plots. The draft genome of isolated specie NY2 was used as reference genome, randomly connected the Contig sequences of this genome to generate a full length pseudo-single genome. The Prokka program was used for annotations and aligned the NY4 and Y1 genomes to the reference genomic group constructed by NY2, using a minimum similarity of 70%, and convert the comparison results into GFF format. The results were imported in the form of image.

Genome annotation

The assembled genome annotation was performed through the Prokka (version 1.14.5, 10.1093/ bioinformatics/btu153) pipeline. The subsystem identification in the genome and functions assigning was determined through the protein-coding sequences (CDS). The comparison of the assembled genome sequences was evaluated in the SEED Subsystem (Overbeek *et al*., 2014). The evaluation of orthologous genes was performed via clusters of orthologous genes (COGs) (Galperin *et al*., 2015) and eggnog (Huerta-Cepas *et al*., 2016). To examine the high-level functions of the LCG group isolates, the Kyoto encyclopedia of genes and genomes (KEGG) database (https://www.genome.jp/kegg/) was accessed.

Pan-genome analysis

The pan-genome analysis of the bacterial genomes was performed by roary (version 3.13.0, 10.1093/ molbev/msz284) pipeline. Briefly, the protein sequences from gff3 files were extracted to a reduced set of protein sequences followed by an all-against-all comparison with BLASTP. The sequence identity was set 98% with e-value 1E-6. Based on BLASTP similarity score, sequences were clustered with a Markov cluster (MCL) algorithm. Further, a concatenated core gene alignment sequence was produced using PRANK (10.1007/978-1-62703-646-7_10).

RESULTS

Sequencing data and assembled genome properties

The summary and statistics of the sequence data and assembly results is shown in Table I. The number of clean reads used in the analysis was in the range of 2303582 to 5418977, with a mean value of3481040. The GC contents were in the range between 46.03to 46.52%, with 46.26% is the mean value. The coverage and GC contents of each identified species/strains in the corresponding sample are shown in Figure 1. The assembled contigs from the genome sequencing data of the three bacterial genomes vary and were in the range of 116 (lowest) for sample NY4 and 201 (highest) for sample NY2. The other important genome features characterized in the current analysis are summarized in Table I.

Table I. Summary of the sequencing data and identified genome features.

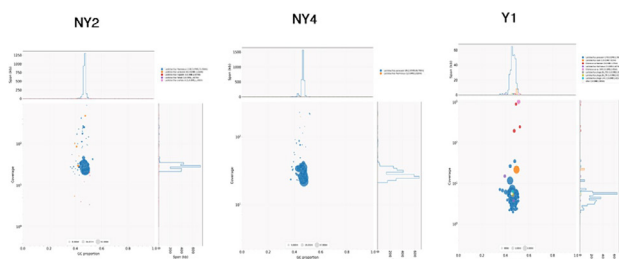| Library name	NY2	NY4	Y1
Species/ strain	*Lactobacillus rhamnosus ATCC 53103*	*Lactobacillus paracasei L9*	*Lactobacillus casei W56*
Raw reads	8878209	6115215	6922700
Clean reads	2303582	2720561	5418977
Raw base (G)	2.66	1.83	2.08
Clean base (G)	0.69	0.82	1.63
GC contents %	46.52	46.25	46.03
Contigs	201	116	121
Bases	2747227	2967516	2988029
CDS	2544	2833	2888
Gene	2647	2928	2985
mRNA	2647	2928	2985
Misc_RNA	31	35	32
rRNA	6	5	7
tRNA	65	54	57
tmRNA	1	1	1

Fig. 1. Blob plots indicate the results of BLAST matches of each of the three *Lactobacillus casei* Group (LCG) assembled genomes. The colors indicate BLAST matches to different species of bacteria. The proportion of GC contents in each assembled genome is also indicated.

