Infection and Drug Resistance

Dovepress
Taylor & Francis Group

Open Access Full Text Article

ORIGINAL RESEARCH

Genomic Investigation of a Bacillus subtilis Strain Sourced from Commercially Available Milk Powder in China Reveals Potential Risk Factors

Ling Zhu[1,*], Keyu Chen[2,*], Liu Xu[2,*], Afeng Wang[2], Haojian Gan[2], Junhong Sun[2], Yujie Wu[2], Yutong Li[2], Yingying Guo[3], Yanfeng Yi[4], Xinhua Qiang[1], Jie He[5], Hongchang Zhou[2,6], Yibin Lin[2,6]

[1]Clinical Laboratory, First People's Hospital Affiliated to Huzhou University, Huzhou, 313000, People's Republic of China; [2]School of Medicine, Huzhou University, Huzhou, 313000, China; [3]School of Life, Huzhou University, Huzhou, 313000, People's Republic of China; [4]School of Life and Health Sciences, Huzhou College, Huzhou, 313000, People's Republic of China; [5]Infectious Diseases Department, First People's Hospital Affiliated to Huzhou University, Huzhou, 313000, China; [6]Key Laboratory for Precise Prevention and Control of Major Chronic Diseases, Huzhou University, Huzhou, 31300, People's Republic of China

*These authors contributed equally to this work

Correspondence: Hongchang Zhou; Yibin Lin, Email zhouhc529@zjhu.edu.cn; yybbll@gmail.com

Background: Milk powder is a key food source, especially for infants and vulnerable groups. However, Bacillus contamination during production, storage, or handling can cause spoilage, quality issues, or health risks. This study identified and isolated *Bacillus subtilis* from commercially available Chinese milk powder.

Methods: A pure colony of *Bacillus subtilis* was isolated from an LB agar plate supplemented with milk powder and identified using mass spectrometry. The genome of this strain was sequenced using third-generation sequencing technology. Following assembly, the genome was functionally annotated and subjected to comprehensive bioinformatic analysis.

Results: Genomic analysis classified the strain as *Bacillus subtilis* via MALDI-TOF and ANI (98.82% with *B. subtilis* AMR1). Its genome features a 4.26 Mbp chromosome and 97.6 kbp plasmid encoding 4,539 genes, including virulence factors (209 genes), antibiotic resistance genes (19 genes), and carbohydrate-active enzymes (253 genes). Key virulence mechanisms include immune modulation, stress adaptation, toxin production, and biofilm formation. Antibiotic resistance involves efflux pumps (eg, *qacJ, bmr*), enzymatic inactivation (eg, *FosBx1, aadK*), and target modification (eg, *vanG* cluster, *tet(45)*). Phylogenetically (LIN78), the strain clusters with foodborne *B. subtilis* isolates (eg, from Korean gochujang and soybean), diverging from *B. cereus* and environmental *Bacillus* clades. Comparative genomics revealed 53 LIN78-specific genes, encompassing defense mechanisms and mobile elements, and synteny in all homologs except *B. subtilis* ATCC 11774. Genomic islands, CRISPR arrays, and recombination-associated repeats indicate adaptive evolution.

Conclusion: This study characterizes *Bacillus subtilis* LIN78, a genomically plastic strain isolated from Chinese milk powder. It exhibits adaptation to food environments via horizontal gene transfer, stress tolerance, and spoilage traits, while carrying antimicrobial resistance risks and potential biotechnological applications. The findings necessitate genomic monitoring to manage food safety, resistance spread, and leverage its dual role as both a spoilage organism and source of bioactive compounds..

Keywords: Bacillus subtilis, whole genome DNA sequencing, virulence factors, resistant genes, milk powder

Introduction

Milk powder serves as a vital nutritional resource, especially for infants, children, and immunocompromised populations. Despite its widespread consumption, microbial contamination can occur during production, processing, or storage, raising concerns about potential health hazards. While rigorous sterilization techniques—including pasteurization and spray drying—are employed during manufacturing, heat-resistant endospores from spore-forming bacteria like *Bacillus* species enable these organisms to persist in the final product. Such contamination can compromise product integrity, accelerate spoilage, or trigger

Received: 1 June 2025
Accepted: 19 August 2025
Published: 25 August 2025

opportunistic infections in susceptible groups.[1] Comprehensive analysis of microbial communities in commercially available milk powder is critical for safeguarding food safety and mitigating risks to public health.

Bacillus subtilis is a Gram-positive bacterium extensively studied for its industrial and probiotic utility, demonstrating considerable potential in applications ranging from antimicrobial compound synthesis to gut health promotion and livestock feed optimization.[2,3] Significantly, species of this genus are among the spore-formers frequently identified in milk powder due to their endospore resilience.[4–6] While many strains are recognized as beneficial, concerns persist regarding the presence of virulence determinants or antibiotic resistance genes in specific isolates, which may compromise their safety in food systems.[7–9] Furthermore, horizontal gene transfer (HGT), mediated by plasmids or mobile genetic elements (MGEs), poses significant risks by enabling the dissemination of detrimental genetic traits among bacteria. Consequently, comprehensive genomic characterization of food-derived *B. subtilis* strains is imperative to differentiate probiotic candidates from those with opportunistic pathogenic potential.

Despite the well-documented utility and potential risks associated with *B. subtilis* and its common occurrence in dairy products, genomic data for strains isolated from Chinese commercial milk powder remain severely limited. This critical gap impedes accurate risk assessment of locally circulating strains, as existing studies lack context-specific insights into genetic markers (eg, virulence, antimicrobial resistance, or plasmid-mediated HGT) within China's dairy supply chain.[10,11] Key limitations of prior work include: (i) Inadequate characterization of genomic architectures (especially plasmid content) of strains from Chinese sources, despite the recognized role of MGEs in risk dissemination; (ii) Inability to evaluate region-specific safety profiles or biotechnological potential due to data scarcity. This study addresses these gaps by presenting a comprehensive genomic characterization of a *B. subtilis* strain obtained from Chinese retail milk powder, with particular emphasis on its plasmid content. Given the propensity of plasmids and other MGEs to harbor genes associated with antimicrobial resistance, virulence, or metabolic adaptation and facilitate HGT—modulating bacterial behavior in food matrices and hosts—our sequencing and annotation establish a foundation for elucidating the strain's genetic risks or benefits.

To achieve this characterization and risk/benefit assessment, this study employs whole-genome sequencing (WGS) to comprehensively define the genomic architecture of the isolated *Bacillus subtilis* strain, comprising both its chromosomal and plasmid sequences. Key objectives include identifying functional genes associated with stress resistance, sporulation, virulence, and antibiotic resistance, as well as examining plasmid features—such as mobility, stability, and cargo gene content—to evaluate their role in bacterial adaptation and the risk of disseminating detrimental genes. Additionally, the work aims to assess the strain's safety profile and biotechnological potential, critically evaluating whether it poses health risks or holds promise for probiotic or industrial applications.

The findings will enhance frameworks for microbial risk assessment in dairy production and inform regulatory policies or strain selection for probiotic development. Furthermore, the genomic data generated will serve as a foundational reference for future research on *Bacillus* ecology and evolution within food systems, providing critical insights into the safety and functional attributes of *B. subtilis* in consumer products.

Materials and Methods

Isolation, Culture, and Identification of Bacterial Strains

This study investigated a *Bacillus subtilis* strain sourced from commercially available milk powder procured in March 2023. The product was acquired from a supermarket in Huzhou, Zhejiang Province, China, and maintained at 4°C until laboratory processing. For microbial isolation, 1 gram of milk powder was mixed with 9 mL of sterile phosphate-buffered saline (PBS) to create a uniform suspension. Sequential dilutions (10^{-1} to 10^{-5}) were prepared and spread onto Luria-Bertani (LB) agar plates, followed by aerobic incubation at 37°C for 24–48 hours. Distinct colonies displaying characteristic *Bacillus* traits—irregular margins, dry texture, and opaque appearance—were selected for purification via subculturing. Gram-staining analysis confirmed Gram-positive, rod-shaped cells capable of endospore formation, aligning with *Bacillus* genus features. Species-level identification was achieved through MALDI-TOF mass spectrometry. Confirmatory biochemical profiling demonstrated catalase production, starch hydrolysis capability, and a positive Voges–Proskauer reaction, collectively substantiating the classification as *B. subtilis*.

Genomic DNA Extraction, Sequencing and Assembly

Bacillus subtilis cultures were incubated overnight on sheep blood agar plates at 37°C, after which bacterial biomass was harvested from the agar surface. Genomic DNA isolation was conducted using the TIANamp Bacteria DNA Kit (TIANGEN Biotech, Beijing), following the supplier's protocol. DNA integrity and concentration were assessed via 1% agarose gel electrophoresis visualized under UV light, using a quantitative DNA ladder to estimate concentration. Samples demonstrating intact genomic DNA bands with estimated concentrations within the optimal range specified by the BGI sequencing protocols (typically >20 ng/μL for PacBio libraries and >50 ng/μL for DNBSEQ libraries) were submitted for whole-genome sequencing (WGS). WGS was carried out at the Beijing Genomics Institute (BGI, Shenzhen, China) using a hybrid approach: long-read data were generated on the PacBio Sequel II platform with four SMRT Cell Zero-Mode Waveguide arrays, while short-read sequencing was performed on the DNBSEQ platform. Raw PacBio subreads shorter than 1 kb were filtered out prior to assembly. A self-corrected circular consensus sequence dataset was produced using Canu software (v2.0), enabling the construction of draft genomic unitigs (contiguous assemblies of non-overlapping fragments). To refine sequence precision, single-nucleotide errors were corrected using the Genome Analysis Toolkit (GATK; Broad Institute).

Genome Component Prediction

Genomic annotation of the *Bacillus subtilis* LIN78 assembly was conducted using Glimmer3 (accessible at http://www.cbcb.umd.edu/software/glimmer/), a Hidden Markov Model-based tool for gene prediction. Non-coding RNA elements, including tRNA, rRNA, and sRNA, were identified through tRNAscan-SE,[12] RNAmmer,[13] and the Rfam database.[14] Tandem repeat regions were detected with the Tandem Repeat Finder (http://tandem.bu.edu/trf/trf.html), while minisatellite and microsatellite sequences were classified according to repeat unit length and frequency. Genomic islands were analyzed via the Genomic Island Suite of Tools (GIST) (http://www5.esu.edu/cpsc/bioinfo/software/GIST/), incorporating algorithms such as IslandPath-DIMOB, SIGI-HMM, and IslandPicker. Prophage elements were predicted using the PHAge Search Tool (PHAST) (http://phast.wishartlab.com), and CRISPR arrays were identified employing CRISPRFinder.[15]

Gene Annotation and Protein Classification

Functional annotation was initiated through identification of top matches using the BLAST alignment tool. For comprehensive functional characterization, seven specialized databases were employed: Kyoto Encyclopedia of Genes and Genomes (KEGG),[16] Clusters of Orthologous Groups (COG),[17] the Non-Redundant Protein Database (NR), Swiss-Prot,[18] Gene Ontology (GO),[19] TrEMBL,[20] and EggNOG.[21] To assess pathogenicity and antimicrobial resistance, four additional databases were utilized. Virulence-associated genes and antibiotic resistance determinants were identified by analyzing core datasets from the Virulence Factors of Pathogenic Bacteria (VFDB)[22] and the Antibiotic Resistance Genes Database (ARDB).[23] Carbohydrate-active enzyme profiling was conducted using the CAZY database,[24] while the Comprehensive Antibiotic Resistance Database (CARD; https://card.mcmaster.ca/analyze)[25] provided insights into resistance mechanisms.

