Advances in Biochemistry and Biotechnology
Saroj D, et al. Adv Biochem Biotechnol 8: 10120.
www.doi.org/10.29011/2574-7258.010120
www.gavinpublishers.com

Research Article

Whole Genome Sequencing-based Genetic Characterization and Safety Assessment of Probiotic *Bacillus subtilis* strain PLSSC

Dina Saroj, Pruthvi Upadhyaya, Vikash Kumar*

Advanced Enzyme Technologies Limited, Sun Magnetica, Louiswadi, Thane-West, Maharashtra 400 604, India

***Corresponding Author:** Vikash Kumar, Advanced Enzyme Technologies Limited, Sun Magnetica, Louiswadi, Thane-West, Maharashtra 400 604, India

Citation: Saroj D, Upadhyaya P, Kumar V (2024) Whole Genome Sequencing-based Genetic Characterization and Safety Assessment of Probiotic *Bacillus subtilis* strain PLSSC. Adv Biochem Biotechnol 9: 10120. DOI: 10.29011/2574-7258.010120

Received Date: 20 June, 2024; **Accepted Date**: 27 June, 2024; **Published Date:** 1 July, 2024

Abstract

Varieties of microbes are beneficial to humans. Probiotic microbes act by balancing the microbial flora in the gut. Whole genome sequencing (WGS) technique has been used to identify and characterize novel microbes. In the present study, genetic characterization of probiotic *Bacillus subtilis (B. subtilis)* strain PLSSC was carried out by WGS. A single scaffold of 4,204,670 bp size having 43.58 mol% G+C was derived by using both short reads and long reads generated by Illumina and Nano-pore respectively. Gene annotation by Prokaryotic Genome annotation pipeline (PGAP) resulted in identification of 4296 coding sequences (CDS), 86 tRNAs, 30 rRNAs and 5 ncRNA. BLASTN of assembled genome revealed that *B. subtilis* KCTC 3135 is the closest strain showing ~99% identity. Analysis of assembled PLSSC genome for the genes related to safety such as antibiotic resistance, virulence factors and toxins revealed that none of the identified genes pose risk to human health. Presence of clustered regularly interspaced short palindromic repeats (CRISPR) and lack of functional prophage sequences appeared to be advantageous in maintaining the genome stability. Additionally, presence of genes contributing to probiotic properties such as acid and bile salt tolerance, anchoring to the gut mucosa and anti-microbial activity in the PLSSC genome ensure strain survivability thereby increasing their colonization and reducing pathogenic adherence in the gut. Overall, genomic analysis strongly suggests that *B. subtilis* strain PLSSC is a safe strain and can be used as a probiotic.

Keywords: Probiotics; Bacteria; *Bacillus*; Whole genome sequencing; Safety

Introduction

The gut microbiota is a complex microecosystem that harbours around 100 trillion microbes derived from over 2,000 diverse species [1]. It plays a vital role in maintaining the energy balance, regulating gut pH and developing gut cells in the host body [2]. Some species of the gut microbiome including *Bacillus, Lactobacillus, Lactococcus, Bifidobacterium, Streptococcus, Enterococcus*, and *Saccharomyces cerevisiae* have been classified as probiotics [2]. Probiotics are microorganisms that provide a health benefit to the host when administered at specific concentrations [3]. They have also been reported to produce antimicrobial compounds having bio-therapeutic potential [4] and are involved in the immunological development and intestinal barrier improvement in the host [1].

In order to be recognized and used as a probiotic, a microbe must possess several characteristics such as adherence, a range of activities against pathogenic microbes, tolerance to harsh conditions, etc. to survive the upper intestinal tract and reach the site of action. The *Bacillus spp*., due to their spore-forming ability and higher rate of secondary metabolism are promising probiotic candidates. The endospore structure is able to survive harsh GIT environment, such as high temperatures, low pH and high bile salts, and the range of active substances synthesized confer anti-cancer and antioxidant properties to the strain [2]. These characteristics are responsible for the higher success rate of *Bacillus* in colonizing the GIT as compared to other genera [5]. *Bacillus subtilis*, first described in the 19th century, is a Gram positive, aerobic, fast growing, spore forming bacterium [6]. *Bacillus subtilis* strain 168

was one of the first bacteria whose genome was fully sequenced and remains as one of the best annotated genomes [7,8]. *B. subtilis* as a probiotic offers multiple health benefits such as anti-allergic effects, increased immunity and reduced bone loss in postmenopausal women [9]. Moreover, *Bacillus* isolates also produce a wide range of antimicrobial compounds, including lipopeptides and Bacteriocin-Like Inhibitory Substances (BLIS) [10]. Several species of the genera including *B. subtilis, B. clausii, B. licheniformis, B. coagulans, B. polyfermenticus, B. cereus and B. pumilus* are widely used as probiotics [6].

Even though consumed regularly, there are a few risks associated with some *B. subtilis* strains, such as the presence of genes responsible for enterotoxin and biogenic amine synthesis, possibility of AMR gene transfer, and cytotoxicity. Therefore, before a strain could be used as a probiotic, it is essential to evaluate its safety and efficacy by assessing parameters like presence of potential virulent or pathogenic factors, toxin and biogenic amine production, and antibiotic resistance. Whole-genome sequencing (WGS) technology has been employed not only in the characterization of promising probiotic strains but also for the evaluation of its safety aspects in an efficient manner. [11-16]. Bacterial strains which differ by even a single nucleotide can be identified by WGS. Particularly, Food and Drug Administration (FDA) has started emphasizing on WGS to identify food borne pathogens and prevent illness. EFSA (European Food Safety Authority) also recommends WGS data for the approval of microbes as feed or food additives [11].

