2025, 65(2): 781-795
CSTR: 32112.14.j.AMS.20240578
DOI: 10.13343/j.cnki.wsxb.20240578

微生物学报
Acta Microbiologica Sinica
http://journals.im.ac.cn/actamicrocn

Research Article 研究报告

泛基因组分析鲁氏芽孢杆菌属的物种多样性与潜在工业应用

邹伟[1*]，杨凌凌[1]，刘超杰[1]，郑佳[2]，张楷正[1]，乔宗伟[2]

1 四川轻化工大学 生物工程学院，四川 宜宾
2 宜宾五粮液股份有限公司，四川 宜宾

邹伟, 杨凌凌, 刘超杰, 郑佳, 张楷正, 乔宗伟 . 泛基因组分析鲁氏芽孢杆菌属的物种多样性与潜在工业应用[J]. 微生物学报, 2025, 65(2): 781-795.
ZOU Wei, YANG Lingling, LIU Chaojie, ZHENG Jia, ZHANG Kaizheng, QIAO Zongwei. Pangenome analysis of *Rummeliibacillus* sp. strains reveals their unexpected diversity and potential for industrial application[J]. Acta Microbiologica Sinica, 2025, 65(2): 781-795.

摘　要：【目的】鲁氏芽孢杆菌属(*Rummeliibacillus*)包含 3 个种，即 *R. stabekisii*、*R. pycnus* 和 *R. suwonensis*，在生物降解、益生菌、动物饲料以及精氨酸、己酸等的生产方面具有一定的应用潜力。本研究旨在从基因组水平研究该属的遗传多样性。【方法】对该属不同来源的 12 个菌株进行比较泛基因组分析，分析内容还包括菌株的系统发育分析、功能注释、基因组代谢途径分析以及可移动遗传元件预测。【结果】泛基因组分析共鉴定出 8 024 个基因家族，核心基因组、附属基因组和菌株特异性基因分别包含 1 550、3 941 和 2 533 个基因家族。在核心基因组中，6 个菌株的精氨酸循环是完整的，7 个菌株具备完全生物合成乙偶姻的能力。然而，只有 *R. pycnus* 和 *R. suwonensis* 3B-1 具有合成己酸的能力。通过系统发育树、DNA-DNA 杂交(DNA-DNA hybridization，DDH)和平均核苷酸一致性(average nucleotide identity, ANI)分析发现，*Rummeliibacillus* sp. G93 和 *Rummeliibacillus* sp. TYF-LIM-RU47 均属于 *R. stabekisii*。*Rummeliibacillus* sp. POC4 和 *Rummeliibacillus* sp. TYF005 可能属于该属的一个新种。此外，在所有 12 个菌株中均鉴定出基因岛，其数量从 4 个(*R. stabekisii* DSM 25578 和 *R. stabekisii* NBRC 104870)到 14 个(*Rummeliibacillus* sp. SL167 菌株和 *Rummeliibacillus* sp. TYF005 菌株)不等，并且在 5 个分析菌株发现了前噬菌体序列。【结论】本研究提供了 *Rummeliibacillus* sp. 较为全面的基因家族谱，有助于对其进行进一步探索。

关键词：平均核苷酸一致性；细菌泛基因组分析；基因岛；泛基因组；鲁氏芽孢杆菌属

资助项目：中国轻工业浓香型白酒固态发酵重点实验室开放基金(2021JJ017)
This work was supported by the Key Laboratory of Wuliangye-flavor Liquor Solid-state Fermentation, China National Light Industry (2021JJ017).
*Corresponding author. E-mail: weizou@suse.edu.cn.
Received: 2024-09-19; Accepted: 2024-11-08; Published online: 2024-12-24

Pangenome analysis of *Rummeliibacillus* sp. strains reveals their unexpected diversity and potential for industrial application

ZOU Wei[1*], YANG Lingling[1], LIU Chaojie[1], ZHENG Jia[2], ZHANG Kaizheng[1], QIAO Zongwei[2]