Comparative genome analysis

The results of comparative genome-analysis for the three assembled annotated genomes are shown in Figures 2-5. The bacterial species analyzed were three different but closely related species. The pan-genome analysis was conducted together for the three assembled genomes. The altogether pan-genome analysis of the three genomes from three different species identified a total of 5449 genes among which 377 were core genes and 5072 were shell genes (Supplementary Table SI). The same and closely related species strains share some similar accessory genome (Fig. 3A). The specie *L. rhamnosus* (NY2) in *L. casei* group represents a bit different pattern of the accessory genome while the remaining two species *L. paracasei* (NY4) and *L. casei* (Y1) also have a varying accessory genome pattern (Fig. 3A). The sequence cluster of the protein-coding gene in each strain was different, ranging from 2544 to 2888 (detail provided in Supplementary Table SI).

Further, the data obtained by evaluating the number of conserved genes versus the number of total genes as well as the number of new genes and unique genes have varying pattern (Figs. 4 and 5). The number of conserved genes decreases as the number of genomes increases (Fig. 4A). However, the number of unique genes increased as the number of genomes embedded for the first three genomes and showed a declining pattern as further genome was embedded in the analysis (Fig. 4B). The pattern of the number of new genes showed a constant decline as the number of genomes embedded in the analysis (Fig. 4B).

Functional genome annotation
Subsystem category distribution

The subsystem category distribution of the dominant systems identified in the three annotated genomes using the SEED database is shown in Figure 5. Among these subsystems, Carbohydrates was the largest subsystem, followed by protein metabolism. The other dominant

systems were DNA metabolism and nucleoside and nucleotide (Fig. 5). Besides this, numerous other subsystems with different numbers were identified see Figure 6 and (Supplementary Table SII-SV).
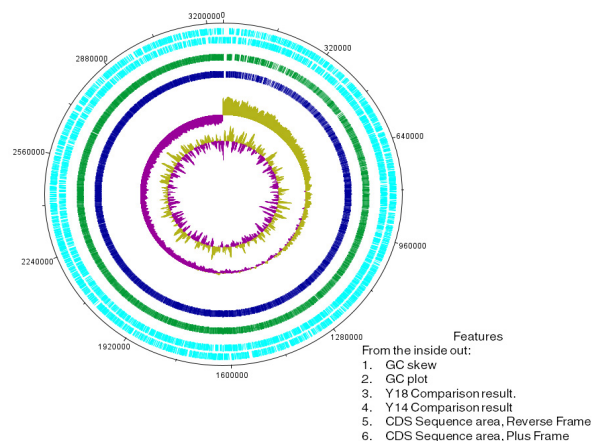

Fig. 2. Circular plot of *Lactobacillus casei* Group (LCG) genome generated by DNA plotter with NY2 as a reference. Circles indicate, from inside outwards: GC skew; GC content; NY4 Comparison results; Y1 comparison result; CDS sequence (reverse frame); CDS sequence (plus frame).
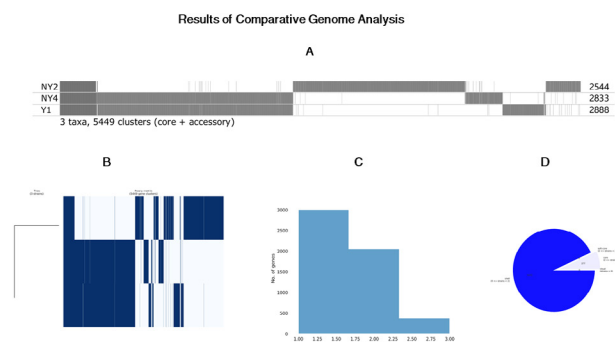