In this study, the WGS of *B. subtilis* strain PLSSC was performed using Illumina and Nanopore technology. The strain was identified and confirmed by analysing the 16S rRNA gene sequence, mol% G+C content, and average nucleotide identity (ANI). Genome analysis suggested that *B. subtilis* strain PLSSC is safe and has genes essential for the probiotic characteristics.

Materials and Methods

Extraction of genomic DNA and purification

GeneAll Kit (GeneAll, Seoul, South Korea) was used for the isolation and purification of *B. subtilis* PLSSC genomic DNA. Bacterial culture (2 ml) was grown overnight and centrifuged at 8000×g for 2 minutes. After centrifugation, pellet was collected and re-suspended in 180 µl of GP buffer (containing 30 mg ml−1 of lysozyme) and subsequently incubated at 37 °C for 30 minutes. After incubation, RNAase and Proteinase K were added to the suspension. Prior to transfer to a binding column, 200 µl of absolute ethanol was added to the lysate. Column was washed using a wash buffer and afterwards, elution of DNA was carried out with 100–200 µl of sterile Milli-Q water. After isolation, genomic DNA was assessed in terms of quantity and quality by NanoDrop-2000 (ThermoFisher Scientific, USA), Qubit, and agarose gel electrophoresis.

Whole genome sequencing and assembly

For the preparation of Illumina and Nanopore Whole Genome Sequencing (WGS) libraries, purified genomic DNA from *B. subtilis* strain PLSSC was used. Illumina WGS library was prepared by using Illumina-compatible SureSelect^{QXT} (Fig. 1) whole genome library prep kit (Agilent, Santa Clara, CA, U.S.A.).

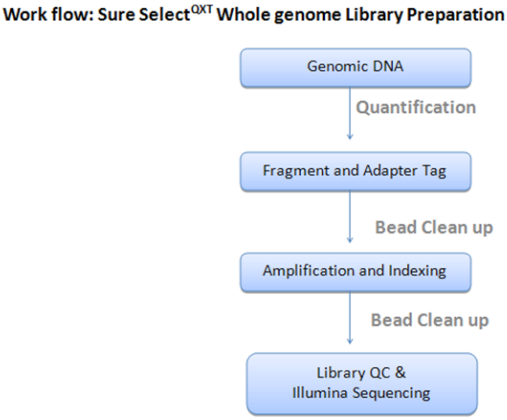

Figure 1: Workflow for illumina SureSelect^{QXT} library preparation.

25 ng DNA was subjected to fragmentation and adapter tagging using Sure Select QXT enzyme. HighPrep PCR beads were used to purify fragmented and adapter-tagged DNA. Amplification and indexing of fragmented and adapter-tagged DNA were carried out using 6-cycles of PCR. Again, HighPrep beads (MAGBIO, MD, USA) were used to purify PCR products. Purified PCR products were subjected to library quality control check. Quantification of Illumina-compatible sequencing library was done by Qubit fluorometer (Thermo Fisher Scientific, MA, USA) and fragment size distribution was analysed on Agilent TapeStation (Table 1 and Fig. S1).

The tape station profile of Illumina library reveals that the fragment size ranges from 187 to 1149 bp. However, larger proportion of the Illumina-compatible sequencing library had a fragment size ranging from 200 to 700 bp. Taking into account the combined adapter size i.e. approximately 120 bp, the effective user-defined insert size was 80 to 580 bp, which was obtained with optimal concentration. This ensured that the library was suitable for Illumina sequencing to get the desired amount of sequencing data. A total of 2,490,092 (R1 + R2) reads were generated for *B. subtilis* strain PLSSC on Illumina MiSeq platform.

1 µg of genomic DNA from *B. subtilis* strain PLSSC was used for the generation of long read library for Nanopore sequencing. Firstly, the genomic DNA was end-repaired with NEBnext ultra II end repair kit (New England Biolabs, MA, USA), followed by

a clean-up with 1x AmPure beads (Beckmann Coulter, USA). Library preparation was then performed using a native barcoding kit (NBD103) wherein barcodes were ligated using NEB blunt/TA ligase (New England Biolabs, MA, USA), followed by a clean-up with 0.5x AmPure beads (Figure 2).

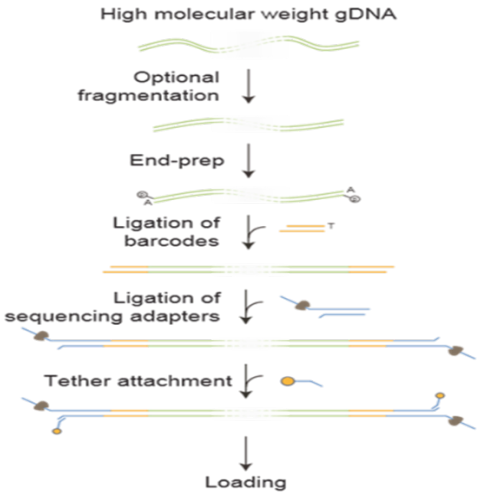

Figure 2: Native barcoding library preparation.

Nanopore sequencing was executed on GridION X5 (Oxford Nanopore Technologies, Oxford, UK) using SpotON flow cell R9.4 (FLO-MIN106) in a 48 hour sequencing protocol on MinKNOW 2.1 v18.05.5. Albacore v2.3.1 was used for base-calling Nanopore raw reads in '*fast5'* format to '*fastq'* format.

MaSuRCA Hybrid Assembler [17] was used for generating a hybrid assembly of Illumina and nanopore reads. Gene annotation of assembled genome was carried out by NCBI Prokaryotic genome Annotation Pipeline [18].