Comparative Genomics and Phylogenetic Analysis

Genomic data from five previously studied *Bacillus subtilis* strains—AMR1 (BSAMR1, NZ_CP050319), A10 (BSA10, NZ_CP054050), ATCC11774 (BS11774, NZ_CP026010), BS16045 (BS16045, NZ_CP017112), and DKU_NT_02 (BSDKU, NZ_CP022890)—served as reference genomes for comparative analysis (Table S1). Syntenic relationships between *B. subtilis* LIN78 and these reference strains were analyzed using MUMmer and BLAST alignment tools. Core and pan-genomic profiles were generated through CD-HIT protein clustering software, applying thresholds of 50% pairwise sequence identity and a 0.7 amino acid length difference ratio. Gene family classification involved a multi-step computational pipeline: initial protein sequence alignment via BLAST, redundancy removal using SOLAR, and hierarchical clustering with Hcluster_sg. Phylogenetic reconstruction was implemented through the TreeBeST platform employing neighbor-joining (NJ) methodology.

For broader evolutionary context, 25 complete *Bacillus* genomes closely related to LIN78 were retrieved from GenBank. Protein sequence alignment utilized the MUSCLE algorithm, while nucleotide alignment was conducted in MEGA11. A maximum likelihood phylogenetic tree was subsequently generated using MEGA11's built-in evolutionary modeling toolkit.

Results

Biological and Phylogenetic Characteristics of B. subtilis LIN78

The bacterial strain LIN78, identified as *Bacillus subtilis*, was isolated from commercial milk powder samples in China and displayed proteolytic activity toward milk components. Strain LIN78 can degrade milk powder, forming a clear halo, whereas *B. subtilis* subsp. subtilis str. 168 cannot degrade milk powder (Table S 1a). Neither strain LIN78 nor *B. subtilis* subsp. subtilis str. 168 produces a hemolytic zone on sheep blood agar plates. On sheep blood agar, strain *B. subtilis* subsp. subtilis str. 168 forms white, dry, flat colonies with irregular margins, while strain LIN78 forms raised, white colonies with irregular margins, a rough surface texture, and a sticky consistency (Figure 1B–C). Strain LIN78 exhibits strong spreading ability; when cultured on 0.3% agar plates at 37°C for 12 hours, it can spread to cover the entire plate (90 mm) (Figure 1D). Using crystal violet staining, it can be observed that the LIN78 strain forms spores located centrally within the bacterial cells (Figure 1E). Taxonomic classification as *B. subtilis* was verified through MALDI-TOF mass spectrometry. Whole-genome comparison with *B. subtilis* AMR1 demonstrated 98.82% average nucleotide identity (ANI), a standard measure of genomic relatedness, confirming their phylogenetic congruence at the species level. Protein functional annotation via the Non-Redundant Protein Database revealed predominant alignment of LIN78-derived proteins with *B. subtilis* orthologs, reinforcing taxonomic assignment (Table S2). This multi-method analysis establishes LIN78 as a *B. subtilis* strain harboring significant antimicrobial resistance determinants.

Genome Features of B. subtilis LIN78

As outlined in Table 1, *Bacillus subtilis* LIN78 exhibits a circular chromosome measuring 4,261,842 base pairs (bp) (Figure 2A) alongside a 97,563 bp plasmid designated pLIN78 (Figure 2B). Coding sequences constitute approximately 87.95% of the genome, encompassing 4,539 protein-coding genes with an average length of 845 bp. Distribution analysis reveals 4,445 chromosomal genes and 94 plasmid-borne genes. The chromosomal complement includes 27 rRNA operons (containing 23S, 5S, and 16S subunits) and 38 small RNA (sRNA) elements. Bioinformatic characterization detected 28 genomic islands exclusively localized to the chromosome (Table S3). Furthermore, four intact prophage sequences were annotated within the chromosomal DNA (Table S4). No mobile genetic elements or virulence-associated loci were identified on the plasmid.

Repetitive DNA elements in LIN78 include 117 tandem repeats (TRs) with motif lengths of 6–253 bp, collectively representing 0.4623% of the genome (Table S5). Additionally, 44 minisatellite sequences (15–60 bp repeats, 0.1053% of the genome) and four microsatellite sequences (6–9 bp repeats, 0.093% of the genome) were observed (Table S5). These repetitive elements may indicate potential genetic recombination events between *B. subtilis* LIN78 and related species. Two putative CRISPR regions were also identified: a 105 bp locus (913,929–914,034 bp) and a 107 bp locus (3,223,-972–3,224,079 bp) on the chromosome (Table S6).

Overview of Gene Annotation Analysis of B. subtilis LIN78

Functional annotation of the *Bacillus subtilis* LIN78 genome offers a detailed profile of its genetic features and potential roles. Among the 4,539 annotated genes (99.42% of the genome), database-specific analyses revealed distinct functional categorizations (Table 2 and Table S7). The IPR database assigned functional domains to 3,766 genes (82.96% of the genome), while SWISSPROT matched 3,936 genes (86.71%) to experimentally validated proteins. Functional classification via COG identified 3,302 genes (72.74%) (Figure 2C), and GO terms categorized 2,578 genes (56.79%) into molecular functions, biological processes, and cellular components (Figure 2D). KEGG pathways mapped 2,734 genes (60.23%), and the NR database annotated 4,510 genes (99.36%) (Figure 2E). Notably, 209 virulence-associated genes (4.6%) were cataloged in VFDB, and 30 genes (0.66%) linked to antibiotic resistance were identified in ARDB. Additionally, CAZy annotations highlighted 253 genes (5.57%) involved in carbohydrate-active enzyme production, and T3SS analysis assigned 475 genes (10.46%) to effector protein secretion systems. Together, these annotations elucidate the genomic architecture and functional capabilities of *B. subtilis* LIN78, underscoring its metabolic versatility, adaptive mechanisms, and potential biotechnological applications.

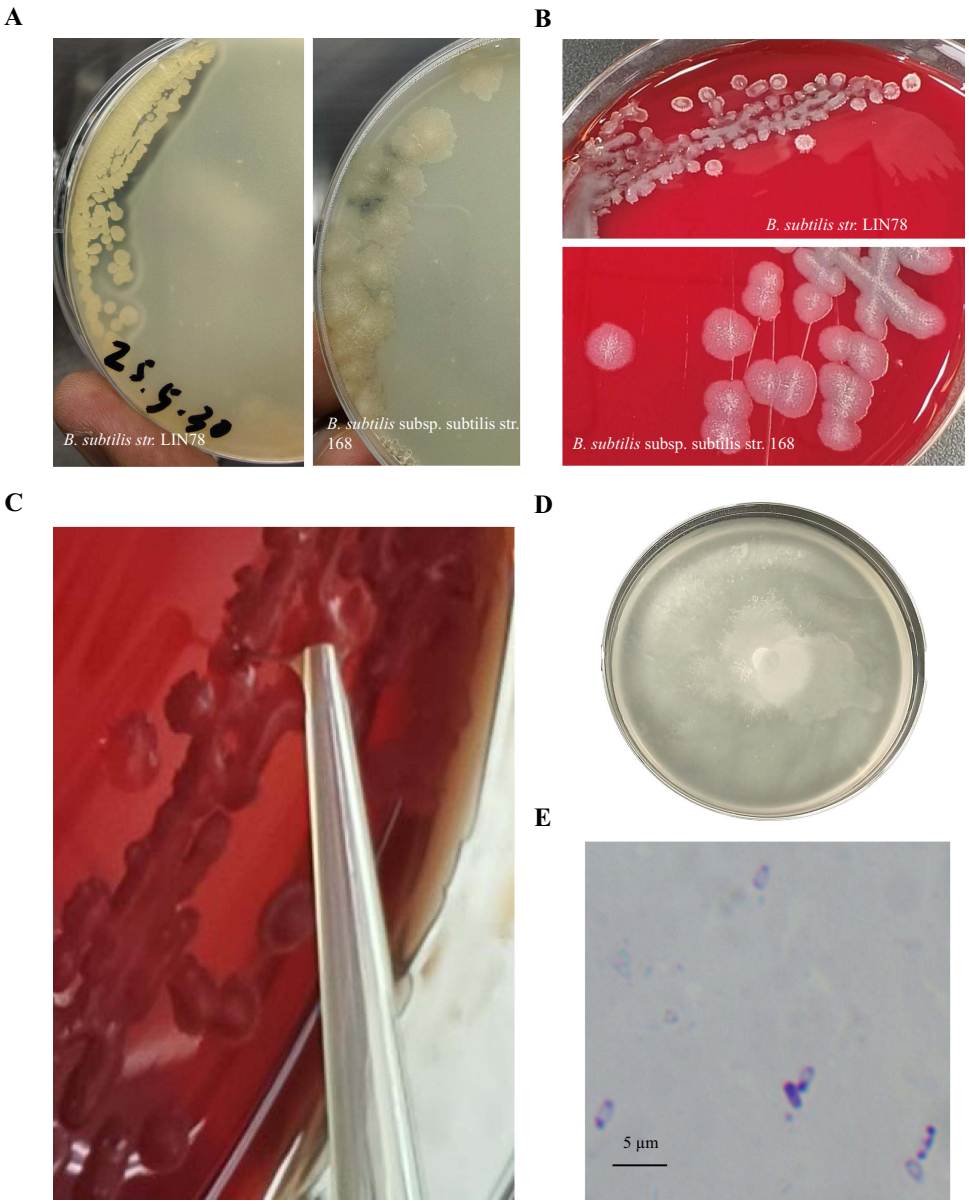

Figure 1 Biological characteristics of *B. subtilis* LIN78. (**A**) The *Bacillus subtilis* strain LIN78 colonies on LB agar plate containing milk, demonstrating milk-degrading capabilities. (**B**) LIN78 colonies on sheep blood agar plates. (**C**) Mucoviscosity qualitative measurement by string test. (**D**) Growth of LIN78 on LB agar plate containing 0.3% (w/v) agar exhibits motility. (**E**) LIN78 strained with crystal violet. LIN78 can produce endospores that are oval-shaped and located centrally in the cell.