Fig. 3. Comparative genome analysis of the three *Lactobacillus casei* Group (LCG) strains. (A) Roary matrix. Pan-genome analysis of three *Lactobacillus casei* Group (LCG) strains annotated genomes. Roary produced the gene presence/absence matrix. A total of 5449 protein-coding gene sequence clusters produced by Roary are shown in the figure. The presence of the gene is represented by grey color while the absence is shown by white color. The numbers represent clusters in each assembled annotated genome. (B) Pan-genome constructed by Roary based on the core and accessory genes showing phylogenetic relatedness of the strains by blue (present) and white (absent (C) Barplot indicating the ratio of number of genome vs the number of genes (D) The pan-genome pie chart showing gene content visualized with the use of Roary software.
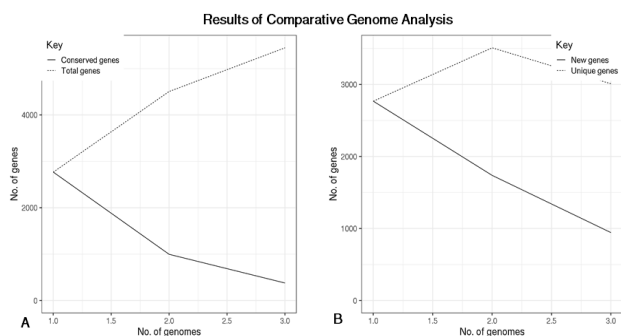
Fig. 4. Heap's law chart representation regarding conserved genes vs total genes in three assembled genomes (A) Diagram of conserved genes per number of genomes. (B) Diagram of new and unique genes per number of genomes.
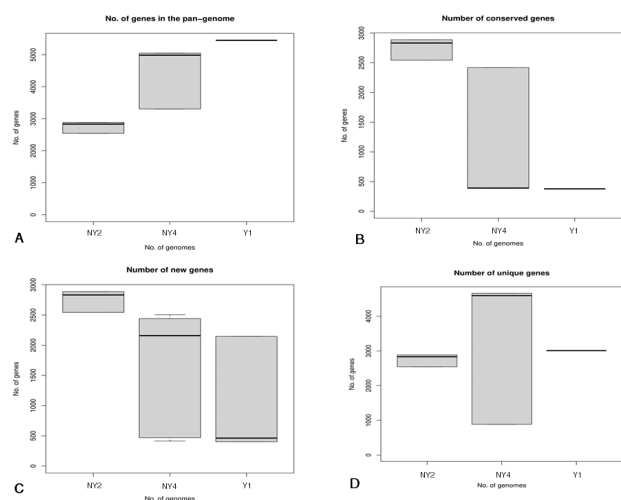

Fig. 5. Box plots representation of pan genome analysis of the three *Lactobacillus casei* Group (LCG) strains. (A) Represent the number of genes in pan genome of the three assembled genome (B) Number of conserved genes identified in three assembled genome (C) Number of new genes identified in three assembled genome (D) Number of unique genes identified in three assembled genomes.

Genome annotation with eggNOG database

The eggnog database, hosted by the European molecular biology laboratory, is a database of orthologous groups and function annotation of the genome. The functional study analysis of the three annotated genomes revealed that the COGs associated with "Carbohydrate transport and metabolism were dominant followed by those associated with Replication, recombination, and repair. The other dominant categories with the majority of COGs associated were amino acid transport and metabolism and Translation ribosomal structure and biogenesis, respectively Figure 7 and Supplementary Table SVI-SVII.
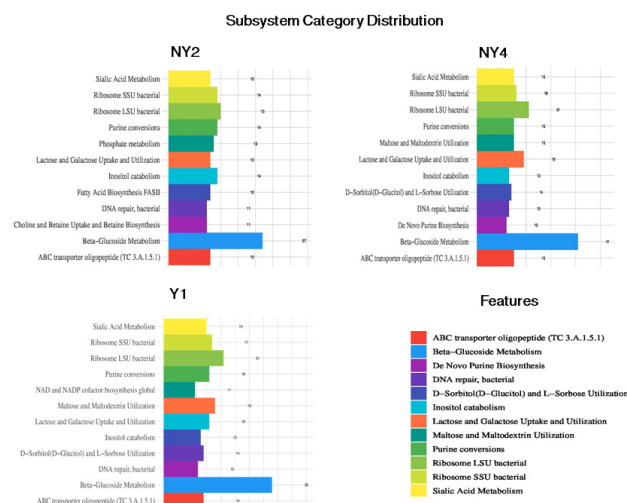