Sample ID	Qubit (ng/µl)	Vol. (µl)	Yield (ng)	Index1	Index1 Sequence	Index2	Index2 Sequence
GT_SO_7892_*B. subtilis* strain PLSSC	9.54	10	95.4	P7i6	TAGGCATG	P7i18	ACTGCATA

Table 1: Description of the library.

Calculation of average nucleotide identity

Average Nucleotide Identity (ANI) is used to measure the relatedness between two genomes. ANI calculation of the assembled genome of *B. subtilis* strain PLSSC against the genome of *B. subtilis* KCTC 3135 was performed according to Goris et al. [19] and calculated using ANI calculator (http://enve-omics.ce.gatech.edu/ani/) with default parameters.

Identification of antibiotic resistance and virulence factor genes

For the identification of antibiotic resistance associated genes, a homology-based search was carried out between the assembled genome of *B. subtilis* strain PLSSC and Comprehensive Antibiotic Resistance Database (CARD) [20]. For the identification of hits, BLASTX was used with similarity >30%, coverage >70% and e-value <1e-02. Additionally, the assembled genome was also compared with the COG database [21] to identify gene function.

Virulence Factor Database (VFDB) [22] having 3072 sequences in the core database was used to search for the virulence factors genes. Search was performed using BLASTX and only those hits which showed a similarity >30%, coverage >70% and e-value <1e-02 were taken into consideration.

Identification of biogenic amine producing genes

Genes involved in biogenic amine production, mainly amino acid decarboxylases were searched in the assembled genome of *B. subtilis* strain PLSSC as described by Salvetti et al. [23]. Protein sequences of the short-listed biogenic amine producing genes (amino acid decarboxylases) were downloaded from the Uniprot database and BLASTX was performed between the assembled genome and biogenic amine producing proteins.

Assessment of Genomic Stability

Stability of the assembled genome was assessed in accordance with Salvetti et al. [23]. Presence of insertion sequences (IS), prophage sequences and clustered regularly interspaced short palindromic repeats (CRISPR) sequences in the genome was investigated. PHASTER, a web-based server was used to identify prophage sequences in the assembled bacterial genome [24]. Mobile elements were searched using ISfinder (web-based software) and ACLAME database (version 0.4). For screening CRISPR sequences, CRISPRCasFinder was used [25].

In silico mining of probiotic genes

The assembled genome of *B. subtilis* strain PLSSC was assessed for the genomic features which contribute to the probiotic properties such as adhesion to gut mucosa, acid tolerance, bile salt tolerance and environmental stress resistance [26]. All the predicted protein sequences (n=4296) were annotated by submitting them to Batch-CD Search web service available in the conserved domain database. In the Batch-CD search, the Pfam database containing 19178 position specific scoring matrices (PSMMs) was selected for functional annotation of predicted proteins.

Results

Assembly of *B. subtilis* strain PLSSC genome and its features

De novo WGS of *B. subtilis* strain PLSSC yielded a single scaffold of 4,204,670 bp in size. Annotation of the assembled genomic sequence by the NCBI Prokaryotic Genome annotation pipeline resulted in 4296 coding sequences (CDS), 86 tRNAs, 30 rRNAs and 5 ncRNA, which were in accordance with the annotation described for *B. subtilis* KCTC 3135 (https://www.ncbi.nlm.nih.gov/nuccore/CP015375.1/). The predicted protein coding genes were annotated against all the proteins (228,533) listed in the Uniprot database for bacteria (Uniprot Consortium, 2014). We observed that, out of 4,296 CDS from the assembled genome, 4,256 got significant hits (identity >30% and e-value <=1e-05) against the Uniprot Bacterial protein database. For these 4,256 proteins, a gene ontology classification (molecular function, cellular component and biological process) was also carried out and represented in the form of a pie chart (Figure 3). Gene ontology analysis indicates that 39.25%, 45.49% and 15.26% of the proteins from the assembled genome represented for molecular function, cellular component and biological process respectively. Major pathway groups associated with the predicted proteins and their functions are indicated in Figure 4 and 5.

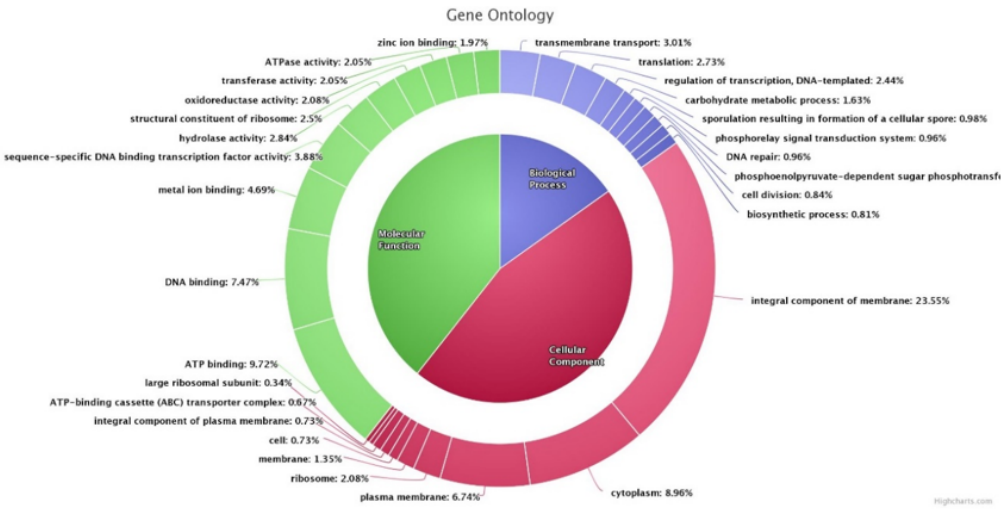

Figure 3: Gene ontology (GO) association of predicted protein-coding genes.