Genomic Analysis of Virulence-Associated Genes in B. subtilis LIN78

Genomic profiling of *B. subtilis* LIN78 revealed 209 genes linked to virulence factor expression or prediction (Figure 3 and Table 3, and Table S8). Functional categorization highlighted critical pathways supporting pathogenicity and survival. Immune evasion mechanisms (61 genes/proteins) featured *lpxA/glmU, cpsA/uppS*, and *galU*, which modulate cell wall biosynthesis, capsular polymer assembly, and host defense evasion. Stress adaptation systems (13 elements) included oxidative stress resistance mediators (*katA*, catalase) and *clp* proteases for protein quality control. Host adhesion (12 genes) involved surface proteins (*tufA, groEL*, *fbpA*) and chaperones promoting bacterial attachment. Toxin delivery pathways (9 components), such as *CBU_1566* and *xcpR*, were implicated in effector secretion. Metabolic flexibility was driven by 52 nutrient acquisition genes, including amino acid biosynthesis (*lysA*), heme production (*hemL*), and substrate uptake (*sugC*). Motility-related genes (*fliC, flrA*; n = 22) and biofilm regulators (2 genes) supported colonization, while cytolytic exotoxins (*hlyA*; 5 genes) enhanced tissue damage. Regulatory networks (15 elements) encompassed two-component systems (*phoP, virR*) and stress-responsive regulators (*relA*).

https://doi.org/10.2147/IDR.S544153

Table 1 The General Genome Features of Bacillus subtilis LIN78

	LIN78	
Category	Chromosome	pLIN78
DNA size (bp)	4,261,842	97,563
GC Content (%)	43.43	34.56
Gene Number	4,445	94
Gene Length	3,764,313	69,603
Number of tRNA	86	0
Number of rRNA (16S-23S-5S)	27	0
Number of sRNA	38	0
Islands	28	0
Prophage	4	0
CRISPR	2	0

Post-translational modifications (7 genes, eg, *srtA* for sortase activity) and exoenzymes (9 genes, including *nuc* nuclease) processed virulence factors, with *secA2* aiding secretion. Collectively, these systems enable pathogenesis through immune evasion, stress resilience, host interaction, and metabolic adaptation.

Secondary Metabolite Gene Clusters in B. subtilis LIN78

Genome mining of *B. subtilis* LIN78 using anti-SMASH (version 7)[26] revealed 13 secondary metabolite biosynthetic gene clusters (Figure 4). These clusters encompassed diverse pathways, including non-ribosomal peptide synthetase (NRPS), lanthipeptide-class-I, terpene, trans-AT polyketide synthase (trans-AT-PKS), type III polyketide synthases (T3PKS), non-ribosomal peptide (NRP) metallophore, cyclodipeptide synthase (CDPS), and sactipeptide. Notably, NRPS-associated clusters were the most abundant among those identified. The 13 gene clusters suggest *B. subtilis* LIN78 can produce multiple types of secondary metabolites, which are compounds often involved in ecological interactions (eg, antibiotics, siderophores). The high number of NRPS clusters implies that this strain specializes in synthesizing non-ribosomal peptides—complex molecules (eg, surfactin) with antimicrobial or metal-chelating properties. The abundance of these clusters highlights the strain's potential for drug discovery or agricultural applications (eg, biocontrol agents).

Identification of Antimicrobial Resistance Genes in the Genome

The genome of *Bacillus subtilis* LIN78 was screened for antimicrobial resistance (AMR) genes using the Comprehensive Antibiotic Resistance Database (CARD).[25] Protein sequences from the *B. subtilis* strain and plasmid were analyzed via BLAST against CARD, applying an E-value cutoff of ≤1e-5. Genomic analysis of *B. subtilis* LIN78 isolates uncovered 19 predicted AMR genes (Table 4 and Table S9), while none were detected in the plasmid LIN78.

A detailed genomic investigation of *Bacillus subtilis* LIN78 identified 19 antibiotic resistance genes, underscoring its capacity for multidrug resistance. These genes represent various defensive strategies, such as efflux pumps, target alteration, enzymatic inactivation, and cell wall modification. Key resistance mechanisms include: A) Macrolide Resistance: Mediated by the *mphK* gene (encoding macrolide phosphotransferase) and ribosomal protection genes *lmrB* and *tmrB*. B) Glycopeptide Resistance: Linked to the *vanG* cluster (*vanT, vanW*, *vanY*), which facilitates peptidoglycan precursor modification, and the *vanW* gene within the *vanI* cluster. C) Multidrug Efflux Systems: Efflux pumps (*qacJ, qacG*, *bmr, blt*) conferring resistance to cationic antimicrobials and biocides. D) Membrane Adaptation: The *mprF* gene, encoding lysyl-phosphatidylglycerol synthase, reduces cationic antimicrobial binding. E) Enzyme-Based Inactivation: Includes *FosBx1* (fosfomycin resistance), *BcI* (β-lactamase inhibition), and *aadK* (aminoglycoside adenyltransferase). F) Tetracycline Resistance: Driven by the ribosomal protection gene *tet(45)*. G) Supplementary Factors: ABC transporters (*ykkC/ykkD*) and the putative methyltransferase *vmlR*.

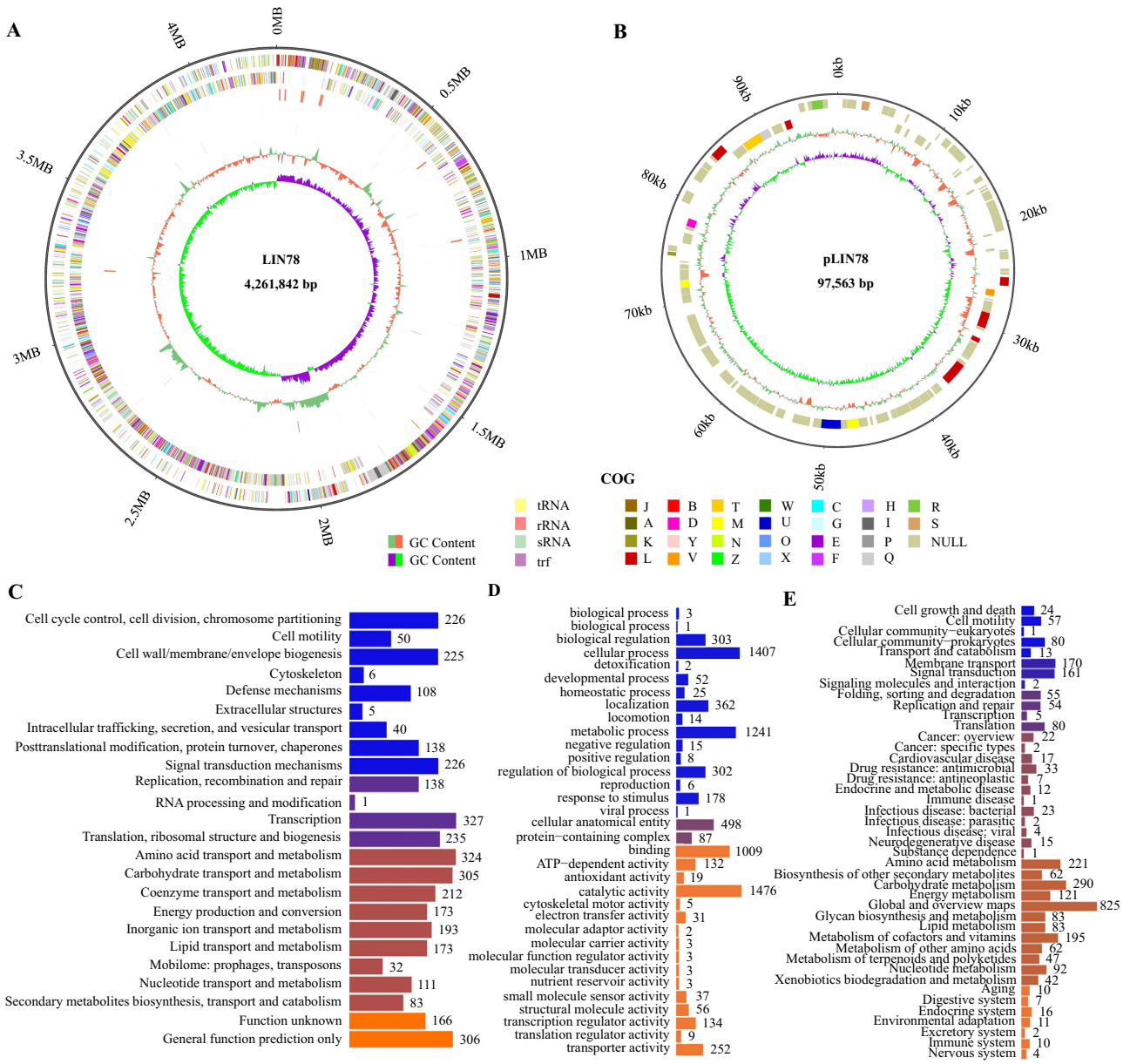

Figure 2 The complete genome sequences of Bacillus subtilis LIN78 chromosome (**A**) and plasmid (**B**) are depicted with the following labeling from outside to inside: contigs, COGs on the forward strand; CDS, tRNAs, and rRNAs on the forward strand; CDS, tRNAs, and rRNAs on the reverse strand; COGs on the reverse strand; GC content; and GC skew. The Cluster of Orthologous Groups (COGs) analysis (**C**) of LIN78 genes categorizes them into four groups: metabolism, cellular processes, information, and poorly characterized. The gene ontology (GO) analysis (**D**) groups the genes into three categories: biological processes, cellular components, and molecular functions. The Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis (**E**) classifies these genes into six categories: cellular processes, environmental information processing, genetic information processing, human diseases, metabolism, and organismal systems.

This genomic inventory illustrates the advanced resistance mechanisms of *B. subtilis*, spanning β-lactams, glycopeptides, macrolides, tetracyclines, aminoglycosides, fosfomycin, and disinfectants. The findings emphasize the urgency of monitoring AMR gene dissemination in clinical and environmental ecosystems to mitigate resistance spread.

Phylogenetic Tree Analysis

Phylogenetic analysis using a whole-genome maximum likelihood (ML) approach was conducted to clarify the evolutionary position of *Bacillus subtilis* strain LIN78 relative to other *Bacillus* lineages (Figure 5). The constructed tree, incorporating 19 closely related strains, revealed that LIN78 and *B. anthracis* share a common ancestor. Furthermore,

https://doi.org/10.2147/IDR.S544153

Table 2 Comparative Analysis of LIN78 Genome Annotation Across Diverse Databases

Database	Annotated Number (%)
VFDB	209 (4.6%)
ARDB	30 (0.66%)
CAZY	253 (5.57%)
IPR	3,766 (82.96%)
Swiss-Prot	3,936 (86.71%)
COG	3,302 (72.74%)
CARD	13 (0.28%)
GO	2,578 (58.79%)
KEGG	2,734 (60.23%)
NR	4,510 (99.36%)
T3SS	475 (10.46%)
Over All	4539 (99.42%)

LIN78 displayed close genetic ties to three food-derived *B. subtilis* strains: BS16045 (isolated from Korean Gochujang),[27] DKU_NT_02 (from traditional soybean-based Korean foods),[28] and a third strain sourced from rice (*Oryza sativa*) in India. Notably, all four strains clustered within a food-associated group distinct from the clade containing *Bacillus cereus* and *B. subtilis* subsp. *subtilis*. This separation highlights the ecological relevance of LIN78