Fig. 6. The subsystem category distribution of the probiotic bacterial genome sequences according to the Seed database. The largest subsystem feature counts belong to Carbohydrates, protein metabolism, DNA metabolism, and nucleoside and nucleotide. The description of various subsystem features are represented in a different color (below right).
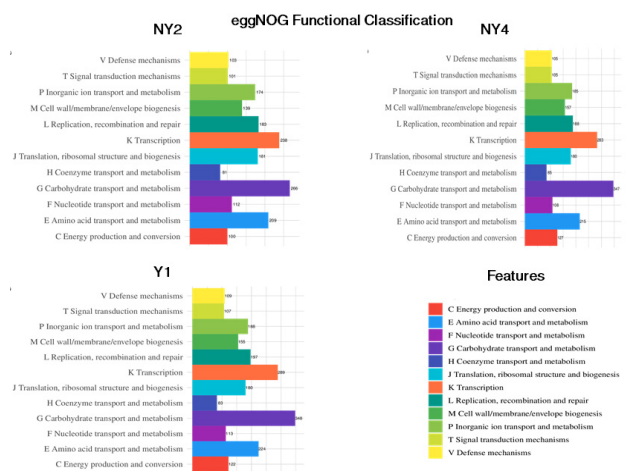
Fig. 7. Classification of eggNOG annotations of the probiotic bacterial genome sequences using eggNOG. The numbers in the plots represent the genes assigned to eggNOG classification. The description of various annotated eggNOG features is represented by a different color (below right).

KEGG analysis

The information provided by the KEGG database is related to the high level functions for high throughput sequencing data regarding genetic information processing, metabolism, cellular processes, human diseases, environmental information processing, drug

development, and organismal systems. Using the KEGG database pathways as a reference, the genome content in the current analysis were characterized to identify the biological pathways. The majority of gene contents were involved in metabolic pathways, biosynthesis of secondary metabolites, microbial metabolism in diverse environments, and biosynthesis of antibiotics respectively Figure 8 and Supplementary Table SVIII-SX.
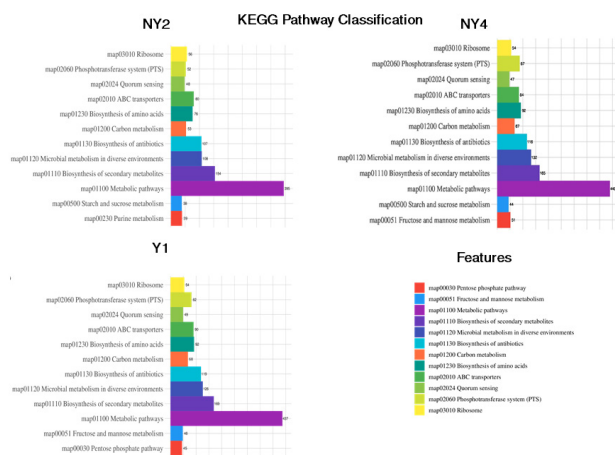

Fig. 8. Classification of the KEGG pathways based on the assembled genome of the three probiotic bacterial strains. The different colors represent the various KEGG pathways identified in the current analysis.

DISCUSSION

The analysis of the current study provides a detailed insight into the genome characterization of important probiotic bacterial species found in natural homemade yogurt samples. Yogurt, the fermented milk product is one of the most famous and widely consumed traditional fermented food and associated with various health benefits. The three probiotic bacterial species identified and characterized in the current analysis have a proven record of ideal starter cultures as well as associated with the health benefits in different health complications (Johansson *et al.*, 2011; Huang *et al.*, 2017; Dimitrellou *et al.*, 2019). Functional genes associated with variety of functions ranging from adaptation and survival to the nutrient metabolism is present on the genomes of each bacterium, but the genome itself is sensitive to environmental conditions and situations (Dong *et al.*, 2019). Moreover, bacteria have niche specific functional characteristics while the genome of lactic acid bacteria is highly fragile and can easily be affected by any external factors (Ceapa *et al.*, 2016). In order to understand bacteria, its functions and associated environment, it is necessary to sequence

and study more new genes. Not only this, but previously sequenced genomes need to be sequenced and revisited (Siezen *et al.*, 2012; Ceapa *et al.*, 2016; Dong *et al.*, 2019). In connection with this need, we adopted an emerging approach to comprehensively analyze the genome contents and its functional aspects of LCG species isolated from traditional yogurt.