Using the same criteria, predicted protein coding genes from the assembled genome were also annotated with Clusters of Orthologous Groups (COG) database (http://www.ncbi.nlm.nih.gov/COG/) [21]. It was observed that, out of 4,296 CDS from the assembled genome, homology was found against the COG database for 4,091.

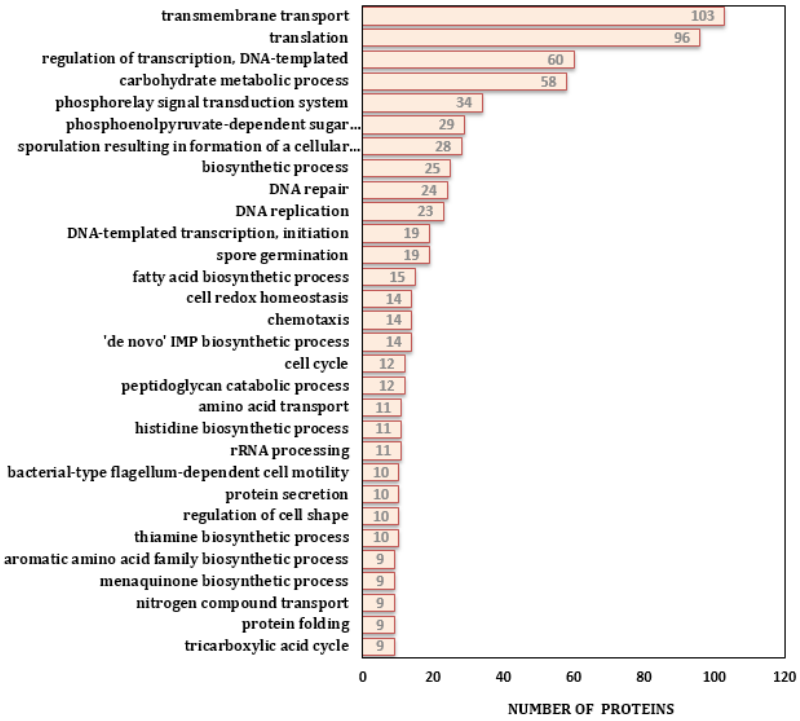

Figure 4: Pathway abundance of predicted proteins. Only top 30 pathways have been shown.

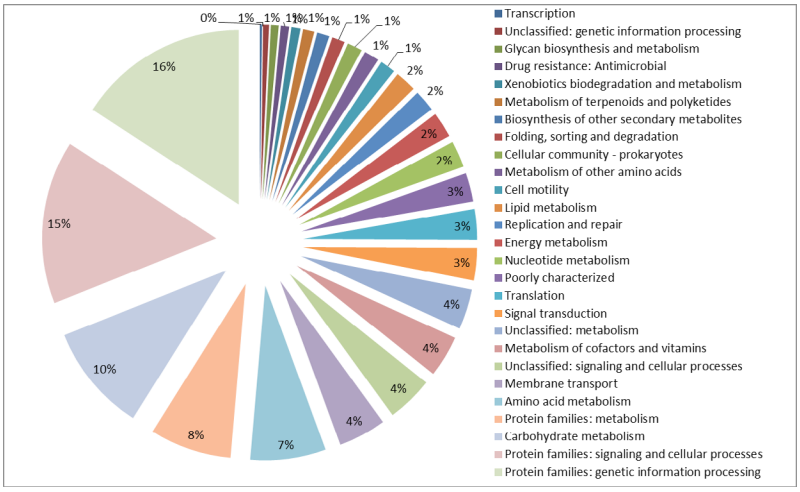

Figure 5: Pathway function associated with predicted proteins.

Taxonomic analysis

As mentioned before, the assembled PLSSC genome contains 30 rRNA genes. Further classification of these 30 rRNA genes revealed that there were equal number (10) of 5S rRNA, 16S rRNA and 23S rRNA genes. Upon alignment to the SILVA 16S Database [27], ten 16S rRNA sequences exhibited significant similarity among themselves. Moreover, phylogenetic tree constructed from 16S rRNAs showed that the assembled strain PLSSC is closely related to *Bacillus subtilis* subsp. subtilis with (Figure 6).

Figure 6: Phylogenetic tree showing relationship of *B. subtilis* strain PLSSC with other *Bacillus* species.

The mol% G+C is an old but useful molecular taxonomy method for bacterial classification. Based on its whole genome sequence, the mol% G+C for strain PLSSC was found to be 43.58%. The average nucleotide identity (ANI) between *B. subtilis* strain PLSSC and *B. subtilis* strain KCTC 3135 was found to be ~99%. BLASTN search of the assembled PLSSC genome against the RefSeq genome database belonging to *B. subtilis* (taxid: 1423) showed ~99% sequence homology with the genome of *B. subtilis* KCTC 3135.

Comparison of *B. subtilis* strain PLSSC and *B. subtilis* KCTC 3135 genomes by BLAST Ring Image Generator (BRIG) [29] has been represented in the form of a circos plot (Figure 7). The taxonomic analyses mentioned above identify the strain as *B. subtilis* and reveal its closeness to *B. subtilis* KCTC 3135.