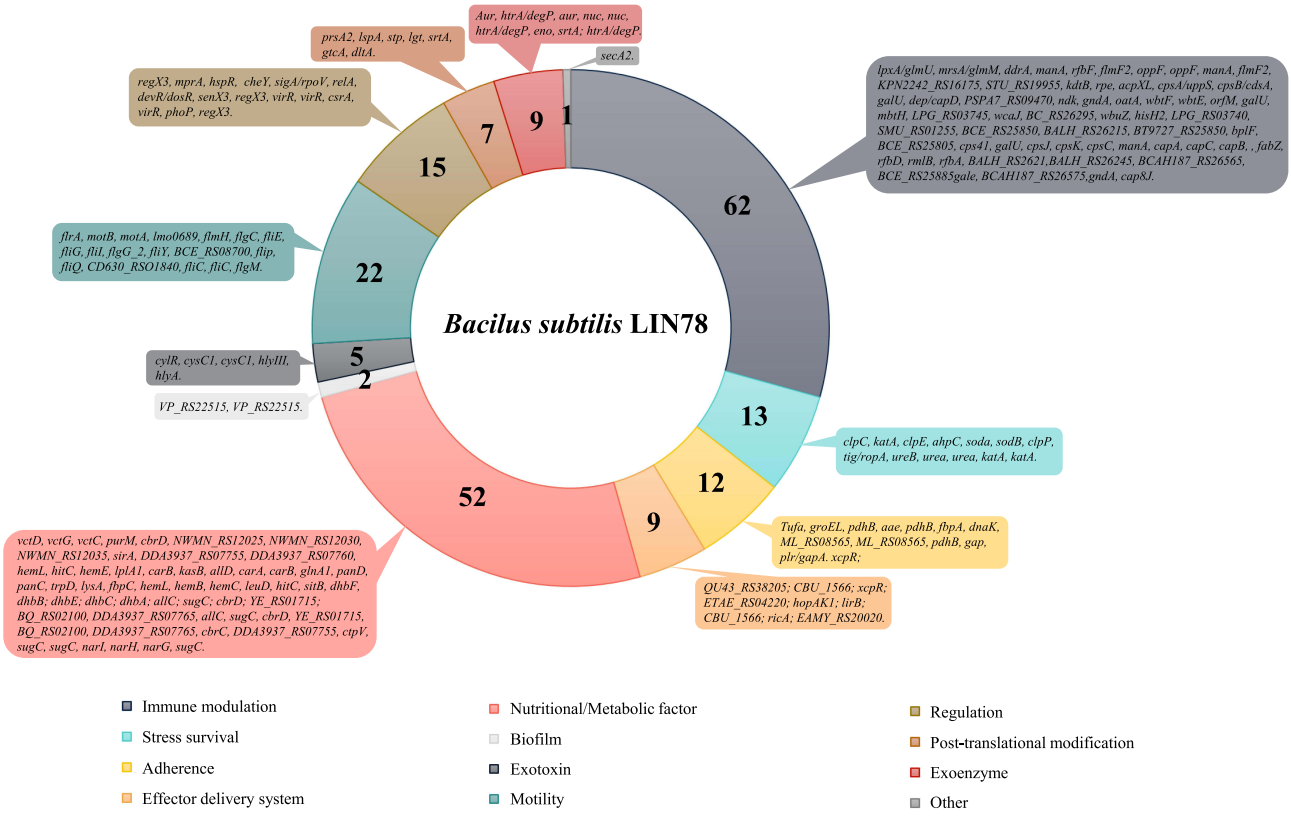

Figure 3 Virulence factor analysis of LIN78 genome. The circular representation displays the 209 identified virulence-associated genes. Genes are color-coded into 12 distinct functional categories. Each category is labeled directly on the figure, along with the corresponding number of genes and the specific gene names belonging to that category.

Table 3 Genes Associated with Homologous to Virulence Factors Found in B. subtilis str. LIN78

Gene_id	Identity	Source	Description
BSLIN78GL003901	100	*B. subtilis* subsp. subtilis str. 168	(*capA*) required for Poly-gamma-glutamate transport
BSLIN78GL003902	100	*B. subtilis* subsp. subtilis str. 168	(*capC*) Poly-gamma-glutamate synthesis
BSLIN78GL003903	100	*B. subtilis* subsp. subtilis str. 168	(*capB*) Poly-gamma-glutamate synthesis
BSLIN78GL003480	99.7	*B. subtilis* subsp. subtilis str. 168	(*dhbB*) isochorismatase
BSLIN78GL003483	99.6	*B. subtilis* subsp. subtilis str. 168	(*dhbA*) 2,3-dihydroxybenzoate-2,3-dehydrogenase
BSLIN78GL002188	99.5	*B. subtilis* subsp. subtilis str. 168	(*dep*/*capD*) gamma-glutamyltranspeptidase
BSLIN78GL002367	99.5	*B. subtilis* subsp. subtilis str. 168	(*hlyIII*) putative membrane hydrolase
BSLIN78GL003481	99.4	*B. subtilis* subsp. subtilis str. 168	(*dhbE*) 2,3-dihydroxybenzoate adenylase
BSLIN78GL003479	99.2	*B. subtilis* subsp. subtilis str. 168	(*dhbF*) non-ribosomal peptide synthetase
BSLIN78GL003482	99.2	*B. subtilis* subsp. subtilis str. 168	(*dhbC*) isochorismate synthase
BSLIN78GL004220	80.4	*B. cereus* AH187	(*galE*) UDP-glucose 4-epimerase
BSLIN78GL003875	79.7	*B. thuringiensis* serovar konkukian str. 97-27	(*galU*) UTP–glucose-1-phosphate uridylyltransferase
BSLIN78GL000094	79	*L. monocytogenes* EGD-e	(*clpC*) endopeptidase Clp ATP-binding chain C
BSLIN78GL003750	76.1	*L. monocytogenes* EGD-e	(*clpP*) ATP-dependent Clp protease proteolytic subunit
BSLIN78GL000120	75	*F. noatunensis* subsp. orientalis str. Toba 04	(*tufA*) elongation factor Tu
BSLIN78GL000624	74.5	*C. thermocellum* ATCC 27405	(*groEL*) chaperonin
BSLIN78GL003682	71.9	*S. gordonii* str. Challis substr. CH1	(*eno*) phosphopyruvate hydratase
BSLIN78GL002584	70.4	*K. pneumoniae* subsp. pneumoniae NTUH-K2044	(*gndA*) NADP-dependent phosphogluconate dehydrogenase
BSLIN78GL001733	68.6	*B. cereus* ATCC 10987	(*BCE_RS08700*) response regulator
BSLIN78GL003721	67.7	*V. vulnificus* YJ016	(*wcaJ*) sugar transferase
BSLIN78GL001453	66.4	*L. monocytogenes* EGD-e	(*clpE*) ATP-dependent protease
BSLIN78GL001066	65.3	*L. monocytogenes* EGD-e	(*lplA1*) lipoate protein ligase
BSLIN78GL003583	64.1	*P. syringae* pv. syringae B728a	(*cbrD*) ABC transporter ATP-binding protein
BSLIN78GL003870	63.2	*S. pneumoniae* TIGR4	(*cps4I*) capsular polysaccharide biosynthesis protein
BSLIN78GL001691	63	*B. melitensis* bv. 1 str. 16M	(*acpXL*) acyl carrier protein
BSLIN78GL003939	62.1	*B. thuringiensis* str. Al Hakam	(*BALH_RS26215*) UDP-glucose/GDP-mannose dehydrogenase
BSLIN78GL003982	61.5	*H. pylori* Shi470	(*ureB*) urease beta subunit UreB, urea amidohydrolase
BSLIN78GL001753	60.2	*E. faecium* Aus0004	(*cpsA*/*uppS*) undecaprenyl diphosphate synthase
BSLIN78GL002465	60.2	*M. ulcerans* Agy99	(*ndk*) nucleoside-diphosphate kinase
BSLIN78GL002753	60.2	*C. trachomatis* D/UW-3/CX	(*dnaK*) chaperone protein DnaK
BSLIN78GL004051	59.7	*M.* sp. JDM601	(*narH*) nitrate reductase subunit beta
BSLIN78GL003719	59.5	*L. pneumophila* subsp. pneumophila str. Philadelphia 1	(*LPG_RS03745*) aminotransferase class I/II-fold pyridoxal phosphate-dependent enzyme
BSLIN78GL002959	59.3	*F. novicida* U112	(*wbtF*) NAD dependent epimerase
BSLIN78GL003798	59	*L. ivanovii* subsp. ivanovii PAM 55	(*lgt*) prolipoprotein diacylglyceryl transferase
BSLIN78GL003876	58.8	*E. faecalis* V583	(*cpsJ*) ABC transporter, ATP-binding protein

in food-related contexts. Interestingly, LIN78 did not group with the industrially significant reference strain ATCC 11774, indicating genomic divergence between these two lineages.

A striking observation emerged in the phylogenetic placement of extremophile *Bacillus* strains. *B. subtilis* subsp. *subtilis* G7 (deep-sea hydrothermal vent isolate),[29] *B. cereus* MB1 (Challenger Deep, Mariana Trench) [28], and *B. subtilis* RO-NN-1 (desert origin) were grouped into a single clade despite their disparate geographic origins. This unexpected clustering suggests high genomic conservation among these environmentally resilient strains, potentially reflecting convergent evolutionary adaptations. Their genetic cohesion contrasts with the distinct branching of LIN78, implying functional differences between food-associated and extremophile *Bacillus* populations. These findings underscore how comparative genomic and proteomic studies can elucidate the evolutionary pathways and ecological specialization of *Bacillus* species.

Comparative Genome Analysis Between LIN78 and Its Close Homolog

Strain LIN78 exhibits similarities to AMR1, DKU_NT_02, BS16045, A10, and ATCC 11774 in genomic features such as gene count, average gene length, G+C content, and non-coding RNA quantity. However, LIN78 features a comparatively

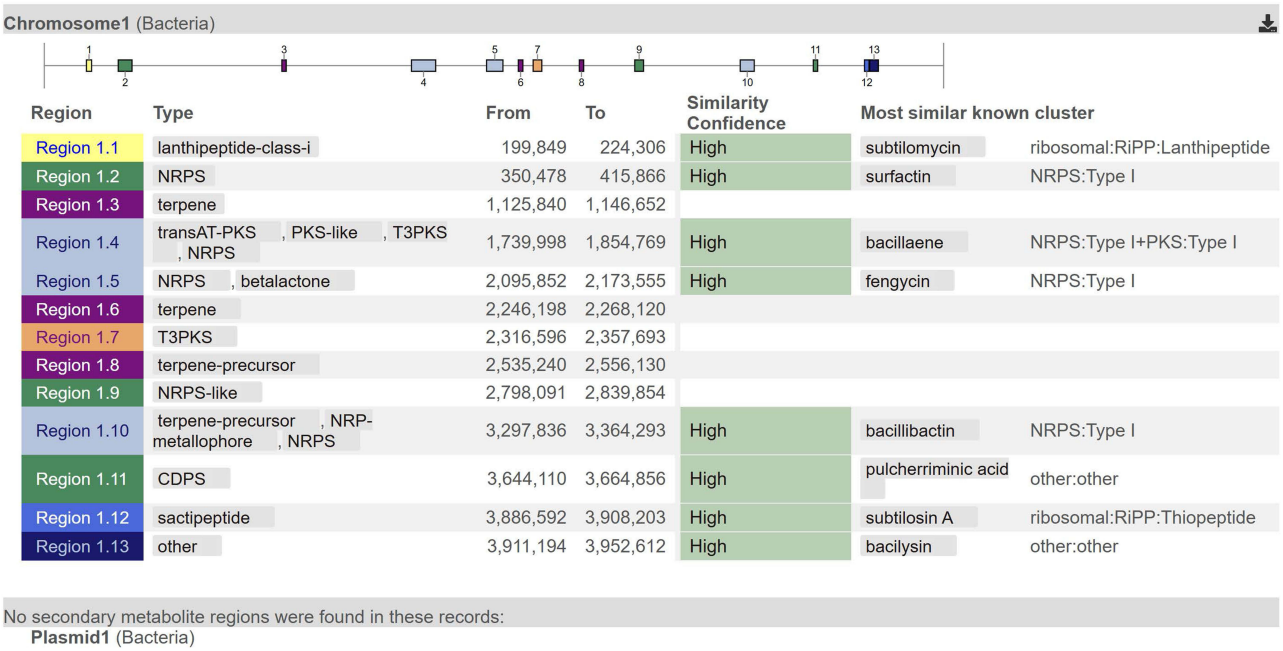

Figure 4 Prediction of secondary metabolite biosynthetic gene clusters in the *B. subtilis* LIN78 genome. This figure details the predicted gene clusters responsible for secondary metabolite production. Each row represents a single predicted cluster region, listing its designated Region identifier, the Type of metabolite predicted, the precise genome location (start-end) nucleotide coordinates, the similarity confidence score indicating the strength of the match to known clusters, and the most similar known cluster identified in public databases. Data source: anti-SMASH (version 7).