The three species of the *Lactobacillus casei* group (LCG), *Lactobacillus casei*, *Lactobacillus paracasei*, and *Lactobacillus rhamnosus*are important group of probiotic bacterial species of lactobacilli. The health-promoting properties and industrial applications of this group have made it an interesting and never-ending research topic. The members of this probiotic group are reported to have a positive role in treating cancer, diabetes, obesity, diarrhea and allergies (Bermudez-Brito *et al.*, 2012). Members of LCG are also used for improving the color texture and flavor of food products (Di Renzo *et al.*, 2018). Similar to others LAB, the members of this probiotic isolated from yogurts and other sources are usually identified and characterized through classical microbiological techniques (Dimitrellou *et al.*, 2019; Bancalari *et al.*, 2020). These techniques are laborious, timing taking are error prone in terms of functional identification. However, the current analysis provides more comprehensive characterization using the high throughput sequencing approach in term of gene contents identification and functional annotation, subsystem distribution, and biological pathway identification (Figs. 3-7).

Next-generation sequencing is a powerful tool to identify and characterize the microorganisms from different sample types. Compare to traditional microbiological techniques, NGS has greater precision and accuracy and provides more in-depth insight into the microbial genome characterization as well as their functional annotation. Further, for closely related bacterial species like *Lactobacillus casei* group, which are difficult to differentiate based on traditional techniques, NGS remains a tool of choice to conduct such type of comparative analysis with greater confidence. Although the expansions of NGS in different research areas have been increased over the last few years, it is still very limited in dairy-related research. Our analysis provides a clear direction to expend this sophisticated tool to comprehensively characterize the other important probiotic microorganism from the dairy-related origin. This can expand our knowledge about the choice of probiotic microorganisms, their health effects, as well as their safety concerns.

The GC contents we analyzed for the genomes were in the range between 46.03 to 46.52%, with 46.26% the mean value. The published literature supports our results as lactic acid bacteria have GC contents ranging from 32

to 52%/ (Mendes-Soares *et al.*, 2014). The difference in percentage of GC content is directly linked with genomic size and structure, and ecological niche to which a microorganism belongs, thus it may vary even for strains (Šmarda *et al.*, 2014). The variation in coding sequences and gene number is also strain dependent trait (Sun *et al.*, 2011).

The assembled contigs from the genome sequencing data of the three bacterial genomes vary significantly. The consequence of contig number on the functional characterization has not been confirmed by literature or there is any typical range for the number of contigs so far. The variation in contig number is associated with technology being used for sequencing and methodology used for assembly (Smits, 2019). The pan-genome analysis of the three strains identified a total of 5449 genes among which 377 were core genes and 5072 were shell genes. Differences in the accessory genes are associated with strain specific characteristics (Pintara *et al.*, 2020). The subsystem category distribution of the dominant systems identified in the three annotated genomes using the SEED database. Among these subsystems, Carbohydrates was the largest subsystem, followed by "protein metabolism. The other dominant systems were DNA metabolism and nucleoside and nucleotide". Milk contain high amount of carbohydrates (i.e. 5g/100ml) followed by proteins (3.5g/100ml) and fats (1.5/100ml). That is carbohydrate metabolism genes are most common in lactic acid bacteria followed by proteins and fats metabolisms (Buron-Moles *et al.*, 2019). The eggnog database, hosted by the European molecular biology laboratory, is a database of orthologous groups and function annotation of the genome. The functional study analysis of the three annotated genomes revealed that the COGs associated with Carbohydrate transport and metabolism were dominant followed by those associated with Replication, recombination, and repair. The other dominant categories with most COGs associated were amino acid transport and metabolism and Translation ribosomal structure and biogenesis, respectively.

Funding

This research received no external funding.

Data availability statement

The raw data used in the current study have been deposited in NCBI and allocated the accession number of PRJNA664263.

Supplementary material

There is supplementary material associated with this article. Access the material online at: https://dx.doi. org/10.17582/journal.pjz/20210711190719

Statement of conflict of interest

The authors have declared no conflict of interest.