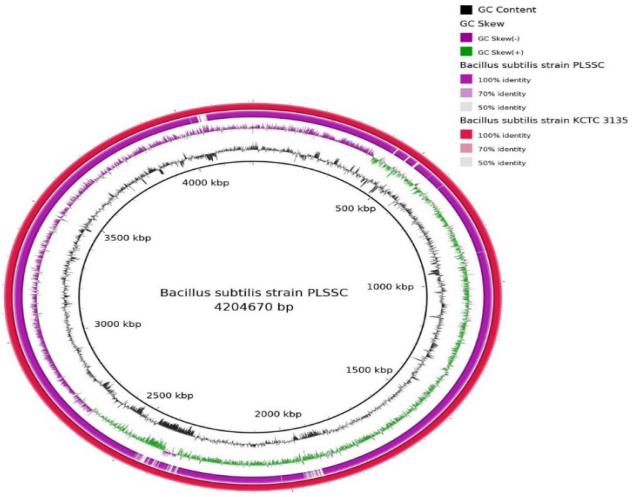

Figure 7: Circos plot comparison of *B. subtilis* strain PLSSC (size of genome is 4204670 bp) with *B. subtilis* strain KCTC 3135 (CP015375.1).

Antimicrobial resistance (AMR) genes

Homology search of the assembled PLSSC genome against the CARD database resulted in the identification of 717 putative antibiotic resistance genes. Identified genes were mainly involved in Defence mechanism (474), Signal transduction mechanisms; Transcription (136), General function prediction only (38), Carbohydrate transport and metabolism; Amino acid transport and metabolism; Inorganic ion transport and metabolism; General function prediction only (32), Cell wall/membrane/envelope biogenesis (10), Coenzyme transport and metabolism; Energy production and conversion (06), Replication, recombination and repair (2), Coenzyme transport and metabolism; Energy production and conversion (1), tunicamycin resistance (1), and Aminoglycoside 6-adenylyltransferase (1).

As per WHO, 2016 and EFSA, 2012 guidelines, assembled PLSSC genome was also screened for genes coding for critically important antimicrobials (CIAs) or highly important antimicrobials (HIAs). Genes for tunicamycin resistance (tmrB: DUT89_01190) and beta–lactamase (aadK: DUT89_13540) were identified in the genome. tmrB and aadK code for an ATP binding tunicamycin resistance protein and a streptomycin modifying enzyme respectively. No mobile element was identified in the flanking regions of these AMR genes which shows that these genes contribute to the intrinsic resistance and these is no risk of horizontal transfer of AMR genes.

Analysis of virulence factor genes

BLAST search against VFDB revealed that a total of 687 virulence factor proteins showed considerable homology with the assembled PLSSC genome. Further analysis with respect to the COG database suggested that these proteins were the products of non-classical virulence factors genes and they were related to diverse COG categories. Gene mining was done to identify genes linked to Diarrheal enterotoxin bceT, Haemolytic enterotoxin operon (hbl genes – hblA, hblC, hblD), Non-haemolytic enterotoxin operon (nhe ABC genes – nheA, nheB, nheC), Cytotoxin K (cytK), Enterotoxin FM (entFM), and Emetic Toxin Cereulide (cesB). None of these genes were present in the *B. subtilis* strain PLSSC assembled genome.

Analysis of biogenic amine producing genes

One gene coding for amino acid decarboxylase i.e. arginine decarboxylase (DUT89_07295) was identified with 100% homology against the biogenic amine producing proteins. In the qualitative detection of activity of arginine decarboxylase gene as described by Chang et al. [30], we did not observe change in the intensity of purple colour around the colony with and without the arginine (**Data not shown**). Above finding confirms that the arginine decarboxylase gene is either non-functional or not expressed at a sufficient level to produce detectable amounts of biogenic amine under the tested conditions.

Analysis of genome stability

Two putative prophage regions were identified at genomic coordinates: 1176131-1209853 (33722 bp) and 2006592-2145332 (138754 bp). However, post-analysis, it was observed that both these putative prophage regions did not contain the genes required for replication/transcription/packaging (topoisomerase, replisome, DNA-binding proteins), morphogenesis (accessory – tail fiber/whisker) or lysis (lysine). This suggests that the prophage sequences in the assembled genome were defective and non-functional.

IS finder [31] identified 12 insertion sites (IS element regions) in the assembled genome. Additionally, a search against the ACLAME database [32] revealed that 497 regions in the assembled genome showed significant hits (coverage >=50% and e-value <=1e-05). However, none of the genes coding for virulence factors had any mobile elements in their flanking regions, and therefore, do not pose any safety concerns to humans or animals. Two CRISPRs were identified from the assembled genome of *B. subtilis* strain PLSSC.

In silico analysis of probiotic features

It was observed that the assembled genome of *B. subtilis* strain PLSSC codes for important proteins which contribute to adhesion, acid and bile salt tolerance, and environmental stress resistance (Table 2).

Analysis using the antiSMASH program [33] showed that the assembled genome of *B. subtilis* strain PLSSC has 11 potential gene clusters responsible for the synthesis of secondary metabolites, including antimicrobial peptides, terpenes, fatty acids and others. The gene clusters are responsible for the synthesis of non-ribosomal cyclic lipopeptides: surfactin, bacillaene, fengycin, bacillibactin and bacilysin (Table S1).

Category	Probiotic feature	Protein ID (*B. subtilis* strain PLSSC genome)	Identified domain using CDD	Pfam ID
Adhesion to gut mucosa	Mucus binding protein	DUT89_04420	Cell wall Anchor	pfam00746
	Sortase	DUT89_04425	Sortase B protein	pfam02063
	Flagellin	DUT89_18655 DUT89_18015	Flagellin C- Terminal	pfam00700
		DUT89_18125	Flagellin N-Terminal	pfam00669
Acid tolerance	F0F1 ATP_synthase	DUT89_18940	ATP synthase subunit A	pfam00119
		DUT89_18930	ATP synthase subunit B	pfam00430
		DUT89_18920	ATP synthase subunit alpha	pfam00006
		DUT89_18910	ATP synthase subunit beta	pfam00006
		DUT89_18915	ATP synthase subunit gamma	pfam00231
		DUT89_18925	ATP synthase subunit delta	pfam00213
		DUT89_18905	ATP synthase subunit epsilon	pfam00401
	Amino acid decarboxylase	DUT89_07295	Orn/Arg decarboxylase	pfam01276
Bile tolerance	Sodium bile acid symporter	DUT89_09815	Bile acid sodium symporter	pfam01758
		DUT89_00370		pfam01758
				pfam01758