larger genome and harbors one plasmid (Figure 2 and Table S10). Whole-genome average nucleotide identity (ANI) analyses indicate that LIN78 shares >95% ANI with these strains, surpassing the established species boundary threshold (Figure 6A), a widely recognized criterion for species classification [28]. Heatmap analysis of dispensable genes further

Table 4 Antimicrobial Resistance Genes Found in B. cereus str. LIN78

ORF_ID	Best_Hit_ARO	Best_Identities	Drug Class
Chromosome1_252	*mphK*	99.02	Macrolide antibiotic
Chromosome1_267	*lmrB*	99.37	Lincosamide antibiotic
Chromosome1_314	*tmrB*	98.48	Nucleoside antibiotic
Chromosome1_470	*vant* gene in vanG cluster	31.78	Glycopeptide antibiotic
Chromosome1_555	*vmlR*	98.54	Lincosamide antibiotic; streptogramin antibiotic; streptogramin B antibiotic
Chromosome1_837	*Bacillus subtilis mprF*	99.88	Peptide antibiotic
Chromosome1_1318	*ykkC*	100	Aminoglycoside antibiotic; tetracycline antibiotic; phenicol antibiotic
Chromosome1_1319	*ykkD*	99.05	Aminoglycoside antibiotic; tetracycline antibiotic; phenicol antibiotic
Chromosome1_1748	*qacJ*	43.75	Disinfecting agents and antiseptics
Chromosome1_1749	*qacG*	47.47	Disinfecting agents and antiseptics
Chromosome1_2023	*Vant* gene in *vanG* cluster	34.59	Glycopeptide antibiotic
Chromosome1_2047	FosBx1	62.04	Phosphonic acid antibiotic
Chromosome1_2143	*Vanw* gene in vanI cluster	40	Glycopeptide antibiotic
Chromosome1_2151	*BcI*	66.34	cephalosporin; penem
Chromosome1_2249	*Vany* gene in vanG cluster	36.44	Glycopeptide antibiotic
Chromosome1_2511	*Bmr*	99.23	Fluoroquinolone antibiotic; nucleoside antibiotic; phenicol antibiotic; disinfecting agents and antiseptics
Chromosome1_2782	*Blt*	99.5	Fluoroquinolone antibiotic; disinfecting agents and antiseptics
Chromosome1_2801	*Aadk*	98.24	Aminoglycoside antibiotic
Chromosome1_4241	*Tet(45)*	75.6	Tetracycline antibiotic

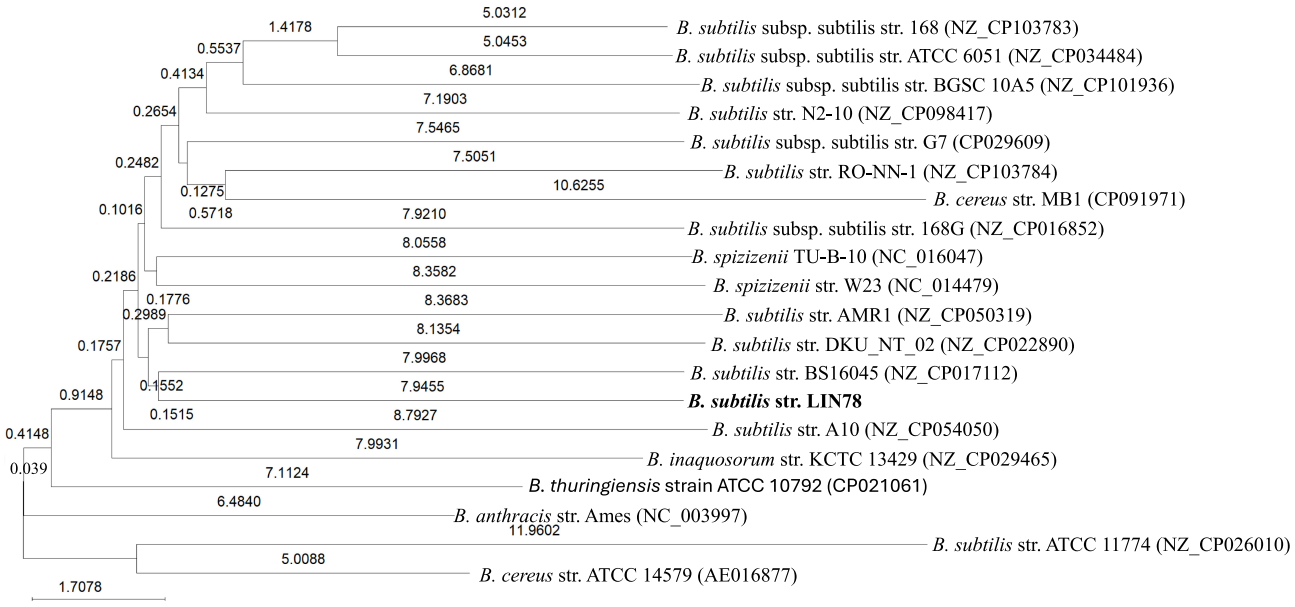

Figure 5 Phylogenetic tree of B. subtilis str. LIN78 and other *B. subtilis* strains. A Neighbor-joining tree was built based on 1,113 highly conserved orthologous genes from 6 annotated genomes of *Bacillus subtilis*. Bootstrap values (shown at nodes) represent the percentage of support from 1,000 resampled datasets. *B. subtilis* str. LIN78 is marked in bold. The scale bar represents 1.7078 nucleotide substitutions per site.

supports a closer evolutionary relationship between LIN78 and AMR1 compared to the other strains (Figure 6B). Comparative genomic analysis reveals 3176 conserved core genes shared among all strains, while LIN78 contains 425 unique dispensable genes, implying functional divergence (Figure 6C). Among these 425 genes, 53 strain-specific genes were assigned to 17 distinct COG categories (Table S11). These include roles in defense mechanisms (6 genes), replication/recombination/repair (5 genes), signal transduction (4 genes), transcription (4 genes), mobilome elements (5), and smaller contributions to cell division, motility, metabolism, and transport (1–3 genes per category), alongside 8 genes with general functional predictions and 2 uncharacterized loci (Figure 6D). Orthology assessment identified 2379 shared gene families across the five *B. subtilis* strains, with LIN78 uniquely possessing 9 gene families (Figure 6E). Synteny analysis demonstrates strong collinearity among LIN78, AMR1 (Figure 7A), A10 (Figure 7B), ATCC11774 (Figure 7C), BS16045 (Figure 7D), DKU_NT_02 (Figure 7E), whereas genomic rearrangements, including translocations and inversions, distinguish LIN78 and ATCC 11774 (Figure 7C).

Discussion

In this study, we characterized the biological traits of *Bacillus subtilis* strain LIN78, isolated from retail milk powder in China. Phylogenetic classification based on average nucleotide identity (ANI) confirmed LIN78 as a member of *B. subtilis*, showing high similarity to type strains AMR1 (98.82%), DKU_NT_02 (98.39%), BS16045 (98.61%), A10 (98.58%), and ATCC 11774 (98.67%). Notably, genome-based phylogenetic analysis revealed that LIN78 formed a distinct clade separate from pathogenic *Bacillus* species, including *B. anthracis* str. Ames and *B. cereus* ATCC 14579, suggesting unique genetic features differentiating it from pathogenic relatives. LIN78 clustered closely with other food-derived *B. subtilis* isolates, highlighting its potential role in food safety, particularly in spoilage prevention, toxin production, and preservation processes.

Comparative genomic analysis between LIN78 and *B. subtilis* ATCC 11774 identified significant structural variations in LIN78, including translocations, inversions, and insertions. These rearrangements, coupled with an abundance of mobile genetic elements (eg, integrases, transposases, and phage-related proteins), suggest that LIN78's genome has undergone extensive genetic remodeling via horizontal gene transfer, phage integration, and recombination. This likely explains the higher number of genomic islands (GIs) in LIN78 compared to ATCC 11774. LIN78 also possesses a substantially larger genome than AMR1, DKU_NT_02, BS16045, A10, and ATCC 11774, encoding 341–475