REFERENCES

Aryana, K.J., and Olson, D.W., 2017. A 100-year review: Yogurt and other cultured dairy products. *J. Dairy Sci.*, **100**: 9987-10013. https://doi.org/10.3168/jds.2017-12981

Atrih, A., Rekhif, N., Moir, A.J., Lebrihi, A., and Lefebvre, G., 2001. Mode of action, purification and amino acid sequence of plantaricin C19, an anti-Listeria bacteriocin produced by *Lactobacillus plantarum* C19. *Int. J. Fd. Microbiol.*, **68**: 93-104. https://doi.org/10.1016/S0168-1605(01)00482-2

Bancalari, E., Castellone, V., Bottari, B., and Gatti, M., 2020. Wild *Lactobacillus casei* group strains: Potentiality to ferment plant derived juices. *Foods*, **9**: 314. https://doi.org/10.3390/foods9030314

Bermudez-Brito, M., Plaza-Diaz, J., Munoz-Quezada, S., Gomez-Llorente, C., and Gil, A., 2012. Probiotic mechanisms of action. *Ann. Nutr. Metab.*, **61**: 160-174. https://doi.org/10.1159/000342079

Bolger, A.M., Lohse, M., and Usadel, B., 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, **30**: 2114-2120. https://doi.org/10.1093/bioinformatics/btu170

Buron-Moles, G., Chailyan, A., Dolejs, I., Forster, J., and Mikš, M.H., 2019. Uncovering carbohydrate metabolism through a genotype-phenotype association study of 56 lactic acid bacteria genomes. *Appl. Microbiol. Biotech.*, **103**: 3135-3152. https://doi.org/10.1007/s00253-019-09701-6

Carver, T., Thomson, N., Bleasby, A., Berriman, M., and Parkhill, J., 2009. DNAPlotter: Circular and linear interactive genome visualization. *Bioinformatics*, **25**: 119-120. https://doi.org/10.1093/bioinformatics/btn578

Ceapa, C., Davids, M., Ritari, J., Lambert, J., Wels, M., Douillard, F.P., Smokvina, T., de Vos, W.M., Knol, J., and Kleerebezem, M., 2016. The variable regions of *Lactobacillus rhamnosus* genomes reveal the dynamic evolution of metabolic and host-adaptation repertoires. *Genome Biol. Evol.*, **8**: 1889-1905. https://doi.org/10.1093/gbe/evw123

Di Renzo, T., Reale, A., Boscaino, F., and Messia, M.C., 2018. Flavoring production in Kamut®, quinoa and wheat doughs fermented by Lactobacillus paracasei, *Lactobacillus plantarum*, and *Lactobacillus brevis*: A SPME-GC/MS Study. *Front. Microbiol.*, **9**: 429. https://doi.org/10.3389/fmicb.2018.00429

Dimitrellou, D., Kandylis, P., and Kourkoutas, Y.,

2019. Assessment of freeze-dried immobilized *Lactobacillus casei* as probiotic adjunct culture in yogurts. *Foods*, **8**: 374. https://doi.org/10.3390/foods8090374

Dong, Y., Zhang, F., Wang, B., Gao, J., Zhang, J., and Shao, Y., 2019. Laboratory evolution assays and whole-genome sequencing for the development and safety evaluation of *Lactobacillus plantarum* with stable resistance to gentamicin. *Front. Microbiol.*, **10**: 1235. https://doi.org/10.3389/fmicb.2019.01235

Doron, S., Snydman, D.R., and Gorbach, S.L., 2005. *Lactobacillus* GG: Bacteriology and clinical applications. *Gastroenterol. Clin. North Am.*, **34**: 483-498. https://doi.org/10.1016/j.gtc.2005.05.011

EFSA (European Food Safety Authority), 2018. The 2016 European Union report on pesticide residues in food. *EFSA Journal,* **16**: e05348. https://doi.org/10.2903/j.efsa.2018.5348