Environmental stress resistance	Universal stress	DUT89_20185	UspA	pfam00582
		DUT89_20185	Usp	pfam00582
	Chaperonins GroEL	DUT89_02670	Cpn60_TCP1	pfam00118
	Chaperonins GroES	DUT89_02665	Cpn10	pfam00166
	Heat Shock protein	DUT89_21575	Hsp33	pfam01430
	Heat resistance	DUT89_12975	GrpE	pfam01025
	Oxidative stress		PMSR	pfam01625
			PMSR	pfam01625
	Hyperosmotic stress	DUT89_12965	DnaJ	pfam01556
	Clp protease	DUT89_06800	CLP_protease	pfam00574
	Cold shock-like protein	DUT89_02150	Csp	pfam00575

Table 2: Proteins of *B. subtilis* strain PLSSC involved in probiotic features.

Discussion

WGS technology has been utilized to uncover the probiotic potentials of many bacterial [12,13,16,34,35] and fungal strains. Detailed annotation of assembled genomes facilitates the identification of genes contributing to the probiotic features as well as those which may raise concerns on the safety. EFSA emphasizes on the WGS analysis of bacterial and yeast strains before they can be used as probiotics. The present study focuses on safety aspects of probiotic *B. subtilis* strain PLSSC.

Despite immense progress in microbial taxonomy, 16S rRNA gene is still in use for the identification and classification of bacterial species. 16s rRNA of PLSSC showed its closeness with *Bacillus subtilis* subsp. subtilis. Comparison of PLSSC genome with the other bacterial genomes in the RefSeq genome database, revealed ~99% sequence homology with genome of the reference strain *B. subtilis* KCTC 3135. Moreover, G+C content of PLSSC (43.58%) was almost identical to the reported G+C content of *B. subtilis* KCTC 3135 (43.51%) [36].

Recent study [37] have reported that many commercial *Bacillus* probiotics contain mobile antibiotic resistance and toxin genes which may pose a health risk. PLSSC genome contains two AMR genes (tmrB and aadK) but absence of any mobile genetic element in their vicinity rule out the possibility of horizontal gene transfer. Moreover, enterotoxins (nheA,B,C, hblCDA, entFM, cytK, and bceT) and emetic toxin (cseB) genes were absent.

Assessment of genome stability is considered to be an important step in the safety workflow. On evolutionary time scale, bacterial genomes can be shaped by horizontal gene transfer and genetic rearrangements [38]. Mobile elements such as insertion sequences are responsible for the capture, accumulation and dissemination of antibiotic-resistance genes [39]. In the *B. subtilis* strain PLSSC assembled genome, none of the insertion sequences were present in the vicinity of putative virulence factors or antibiotic resistance genes. Although putative virulence factors genes were identified, they cannot be considered as harmful as majority of them were related to transport mechanism. Prophage sequence analysis revealed the presence of two prophage regions but their annotation suggested them to be defective and non-functional. CRISPRs are short direct repeats (23- 47 bp in length) found in the DNA of several bacteria (~40% of sequenced bacterial genomes) that play a role in controlling genome stability by providing immunity against previously encountered bacteriophages and plasmids [39]. Each of these repeats is separated by spacers of similar length that are unique for each of the genomes. The presence of two CRISPR systems in the *B. subtilis* strain PLSSC genome, indicates an advantage in promoting genome stability by acting as a barrier to the entry of foreign DNA elements.

For *In-silico* analysis of probiotic traits, the assembled PLSSC genome was searched for the presence of probiotic related proteins. Table 2 shows the list of probiotics related proteins. Genes related to extracellular structure were identified which could be correlated to the adhesion property of the probiotic. Mucus binding protein (MUB), known to facilitate the attachment of probiotic bacteria

[40] to the host mucus was found as a single copy (DUT89_04420) on the genome. Moreover, flagellar hook associated proteins (DUT89_18015 and DUT89_18125) responsible for mucin specific adhesion [41] were also identified. In order to thrive in the acidic gut environment, *B. subtilis* strain PLSSC can use F0F1 ATP synthase and aminoacid decarboxylase, which play a role in maintaining H+ homeostasis and alkalinization of the cytosol [15, 42]. We also identified sodium bile acid symporter, which has a role in providing resistance to bile salts. Environmental stress such as high or low temperatures may lead to loss of function in some proteins or nucleic acids, inhibition of metabolism or disruption of cellular activities. At high temperature, increase in membrane fluidity may lead to disruption of cellular activities [43]. The presence of proteins such as chaperonins, heat-shock proteins in the PLSSC genome confers resistance to temperature stress, thereby preventing denaturation and degradation [43]. Moreover, the presence of cold shock-like protein may help PLSSC to survive at low temperature. Similar to *Bacillus claussi* and *Pediococcus pentosaceus* 1101, PLSSC genome contains a gene for chaperone DnaJ which participates actively in the response to hyperosmotic shock [44,45]. Probiotic bacteria have been reported to produce secondary metabolites with potential antimicrobial activity which inhibit the growth of pathogens in the gut [46]. We found that PLSSC genome contains gene clusters involved in the synthesis of secondary metabolites having antimicrobial properties. These antimicrobials can reduce the adherence of pathogenic microbes in the gut and thereby preventing the dysbiosis. The In silico analysis of the whole genome sequence indicated *B. subtilis* strain PLSSC be a safe probiotic.