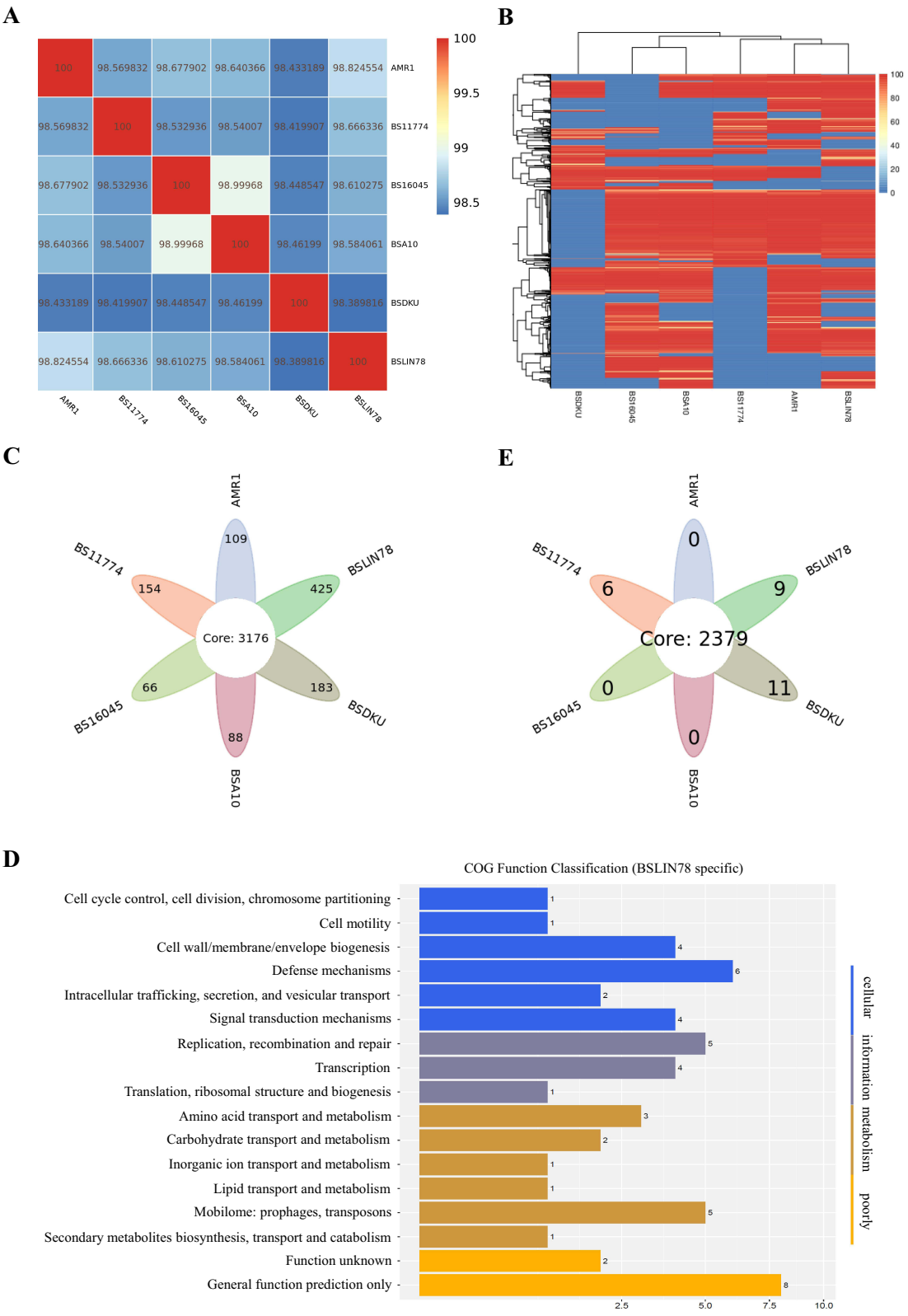

Figure 6 Comparative genomics of LIN78 and other *B. subtilis* strains. (**A**) ANI heatmap between strains. (**B**) Dispensable gene heat map. (**C**) Venn graph of pan genes. (**D**) Venn graph of orthologs in different gene family. (**E**) COG function classification of LIN78 specific genes. BSLIN78: LIN78; BSAMR1: AMR1; BSA10: A10; BS11774: ATCC11774; BS16045: BS16045; BSDKU: DKU_NT_02.

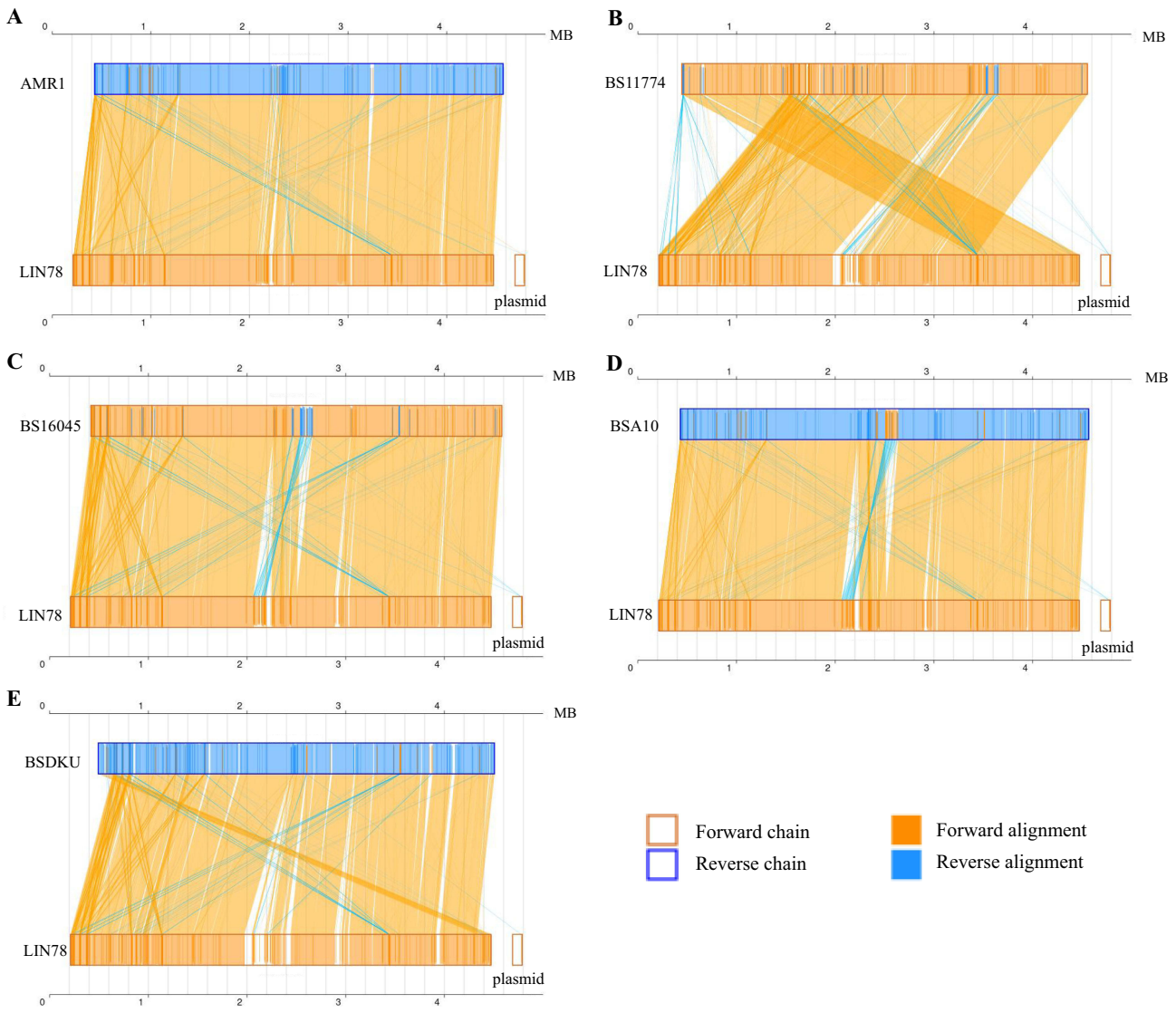

Figure 7 Genome alignments for LIN78 and its close homolog using MUMmer. Upper and lower axes of linear synteny graph are constructed after the same proportion of size reduction in length of both sequences. According to BLAST results, each pair nucleic acid sequence of the two alignments is marked in the coordinate diagram according to its position information, and the height of the filled color in the block indicates similarity of sequence alignment. The color of the lines between the two axes indicate the type of comparison. Genome alignments for (**A**) LIN78 and AMR1 (NZ_CP050319); (**B**) LIN78 and BS11774 (ZN_CP026010); (**C**) LIN78 and BS16045 (NZ_CP017112); (**D**) LIN78 and BSA10 (NZ_CP054050); (**E**) LIN78 and BSDKU (NZ_CP022890).

additional genes. While LIN78 shares 3,176 core genes with these strains, it harbors 425 strain-specific dispensable genes, indicating divergent functional capabilities.

Functional annotation of LIN78's 425 dispensable genes revealed enrichment in COG categories such as defense mechanisms (V), mobilome elements (prophages, transposons; X), replication/recombination/repair (L), and signal transduction (T). Notably, genes linked to defense mechanisms—often associated with antimicrobial resistance (AMR)—were overrepresented in its accessory genome. The abundance of mobilome-related genes aligns with the observed mobile elements, while enrichment in transcription (K) and signal transduction (T) genes may reflect adaptations to environmental stressors, such as nutrient acquisition and survival during food sterilization. Additionally, LIN78's genomic plasticity, underscored by replication/repair systems (L), likely enhances resilience in challenging environments.

Despite sharing high ANI values with other food-derived *B. subtilis* strains, LIN78 exhibits unique genomic features, including a remarkably large plasmid (pLIN78), which contributes to its expanded genome size. These findings

https://doi.org/10.2147/IDR.S544153

collectively suggest that LIN78 has evolved specialized genetic traits to thrive in food-related niches, distinguishing it from both pathogenic and non-pathogenic *Bacillus* strains.

While primarily utilized as a model system for studying cellular differentiation and for biotechnological applications,[2,30] *Bacillus subtilis* has historically been considered of limited clinical significance. Its pathogenic potential is typically restricted to opportunistic infections in immunocompromised hosts.[31,32] Studies indicate that pathogenic *Bacillus* species employ various virulence strategies. These include toxin secretion (eg, hemolysins, proteases, phospholipases, cytotoxins),[33–35] motility through swimming and swarming,[36,37] along with mechanisms facilitating intracellular survival and evasion of host immunity.[38,39] Genomic analysis of the LIN78 strain identified 209 genes potentially associated with virulence. These genes relate to functions such as immune modulation (eg, *capA, capC, capB*), stress adaptation (eg, *clpC, katA*), host adhesion (eg, *tufa, groEL, pdhB*), toxin synthesis (eg, *hlyIII, hlyA*), biofilm development (eg, *VP_RS22515, VP_RS22515*), motility (eg, *flrA, motA, flmH*), and enzymatic activity (eg, *aur, htrA*/*degP, srtA*). Significantly, LIN78 possesses multiple genes required for flagellar assembly, aligning with its observed polar and lateral flagella and demonstrated swimming and swarming motility. Such flagellum-driven movement is essential for bacterial chemotaxis towards favorable environments and is critical for host-pathogen dynamics.[40–44] LIN78 also carries genes encoding hemolysin III and cytolysin, which are implicated in tissue damage through plasma membrane disruption.[45]

The capacity of pathogens to adhere to host cells is a critical prerequisite for successful infection.[46,47] In this study, LIN78 was identified to encode multiple putative adhesion factors, including *tufA, groEL*, *pdhB, aae*,*fbpA, dnaK*, *gap, plr/gapA*—all previously associated with host-cell attachment mechanisms. Genomic analysis further revealed 61 predicted immune modulation genes in LIN78, linked to processes such as cell wall synthesis, capsule formation, and evasion of host defenses. Notably, LIN78 harbors genes involved in hyaluronic acid capsule production, a feature shared with pathogens like *Streptococcus pyogenes* and *Bacillus anthracis*, which exploit capsules to resist phagocytosis[48,49] and evade complement-mediated killing.[50] Additionally, LIN78 carries intracellular survival-associated genes such as *lplA1*,[51] *lspA*,[52] *panD*/*panC*,[53] *sodA*,[54] and *sodB*,[55] suggesting strategies for persistence within host cells.

The strain also encodes 14 iron acquisition systems, including ABC transporters for iron(III), iron chelates, and free iron, underscoring its ability to compete for this essential nutrient. Iron plays a central role in host-pathogen dynamics, influencing both microbial virulence and host immune responses.[56] The combination of robust iron-scavenging mechanisms and intracellular survival genes likely enhances LIN78's capacity to replicate within phagocytic cells, highlighting its adaptability in hostile host environments.