Fernández, M., Hudson, J.A., Korpela, R., and de los Reyes-Gavilán, C.G., 2015. Impact on human health of microorganisms present in fermented dairy products: An overview. *Biomed. Res. Int.*, **2015**: 412714. https://doi.org/10.1155/2015/412714

Gaggìa, F., Mattarelli, P., and Biavati, B., 2010. Probiotics and prebiotics in animal feeding for safe food production. *Int. J. Fd. Microbiol.* **141**: S15-S28. https://doi.org/10.1016/j.ijfoodmicro.2010.02.031

Galperin, M.Y., Makarova, K.S., Wolf, Y.I., and Koonin, E.V., 2015. Expanded microbial genome coverage and improved protein family annotation in the COG database. *Nucl. Acids Res.*, **43**: D261-269. https://doi.org/10.1093/nar/gku1223

Gurevich, A., Saveliev, V., Vyahhi, N., and Tesler, G., 2013. Quast: Quality assessment tool for genome assemblies. *Bioinformatics*, **29**: 1072-1075. https://doi.org/10.1093/bioinformatics/btt086

Huang, R., Ning, H., Shen, M., Li, J., Zhang, J., and Chen, X., 2017. Probiotics for the treatment of atopic dermatitis in children: A systematic review and meta-analysis of randomized controlled trials. *Front. Cell Infect. Microbiol.*, **7**: 392. https://doi.org/10.3389/fcimb.2017.00392

Huerta-Cepas, J., Szklarczyk, D., Forslund, K., Cook, H., Heller, D., Walter, M.C., Rattei, T., Mende, D.R., Sunagawa, S., Kuhn, M., Jensen, L.J., von Mering, C., and Bork, P., 2016. eggNOG 4.5: A hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucl. Acids Res.*, **44**: D286-293. https://doi.org/10.1093/nar/gkv1248

Johansson, M.A., Sjogren, Y.M., Persson, J.O., Nilsson,

C., and Sverremark-Ekstrom, E., 2011. Early colonization with a group of Lactobacilli decreases the risk for allergy at five years of age despite allergic heredity. *PLoS One*, **6**: e23031. https://doi.org/10.1371/journal.pone.0023031

Joishy, T.K., Dehingia, M., and Khan, M.R., 2019. Bacterial diversity and metabolite profiles of curd prepared by natural fermentation of raw milk and back sloping of boiled milk. *World J. Microbiol. Biotechnol.*, **35**: 102. https://doi.org/10.1007/s11274-019-2677-y

Lee, B., 1996. Bacteria-based processes and products. In: *Fundamentals of food biotechnology* (eds. B. Lee). VEH, New York. pp. 219-290.

Mendes-Soares, H., Suzuki, H., Hickey, R.J., and Forney, L.J., 2014. Comparative functional genomics of Lactobacillus spp. reveals possible mechanisms for specialization of vaginal *Lactobacilli* to their environment. *J. Bact.*, **196**: 1458-1470. https://doi.org/10.1128/JB.01439-13

Messens, W., and De, V.L., 2002. Inhibitory substances produced by *Lactobacilli* isolated from sourdoughs. A review. *Int. J. Fd. Microbiol.*, **72**: 31-43. https://doi.org/10.1016/S0168-1605(01)00611-0

Mohd Adnan, A.F., and Tan, I.K.P., 2007. Isolation of lactic acid bacteria from Malaysian foods and assessment of the isolates for industrial potential. *Bioresour. Technol.*, **98**: 1380-1385. https://doi.org/10.1016/j.biortech.2006.05.034

Nawaz, F., Khan, M.N., Javed, A., Ahmed, I., Ali, N., Ali, M.I., Bakhtiar, S.M., and Imran, M., 2019. Genomic and functional characterization of *Enterococcus mundtii* QAUEM2808, isolated from artisanal fermented milk product dahi. *Front. Microbiol.*, **10**: 434. https://doi.org/10.3389/fmicb.2019.00434

Overbeek, R., Olson, R., Pusch, G.D., Olsen, G.J., Davis, J.J., Disz, T., Edwards, R.A., Gerdes, S., Parrello, B., Shukla, M., Vonstein, V., Wattam, A.R., Xia, F., and Stevens, R., 2014. The SEED and the Rapid Annotation of microbial genomes using subsystems technology (RAST). *Nucl. Acids Res.*, **42**: D206-214. https://doi.org/10.1093/nar/gkt1226

Patil, M., Pal, A., Anand, T., and Ramana, K., 2010. Isolation and characterization of lactic acid bacteria from curd and cucumber. *Ind. J. Biotechnol.*, **9**: 166-172.