Conclusion

The de novo assembled genome of *B. subtilis* strain PLSSC was generated using a hybrid assembly of nanopore long reads and Illumina short reads. Sequence homology between the de novo assembled genome of *B. subtilis* strain PLSSC with respect to the reference strain KCTC 3135 was ~99%. Through the whole genome based safety studies, the strain was analyzed for both, genetic elements that are absent, and elements that are present but not a safety concern. The full genome information confirms that the risk-associated genetic elements found in the genome of *B. subtilis* strain PLSSC do not raise any safety concern. The lack of IS and presence of CRISPR elements in the PLSSC genome may maintain its stability. Moreover, the presence of anchorage-related proteins increase their colonization and eventually, reduce pathogenic adherence in the gut. In conclusion, based on the genotypic properties of *B. subtilis* strain PLSSC, it is safe to be used as a probiotic.

Acknowledgements

The authors would like to thank Mr. VL Rathi, and Mr. MM Kabra, Advanced Enzyme Technologies Limited for providing the laboratory facilities.

Funding: This study was funded by Advanced Enzyme Technologies, Limited, Thane, India.

Author contributions: Conceptualization: Dina Saroj; methodology: Dina Saroj; analysis of data and writing: Vikash Kumar and Pruthvi Upadhyaya; review and editing: Dina Saroj, Vikash Kumar and Pruthvi Upadhyaya. All authors read and approved the manuscript

Availability of data: The whole genome sequence of *Bacillus subtilis* strain PLSSC and corresponding annotation are available in NCBI GenBank (accession number NZ_CP031129).

References

1. Li A, Wang M, Zhang Y, Lin Z, Xu M, Wang L, Kulyar MF-e-A and Li J (2023) Complete genome analysis of *Bacillus subtilis* derived from yaks and its probiotic characteristics. Front Vet Sci 9:1099150.
2. Ren J, Yu D, Li N, Liu S, Xu H, Li J, He F, Zou L, Cao Z, Wen J (2023) Biological Characterization and Whole-Genome Analysis of *Bacillus subtilis* MG-1 Isolated from Mink Fecal Samples. Microorganisms 11:2821.
3. Hong HA, Duc LH, Cutting SM (2005) The use of bacterial spore formers as probiotics. FEMS microbiology reviews29:813-835.
4. Hill C, Guarner F, Reid G, Gibson GR, Merenstein DJ, Pot B, Morelli L, Canani RB, Flint HJ, Salminen S, Calder PC (2014) The international scientific association for probiotics and prebiotics consensus statement on the scope and appropriate use of the term probiotic. Nat Rev Gastroenterol Hepatol 11:506–514.
5. Ahire JJ (2012) Studies on probiotic microorganism(s) and its biogenic metabolite(s). (Doctoral dissertation, Ph. D. Dissertation, North Maharashtra University, India).
6. Lee NK, Kim WS, Paik HD (2019) *Bacillus* strains as human probiotics: characterization, safety, microbiome, and probiotic carrier. Food science and biotechnology 28:1297-1305.
7. Errington J, Aart LTV (2020) Microbe Profile: *Bacillus subtilis*: model organism for cellular development, and industrial workhorse. Microbiology 166:425-427.
8. Bremer E, Calteau A, Danchin A, Harwood C, Helmann JD, Médigue C et al (2023) A model industrial workhorse: *Bacillus subtilis* strain 168 and its genome after a quarter of a century. Microbial Biotechnology 16:1203–1231.
9. Takimoto T, Hatanaka M, Hoshino T, Takara T, Tanaka K, Shimizu A, Morita H, Nakamura T (2018) Effect of *Bacillus subtilis* C-3102 on bone mineral density in healthy postmenopausal Japanese women: a randomized, placebo-controlled, double-blind clinical trial. Bioscience of Microbiota, Food and Health 37:87-96.
10. Ramachandran R, Chalasani AG, Lal R, Roy U (2014) A broad-spectrum antimicrobial activity of *Bacillus subtilis* RLID 12.1. Sci. World J. 968487.

11. European Food Safety Authority (EFSA) (2021) EFSA statement on the requirements for whole genome sequence analysis of microorganisms intentionally used in the food chain. EFSA Journal 2021, 19:e06506.

12. Saroj DB, Ahire JJ, Shukla R (2023) Genetic and phenotypic assessments for the safety of probiotic Bacillus clausii 088AE. 3 Biotech 13:238.

13. Ahire JJ, Kashikar MS, Madempudi RS (2021) Comparative accounts of probiotic properties of spore and vegetative cells of *Bacillus clausii* UBBC07 and in silico analysis of probiotic function. 3 Biotech 11:116.

14. Jackson SA, Schoeni JL, Vegge C, Pane M, Stahl B, Bradley M, Goldman VS, Burguière P, Atwater JB, Sanders ME (2019) Improving end-user trust in the quality of commercial probiotic products. Front Microbiol 10:739.

15. Lakshmi SG, Jayanthi N, Saravanan M, Ratna Sudha M (2017) Safety assessment of *Bacillus clausii* UBBC07, a spore forming probiotic. Toxicol Rep 4:62–17.

16. Saroj D, Gupta AK (2020) Genome based safety assessment for *Bacillus coagulans* strain LBSC (DSM 17654) for probiotic application. Int J Food Microbiol 318:108523.

17. Aleksey VZ, Guillaume M, Daniela P, Michael R, Steven LS, James AY (2013) The MaSuRCA genome assembler. Bioinformatics 21:2669–2677.

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

19. Goris J, Konstantinidis KT, Klappenbach JA, Coenye T, Vandamme P, Tiedje JM (2007) DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. Int J Syst Evol Microbiol 57:81–91.