The identification of *Bacillus subtilis* strain LIN78 in retail milk powder and its capacity to degrade dairy components highlights potential challenges in food quality control and microbial safety. Whole-genome sequencing revealed a notable repertoire of genes linked to proteolytic and stress adaptation functions, including *aur* (encoding metalloproteases), *htrA/degP* (a chaperone-protease critical for protein quality control under stress), *nuc* (nucleases), *eno* (enolase), and *srtA* (sortase involved in cell wall anchoring). The presence of multiple copies of *aur* and *htrA/degP* suggests enhanced proteolytic activity and stress tolerance, likely enabling LIN78 to thrive in desiccated milk powder environments by degrading casein and other milk proteins for nutrient acquisition. These enzymatic capabilities pose direct risks to milk powder quality: (1) Extensive casein hydrolysis by Aur metalloproteases may cause bitterness and textural defects (eg, gelation or sedimentation upon reconstitution); (2) HtrA/DegP-mediated tolerance to heat, desiccation, and oxidative stress could allow survival during pasteurization or spray-drying, leading to post-process enzymatic activity that accelerates spoilage during storage—potentially shortening shelf-life. This aligns with reports of *Bacillus* spp. utilizing proteases to hydrolyze dairy matrices, inducing spoilage through pH shifts, off-flavor development, or functional property loss. The coexistence of *nuc* and *srtA* may further indicate adaptive strategies for biofilm formation or extracellular DNA utilization, which could facilitate persistent contamination in dairy processing equipment, perpetuating quality issues.

LIN78's genomic toolkit for stress survival (eg, *htrA/degP*, oxidative stress regulators) and nutrient scavenging (eg, proteases, nucleases) directly threatens milk powder stability. The strain's resilience during thermal processing (pasteurization/spray-drying) may permit residual enzymatic activity in finished products. Proteolysis of milk proteins during storage not only depletes nutritional value but also generates bitter peptides and aggregates that compromise solubility and sensory acceptance. Lipid hydrolysis by lipases (implied by stress-linked regulators) could drive rancidity. These factors collectively contribute to shelf-life reduction—a key concern for manufacturers.

While *B. subtilis* is generally regarded as non-pathogenic, the strain's genetic profile underscores the need to reassess its role in food spoilage and possible risk of opportunistic cytotoxicity under specific conditions. These findings emphasize the importance of monitoring stress-responsive and hydrolytic genes in foodborne *Bacillus* strains. Future studies should correlate gene expression with enzymatic activity in situ and evaluate the strain's impact on sensory and nutritional properties of dairy products. Implementing genomic surveillance for such strains could refine predictive models for shelf-life estimation and contamination prevention in dairy supply chains.

The genomic analysis of *Bacillus subtilis* Lin78 using anti-SMASH 7.0 revealed 13 predicted secondary metabolite biosynthetic gene clusters (BGCs), suggesting a considerable potential for the production of diverse bioactive compounds. Notably, non-ribosomal peptide synthetase (NRPS)-related clusters dominated the BGC profile, consistent with the well-documented role of NRPS pathways in *Bacillus* species for synthesizing antimicrobial peptides, siderophores (crucial for iron acquisition as discussed earlier), and surfactants. This genomic evidence implies that strain Lin78 may prioritize resource allocation towards pathways for synthesizing NRPS-derived metabolites, which are often critical for ecological competition and stress adaptation in bacterial communities, potentially including the milk powder environment. The detection of additional clusters, including those for lanthipeptides, terpenes, trans-AT polyketide synthases (trans-AT-PKS), sactipeptides, and a cyclodipeptide synthase (CDPS), further underscores the predicted metabolic versatility of this strain. For instance, lanthipeptides and sactipeptides are ribosomally synthesized and post-translationally modified peptides (RiPPs) known to possess antimicrobial properties in other contexts, while terpenes and T3PKS products may contribute to niche specialization or signaling. The CDPS cluster could indicate potential for generating cyclodipeptides with pharmaceutical relevance. While these predictions highlight LIN78's potential as a candidate for exploring novel bioactive molecules relevant to biotechnological applications (eg, antimicrobial discovery, agricultural biocontrol), the functional expression and specific ecological or applied roles of metabolites encoded by these 13 BGCs remain to be experimentally determined. Confirmation of cluster activity and metabolite production under relevant conditions (eg, milk powder storage) would require future investigation employing transcriptomic and metabolomic approaches. This validation is crucial to move beyond genomic prediction and assess the concrete contribution of these pathways to LIN78's physiology, potential competitive advantage in food products, and its safety or utility profile.

The genomic analysis of *Bacillus subtilis* LIN78 revealed a striking array of 19 antimicrobial resistance (AMR) genes, underscoring its potential as a reservoir for multidrug resistance. Notably, the absence of AMR genes in the plasmid LIN78 suggests that resistance determinants are chromosomally integrated, which may enhance their stability and persistence within this strain. The identified genes span diverse resistance mechanisms, including efflux pumps, enzymatic inactivation, and target modification, reflecting a sophisticated evolutionary adaptation to antimicrobial pressures. Of particular concern is the coexistence of macrolide resistance genes (*mphK, lmrB*, *tmrB*) and the *vanG* cluster—a genetic element previously associated with glycopeptide resistance in enterococci. This finding raises questions about horizontal gene transfer between *Bacillus* spp. and clinically relevant pathogens, as mobile genetic elements in soil-dwelling bacilli could act as conduits for resistance dissemination. The presence of multidrug efflux systems (*qacJ, bmr*), β-lactamase inhibitors (*Bcl*), and tetracycline resistance gene *tet* highlights the strain's capacity to resist both therapeutic antibiotics (eg, β-lactams, aminoglycosides) and biocides used in agriculture or healthcare. The *vanW* and *vanY* homologs within the *vanI* cluster further suggest a cryptic glycopeptide resistance pathway, warranting experimental validation of its phenotypic expression. While *B. subtilis* is generally considered non-pathogenic, the genomic plasticity evidenced here—coupled with its environmental ubiquity—emphasizes the need to monitor resistance gene flow between commensal soil microbes and clinically critical pathogens. Future studies should prioritize functional characterization of these AMR genes and investigate their transfer dynamics in polymicrobial environments to mitigate risks of resistance spread.

Conclusion

This study provides a comprehensive genomic and functional characterization of *Bacillus subtilis* strain LIN78, isolated from retail milk powder in China. Phylogenetic analysis confirmed LIN78's placement within the *B. subtilis* clade, distinct from pathogenic *Bacillus* species, while highlighting its close relationship to food-derived strains. LIN78's genome exhibits remarkable plasticity, marked by structural rearrangements, mobile genetic elements, and a significantly larger genome size compared to other *B. subtilis* strains, driven in part by the presence of a large plasmid (pLIN78).

These features, coupled with an abundance of strain-specific dispensable genes, suggest extensive horizontal gene transfer and adaptation to food-related niches, particularly in stress tolerance, nutrient acquisition, and spoilage-related proteolytic activity. Functional annotation revealed genomic traits aligned with environmental resilience, including robust iron-scavenging systems, adhesion factors, and immune modulation mechanisms, which may enhance survival in host environments. While LIN78 encodes virulence-associated genes and shares capsule biosynthesis pathways with pathogens, its primary ecological role appears geared toward food matrix degradation, posing a significant potential risk to dairy quality control through spoilage. The strain's proteolytic (eg, *aur*, *htrA/degP*) and thermal tolerance genes underscore its capacity to persist through food-processing conditions, necessitating stricter monitoring in industrial settings. Critically, the identification of 19 chromosomally encoded antimicrobial resistance (AMR) genes, including efflux pumps (eg, *qacJ*, *bmr*), glycopeptide resistance markers (*vanG* cluster), and macrolide resistance genes (*mphK, lmrB*), constitutes a major potential risk factor for resistance dissemination in food systems. Additionally, the strain's diverse secondary metabolite biosynthetic clusters, particularly NRPS-derived compounds, highlight its biotechnological potential for antimicrobial and biocontrol applications. In summary, LIN78 exemplifies the genomic adaptability of *B. subtilis* to food environments while carrying traits that blur the line between commensal and opportunistic pathogen. These findings emphasize the need for targeted genomic surveillance of foodborne *Bacillus* strains, prioritizing markers linked to AMR (eg, glycopeptide/macrolide resistance), virulence (eg, capsule biosynthesis, toxins, adhesion factors), stress-driven spoilage (eg, proteases, heat-shock regulators), and mobile genetic elements, to mitigate spoilage risks, rigorously assess safety implications, and curb the spread of resistance determinants. Future research should focus on correlating genomic traits with phenotypic expression in food systems and host interactions to refine risk assessments and leverage biotechnological opportunities.

Data Sharing Statement

The data supporting this study's findings are available from the corresponding author, Dr. Yibin Lin upon reasonable request.

Ethics Declaration

This work did not involve ethical issues. The study does not include patients' privacy. No ethical approval was required.

Acknowledgments

We thank Dr. Shengwen Shao for assistance with experimental design and help with manuscript preparation. We would like to thank the reviewers for their helpful remarks. We used Deepseek-R1 for language refinement during manuscript preparation.

Author Contributions

All authors made a significant contribution to the work reported, whether that is in the conception, study design, execution, acquisition of data, analysis, and interpretation, or all these areas; took part in drafting, revising, or critically reviewing the article; gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and agree to be accountable for all aspects of the work.

Funding

This research was supported by the National Natural Science Foundation of China under Grant No. 82472291, Zhejiang Provincial Natural Science Foundation of China under Grant No. LTGD24C010001, the Medical & Health Technology Program of Zhejiang Province under Grant No. 2024KY424, Huzhou Municipal Science and Technology Bureau (2024GZ74) to Y.L.; the National Undergraduate Training Program for Innovation (202410347013) to K.C; Zhejiang Science and Technology Plan for Disease Prevention and Control (Project No. 2025JK291) to L.Z.

Disclosure

The authors declare no competing interests in this work.