Patro, J.N., Ramachandran, P., Barnaba, T., Mammel, M.K., Lewis, J.L., and Elkins, C.A., 2016. Culture-independent metagenomic surveillance of commercially available probiotics with high throughput next generation sequencing. *mSphere*, **1**:

e00057. https://doi.org/10.1128/mSphere.00057-16

Pintara, A., Jennison, A., Rathnayake, I.U., Mellor, G., and Huygens, F., 2020. Core and accessory genome comparison of Australian and international strains of O157 shiga toxin-producing *Escherichia coli*. *Front. Microbiol*., **11**: 2162. https://doi.org/10.3389/fmicb.2020.566415

Segers, M.E., and Lebeer, S., 2014. Towards a better understanding of *Lactobacillus rhamnosus* GG-host interactions. *Microb. Cell Fact.*, **13**: 1-16. https://doi.org/10.1186/1475-2859-13-S1-S7

Siezen, R.J., Francke, C., Renckens, B., Boekhorst, J., Wels, M., Kleerebezem, M., and van Hijum, S.A.F.T., 2012. Complete resequencing and reannotation of the *Lactobacillus plantarum* WCFS1 genome. *J. Bact.*, **194**: 195-196. https://doi.org/10.1128/JB.06275-11

Šmarda, P., Bureš, P., Horová, L., Leitch, I.J., Mucina, L., Pacini, E., Tichý, L., Grulich, V., Rotreklová, O., 2014. Ecological and evolutionary significance of genomic GC content diversity in monocots. *Proc. natl. Acad. Sci.*, **111**: E4096-E4102. https://doi.org/10.1073/pnas.1321152111

Smits, T.H.M., 2019. The importance of genome sequence quality to microbial comparative genomics. *BMC Genom.*, **20**: 662. https://doi.org/10.1186/s12864-019-6014-5

Sun, Z., Chen, X., Wang, J., Zhao, W., Shao, Y., Guo, Z., Zhang, X., Zhou, Z., Sun, T., Wang, L., Meng, H., Zhang, H., and Chen, W., 2011. Complete genome sequence of *Lactobacillus delbrueckii* subsp. *bulgaricus* strain ND02. *J. Bact.*, **193**: 3426-3427. https://doi.org/10.1128/JB.05004-11

Sun, Z., Harris, H.M.B., McCann, A., Guo, C., Argimon, S., Zhang, W., Yang, X., Jeffery, I.B., Cooney, J.C., Kagawa, T.F., Liu, W., Song, Y., Salvetti, E., Wrobel, A., Rasinkangas, P., Parkhill, J., Rea, M.C., O'sullivan, O., Ritari, J., Douillard, F.P., Ross, R.P., Yang, R., Briner, A.E., Felis, G.E., de Vos, W.M., Barrangou, R., Klaenhammer, T.R., Caufield, P.W., Cui, Y., Zhang, H., and O'Toole, P.W., 2015. Expanding the biotechnology potential of lactobacilli through comparative genomics of 213 strains and associated genera. *Nat. Commun.*, **6**: 8322. https://doi.org/10.1038/ncomms9322

Vinderola, G., Ouwehand, A.C., Salminen, S., and von Wright, A., 2019. *Lactic acid bacteria: Microbiological and functional aspects.* CRC press. https://doi.org/10.1201/9780429057465

Zheng, J., Ruan, L., Sun, M., and Gänzle, M., 2015. A genomic view of lactobacilli and pediococci demonstrates that phylogeny matches ecology and physiology. *Appl. environ. Microbiol.*, **81**: 7233-7243. https://doi.org/10.1128/AEM.02116-15