20. Jia B, Raphenya AR, Alcock B, et al (2017) CARD: expansion and model-centric curation of the comprehensive antibiotic resistance database. Nucleic Acids Res 45:D566-D573.

21. Tatusov RL, Galperin MY, Natale DA, Koonin EV (2000) The COG database: a tool for genome-scale analysis of protein functions and evolution. Nucleic Acids Res 28:33-6.

22. Chen L, Yang J, Yu J et al (2004) VFDB: a reference database for bacterial virulence factors. Nucleic Acids Res 33 (suppl_1): D325-8.

23. Salvetti E, Orru L, Capozzi V, Martina A, Lamontanara A, Keller D, Cash H, Felis GE, Cattivelli L, Torriani S, Spano G (2016) Integrate genome-based assessment of safety for probiotic strains: *Bacillus coagulans* GBI-30, 6086 as a case study. Appl Microbiol Biotechnol 100:4595–4605.

24. Arndt D, Grant JR, Marcu A, Sajed T, Pon A, Liang Y, Wishart DS (2016) PHASTER: a better, faster version of the PHAST phage search tool. Nucleic Acids Res 44:W16-21.

25. Couvin D, Bernheim A, Toffano-Nioche C et al (2018) CRISPRCasFinder, an update of CRISRFinder, includes a portable version, enhanced performance and integrates search for Cas proteins. Nucleic Acids Res 46:W246-W251.

26. Khatri I, Sharma S, Ramya TNC, and Subramaninan S (2016) Complete genome of *Bacillus coagulans* S-lac and *Bacillus subtilis* TO-AJPC, two phylogenetically distinct probiotics. PLoS One 11:e0156745.

27. Pruesse E, Quast C, Knittel K, et al (2007) SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. Nucleic Acids Res 35:7188-96.

28. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410.

29. Alikhan NF, Petty NK, Ben Zakour NL, Beatson SA (2011) BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. BMC Genom 12:402.

30. Chang M, Chang HC (2012) Development of a screening method for biogenic amine producing *Bacillus* spp. International Journal of Food Microbiology 153: 269–274.

31. Siguier P, Pérochon J, Lestrade L, Mahillon J, Chandler M (2006) ISfinder: the reference centre for bacterial insertion sequences. Nucleic Acids Res 34:D32–D36.

32. Leplae R, Lima-Mendez G, Toussaint A (2009) ACLAME: a classification of mobile genetic elements, update 2010. Nucleic Acids Res 38:D57–D61.

33. Medema MH, Blin K, Cimermancic P, de Jager V, Zakrzewski P, Fischbach MA, Weber T, Takano E, Breitling R (2011) antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. Nucleic Acids Res 39:W339-46.

34. Sylvere N, Mustopa AZ, Budiarti S, Meilina L, Hertati A, Handayani I (2023) Whole-genome sequence analysis and probiotic characteristics of *Lactococcus lactis* Subsp. lactis strain Lac3 isolated from traditional fermented buffalo milk (Dadih). J Genet Eng Biotechnol 21:49.

35. Dhanya Raj CT, Suryavanshi MV, Kandaswamy S, Ramasamy KP, James RA (2023) Whole genome sequence analysis and in-vitro probiotic characterization of *Bacillus velezensis* FCW2 MCC4686 from spontaneously fermented coconut water. Genomics 115(4):110637.

36. Ahn S, Jun S, Ro HJ, Kim JH, Kim S (2018) Complete genome of *Bacillus subtilis* subsp. subtilis KCTC 3135T and variation in cell wall genes of B. subtilis strains. J Microbiol Biotechnol 28: 1760-1765.

37. Cui Y, Wang S, Ding S, Shen J, Zhu K (2020) Toxins and mobile antimicrobial resistance genes in *Bacillus* probiotics constitute a potential risk for One Health. J Hazard Mater 382:121266.

38. Darmon E, David RFL (2014) Bacterial genome instability. Microbiol Mol Biol Rev 78:1–39.

39. Partridge SR, Kwong SM, Firth N, Jensen SO (2018) Mobile Genetic Elements Associated with Antimicrobial Resistance. Clin Microbiol Rev 31:10-1128.

40. Richter C, Chang JT, Fineran PC (2012) Function and regulation of clustered regularly interspaced short palindromic repeats (CRISPR)/ CRISPR associated (Cas) systems. Viruses 4:2291-311.

41. Muscariello L, De Siena B, Marasco R (2020) *Lactobacillus* Cell Surface Proteins Involved in Interaction with Mucus and Extracellular Matrix Components. Curr Microbiol 77:3831-3841.

42. Sánchez B, López P, González-Rodríguez I, Suárez A, Margolles A, Urdaci MC (2011) A flagellin-producing *Lactococcus* strain: interactions with mucin and enteropathogens. FEMS Microbiol Lett 318:101-7.

43. Cotter PD, Hill C (2003) Surviving the acid test: responses of Gram positive bacteria to low pH. Microbiol Mol Biol Rev 67:429–453.

44. Escobar-Sánchez M, Carrasco-Navarro U, Juárez-Castelán C, Lozano-Aguirre Beltrán L, Pérez-Chabela ML, Ponce-Alquicira E (2022) Probiotic Properties and Proteomic Analysis of *Pediococcus pentosaceus* 1101. Foods 12:46.

45. Khatri I, Gaurav S, Srikrishna S (2019) Composite genome sequence of *Bacillus clausii*, a probiotic commercially available as Enterogermina®, and insights into its probiotic properties. BMC Microbiol 19:307.

46. Harwood CR, Mouillon JM, Pohl S, Arnau J (2018) Secondary metabolite production and the safety of industrially important members of the *Bacillus subtilis* group. FEMS Microbiol Rev 42(6):721-738.