References

1. Karanth S, Feng S, Patra D, Pradhan AK. Linking microbial contamination to food spoilage and food waste: the role of smart packaging, spoilage risk assessments, and date labeling. *Front Microbiol*. 2023;14:1198124. doi:10.3389/fmicb.2023.1198124
2. Su Y, Liu C, Fang H, Zhang D. *Bacillus subtilis*: a universal cell factory for industry, agriculture, biomaterials and medicine. *Microb Cell Fact*. 2020;19(1):173. doi:10.1186/s12934-020-01436-8
3. Stulke J, Gruppen A, Bramkamp M, Pelzer S. *Bacillus subtilis*, a Swiss army knife in science and biotechnology. *J Bacteriol*. 2023;205(5): e0010223. doi:10.1128/jb.00102-23
4. Hariram U, Labbe R. Spore prevalence and toxigenicity of *Bacillus cereus* and *Bacillus thuringiensis* isolates from U.S. retail spices. *J Food Prot*. 2015;78(3):590–596. doi:10.4315/0362-028X.JFP-14-380
5. Setlow P, Eichenberger P, Driks A. Spore resistance properties. *Microbiol Spectr*. 2014;2(5). doi:10.1128/microbiolspec.TBS-0003-2012
6. Gopal N, Hill C, Ross PR, Beresford TP, Fenelon MA, Cotter PD. The prevalence and control of bacillus and related spore-forming bacteria in the dairy industry. *Front Microbiol*. 2015;6:1418. doi:10.3389/fmicb.2015.01418
7. Wang C, Du Q, Yao T, et al. spoilage bacteria identification and food safety risk assessment of whole soybean curd. *Indian J Microbiol*. 2019;59 (2):250–253. doi:10.1007/s12088-019-00778-1
8. Mathur S, Singh R. Antibiotic resistance in food lactic acid bacteria--a review. *Int J Food Microbiol*. 2005;105(3):281–295. doi:10.1016/j.ijfoodmicro.2005.03.008
9. Franz CM, Huch M, Abriouel H, Holzapfel W, Galvez A. Enterococci as probiotics and their implications in food safety. *Int J Food Microbiol*. 2011;151(2):125–140. doi:10.1016/j.ijfoodmicro.2011.08.014
10. Yuan L, Fan L, Liu S, et al. Bacterial community analysis of infant foods obtained from Chinese markets by combining culture-dependent and high-throughput sequence methods. *Food Res Int*. 2022;162(Pt B):112060. doi:10.1016/j.foodres.2022.112060
11. Liu Y, Zhang B, Wang S, Yang Z, Yuan L. Different risk potential of *Bacillus cereus* to the dairy industry − keeping up with the old story. *Food Sci Anim Prod*. 2024;2:9240085. doi:10.26599/FSAP.2024.9240085
12. Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res*. 1997;25 (5):955–964. doi:10.1093/nar/25.5.955
13. Lagesen K, Hallin P, Rodland EA, Staerfeldt HH, Rognes T, Ussery DW. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res*. 2007;35(9):3100–3108. doi:10.1093/nar/gkm160
14. Ontiveros-Palacios N, Cooke E, Nawrocki EP, et al. Rfam 15: RNA families database in 2025. *Nucleic Acids Res*. 2025;53(D1):D258–D267. doi:10.1093/nar/gkae1023
15. Grissa I, Vergnaud G, Pourcel C. CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Res*. 2007;35(Web Server):W52–W57. doi:10.1093/nar/gkm360
16. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*. 2000;28(1):27–30. doi:10.1093/nar/28.1.27
17. Galperin MY, Kristensen DM, Makarova KS, Wolf YI, Koonin EV. Microbial genome analysis: the COG approach. *Brief Bioinform*. 2019;20 (4):1063–1070. doi:10.1093/bib/bbx117
18. Boutet E, Lieberherr D, Tognolli M, Schneider M, Bairoch A. UniProtKB/Swiss-Prot. *Methods Mol Biol*. 2007;406:89–112. doi:10.1007/978-1-59745-535-0_4
19. The Gene Ontology Consortium. The gene ontology resource: 20 years and still going strong. *Nucleic Acids Res*. 2019;47(D1):D330–D338. doi:10.1093/nar/gky1055
20. O'Donovan C, Martin MJ, Gattiker A, Gasteiger E, Bairoch A, Apweiler R. High-quality protein knowledge resource: swiss-PROT and TrEMBL. *Brief Bioinform*. 2002;3(3):275–284. doi:10.1093/bib/3.3.275
21. Huerta-Cepas J, Szklarczyk D, Heller D, et al. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res*. 2019;47(D1):D309–D314. doi:10.1093/nar/gky1085
22. Liu B, Zheng D, Zhou S, Chen L, Yang J. VFDB 2022: a general classification scheme for bacterial virulence factors. *Nucleic Acids Res*. 2022;50 (D1):D912–D917. doi:10.1093/nar/gkab1107
23. Liu B, Pop M. ARDB–antibiotic resistance genes database. *Nucleic Acids Res*. 2009;37:D443–447. doi:10.1093/nar/gkn656
24. Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Res*. 2009;37:D233–238. doi:10.1093/nar/gkn663
25. Alcock BP, Huynh W, Chalil R, et al. CARD 2023: expanded curation, support for machine learning, and resistome prediction at the comprehensive antibiotic resistance database. *Nucleic Acids Res*. 2023;51(D1):D690–D699. doi:10.1093/nar/gkac920
26. Blin K, Shaw S, Augustijn HE, et al. antiSMASH 7.0: new and improved predictions for detection, regulation, chemical structures and visualisation. *Nucleic Acids Res*. 2023;51(W1):W46–W50. doi:10.1093/nar/gkad344
27. Jeon S, Heo J, Uhm T-B. Complete Genome Sequence of *Bacillus subtilis* BS16045 Isolated from Gochujang. *Korean J Microbiol*. 2017;53(1):3. doi:10.7845/kjm.2017.6076
28. Bang MS, Jeong HW, Lee YJ, et al. Complete genome sequence of *Bacillus subtilis* strain dku_nt_02, isolated from traditional korean food using soybean (chung-gook-jang) for high-quality poly-gamma-glutamic acid activity. *Genome Announc*. 2018;6(25). doi:10.1128/genomeA.00525-18.
29. Gu HJ, Sun QL, Luo JC, Zhang J, Sun L. A first study of the virulence potential of a *Bacillus subtilis* isolate from deep-sea hydrothermal vent. *Front Cell Infect Microbiol*. 2019;9:183. doi:10.3389/fcimb.2019.00183
30. Gu Y, Xu X, Wu Y, et al. Advances and prospects of *Bacillus subtilis* cellular factories: from rational design to industrial applications. *Metab Eng*. 2018;50:109–121. doi:10.1016/j.ymben.2018.05.006
31. Oggioni MR, Pozzi G, Valensin PE, Galieni P, Bigazzi C. Recurrent septicemia in an immunocompromised patient due to probiotic strains of *Bacillus subtilis*. *J Clin Microbiol*. 1998;36(1):325–326. doi:10.1128/JCM.36.1.325-326.1998
32. Logan NA. Bacillus and relatives in foodborne illness. *J Appl Microbiol*. 2012;112(3):417–429. doi:10.1111/j.1365-2672.2011.05204.x
33. Bottone EJ. Bacillus cereus, a volatile human pathogen. *Clin Microbiol Rev*. 2010;23(2):382–398. doi:10.1128/CMR.00073-09
34. Slamti L, Perchat S, Gominet M, et al. Distinct mutations in PlcR explain why some strains of the *Bacillus cereus* group are nonhemolytic. *J Bacteriol*. 2004;186(11):3531–3538. doi:10.1128/JB.186.11.3531-3538.2004
35. Collier RJ, Young JA. Anthrax toxin. *Annu Rev Cell Dev Biol*. 2003;19:45–70. doi:10.1146/annurev.cellbio.19.111301.140655

36. Senesi S, Ghelardi E, Celandroni F, Salvetti S, Parisio E, Galizzi A. Surface-associated flagellum formation and swarming differentiation in Bacillus subtilis are controlled by the ifm locus. *J Bacteriol*. 2004;186(4):1158–1164. doi:10.1128/JB.186.4.1158-1164.2004
37. Kearns DB, Losick R. Swarming motility in undomesticated *Bacillus subtilis*. *Mol Microbiol*. 2003;49(3):581–590. doi:10.1046/j.1365-2958.2003.03584.x
38. Dixon TC, Fadl AA, Koehler TM, Swanson JA, Hanna PC. Early *Bacillus anthracis*-macrophage interactions: intracellular survival survival and escape. *Cell Microbiol*. 2000;2(6):453–463. doi:10.1046/j.1462-5822.2000.00067.x
39. Tran SL, Ramarao N. *Bacillus cereus* immune escape: a journey within macrophages. *FEMS Microbiol Lett*. 2013;347(1):1–6. doi:10.1111/1574-6968.12209
40. Wadhams GH, Armitage JP. Making sense of it all: bacterial chemotaxis. *Nat Rev Mol Cell Biol*. 2004;5(12):1024–1037. doi:10.1038/nrm1524
41. Sourjik V, Wingreen NS. Responding to chemical gradients: bacterial chemotaxis. *Curr Opin Cell Biol*. 2012;24(2):262–268. doi:10.1016/j.ceb.2011.11.008
42. Renier S, Hebraud M, Desvaux M. Molecular biology of surface colonization by Listeria monocytogenes: an additional facet of an opportunistic Gram-positive foodborne pathogen. *Environ Microbiol*. 2011;13(4):835–850. doi:10.1111/j.1462-2920.2010.02378.x
43. Chaban B, Hughes HV, Beeby M. The flagellum in bacterial pathogens: for motility and a whole lot more. *Semin Cell Dev Biol*. 2015;46:91–103. doi:10.1016/j.semcdb.2015.10.032
44. Josenhans C, Suerbaum S. The role of motility as a virulence factor in bacteria. *Int J Med Microbiol*. 2002;291(8):605–614. doi:10.1078/1438-4221-00173
45. Baida GE, Kuzmin NP. Mechanism of action of hemolysin III from *Bacillus cereus*. *Biochim Biophys Acta*. 1996;1284(2):122–124. doi:10.1016/S0005-2736(96)00168-X
46. Finlay BB, Falkow S. Common themes in microbial pathogenicity revisited. *Microbiol Mol Biol Rev*. 1997;61(2):136–169. doi:10.1128/mmbr.61.2.136-169.1997
47. Ribet D, Cossart P. How bacterial pathogens colonize their hosts and invade deeper tissues. *Microbes Infect*. 2015;17(3):173–183. doi:10.1016/j.micinf.2015.01.004
48. Dale JB, Washburn RG, Marques MB, Wessels MR. Hyaluronate capsule and surface M protein in resistance to opsonization of group A streptococci. *Infect Immun*. 1996;64(5):1495–1501. doi:10.1128/iai.64.5.1495-1501.1996
49. Ezzell JW, Welkos SL. The capsule of bacillus anthracis, a review. *J Appl Microbiol*. 1999;87(2):250. doi:10.1046/j.1365-2672.1999.00881.x
50. O'Brien DK, Ribot WJ, Chabot DJ, et al. The capsule of *Bacillus anthracis* protects it from the bactericidal activity of human defensins and other cationic antimicrobial peptides. *PLoS Pathog*. 2022;18(9):e1010851. doi:10.1371/journal.ppat.1010851
51. Keeney KM, Stuckey JA, O'Riordan MX. LplA1-dependent utilization of host lipoyl peptides enables Listeria cytosolic growth and virulence. *Mol Microbiol*. 2007;66(3):758–770. doi:10.1111/j.1365-2958.2007.05956.x
52. Rahman MS, Ceraul SM, Dreher-Lesnick SM, Beier MS, Azad AF. The lspA gene, encoding the type II signal peptidase of Rickettsia typhi: transcriptional and functional analysis. *J Bacteriol*. 2007;189(2):336–341. doi:10.1128/JB.01397-06
53. Rojony R, Danelishvili L, Campeau A, Wozniak JM, Gonzalez DJ, Bermudez LE. Exposure of mycobacterium abscessus to environmental stress and clinically used antibiotics reveals common proteome response among pathogenic mycobacteria. *Microorganisms*. 2020;8(5):698. doi:10.3390/microorganisms8050698
54. Mittra B, Laranjeira-Silva MF, Miguel DC, Perrone Bezerra de Menezes J, Andrews NW. The iron-dependent mitochondrial superoxide dismutase SODA promotes Leishmania virulence. *J Biol Chem*. 2017;292(29):12324–12338. doi:10.1074/jbc.M116.772624
55. Courtney CM, Goodman SM, Nagy TA, et al. Potentiating antibiotics in drug-resistant clinical isolates via stimuli-activated superoxide generation. *Sci Adv*. 2017;3(10):e1701776. doi:10.1126/sciadv.1701776
56. Sheldon JR, Laakso HA, Heinrichs DE. Iron acquisition strategies of bacterial pathogens. *Microbiol Spectr*. 2016;4(2). doi:10.1128/microbiolspec.VMBF-0010-2015