1 College of Biological Engineering, Sichuan University of Science & Engineering, Yibin, Sichuan, China
2 Wuliangye Yibin Co., Ltd., Yibin, Sichuan, China

Abstract: [Objective] *Rummeliibacillus*, a genus encompassing three known species, *R. stabekisii*, *R. pycnus*, and *R. suwonensis*, has a wide range of potential applications in biodegradation, probiotics, animal feed, and production of arginine, caproic acid, and other compounds. This study aims to explore the genetic diversity of this genus at the genomic level. **[Methods]** A comparative pangenome analysis of 12 strains isolated from different sources was conducted. In addition, the phylogenetic analysis, functional annotation, genomic metabolic pathway analysis, and prediction of mobile genetic elements were carried out. **[Results]** A total of 8 024 gene clusters were identified. The core genome, accessory genome, and strain-specific genes comprised 1 550, 3 941, and 2 533 gene clusters, respectively. In the core genome, the arginine cycle of six strains was complete. Seven strains had the ability to completely biosynthesize acetoin. However, only *R. pycnus* and *R. suwonensis* 3B-1 were able to completely biosynthesize caproic acid. The phylogenetic tree, DNA-DNA hybridization, and average nucleotide identity showed that *Rummeliibacillus* sp. G93 and *Rummeliibacillus* sp. TYF-LIM-RU47 were strains of *R. stabekisii*. *Rummeliibacillus* sp. POC4 and *Rummeliibacillus* sp. TYF005 may belong to a new species of this genus. In addition, genomic islands were identified in all the 12 strains, with the number ranging from four (*R. stabekisii* DSM 25578 and *R. stabekisii* NBRC 104870) to 14 (*Rummeliibacillus* sp. SL167 and *Rummeliibacillus* sp. TYF005), and prophage sequences were found in five of the 12 strains. **[Conclusion]** This study provides a genomic framework for *Rummeliibacillus* that could assist the further exploration of this genus.

Keywords: average nucleotide identity; Bacterial pan-genome analysis (BPGA); genomic islands; pangenome; *Rummeliibacillus*

Rummeliibacillus was first described in the United States in 2009[1]. The genus includes three species, namely, *R. stabekisii*, *R. pycnus*[1] and *R. suwonensis*[2]. Although minimal research has been conducted on this genus, it has enormous potential for biotechnological applications[3]. *R. pycnus* can be used to produce arginine with higher catalytic efficiency than other arginases reported[4-5]. *R. pycnus* is able to convert palm oil plant wastewater into a terpolymer of polyhydroxyalkanoates and biodiesel[6]. *R. stabekisii* has the potential for biomineralization[7]. *R. suwonensis* can produce caproic acid using carbon sources such as sodium acetate[8]. Microbial communities consisting of *Rummeliibacillus* sp., *Caproiciproducens*, and *Clostridium_sensu_stricto_12* have been used to produce caproic acid from food waste by high-temperature fermentation[9]. *Rummeliibacillus* sp.

✉ actamicro@im.ac.cn, ☏ 010-64807516

can produce acetoin using a variety of carbon sources, including pentose, hexose, and lignocelluloses[10]. In addition, the genus can also be co-cultured with some *Clostridium* and *Bacillus* to produce hydrogen[11]. *Rummeliibacillus* sp. also has probiotic properties and can be used for the production of food biological preservatives[12] and animal feed additives[13]. Overall, *Rummeliibacillus* is an important genus with many potential industrial applications.

Now, the availability of next-generation sequencing technologies and decreasing sequencing costs have allowed genome-wide approaches for microbial analysis[14]. Genome sequencing and annotation and comparative genomics have accelerated the study of industrial microorganisms. Genomic information and bioinformatics software and databases can be used to compare multiple genomes of different species or genera[15]. The pangenome represents the entire genome of a species or genus or given phylogenetic branch, describes genetic variation, and allows the definition of core genomes, which could help to understand species diversity and metabolic capacity[16-17].

In this study, we systematically studied the specific taxonomic status of *Rummeliibacillus* sp. through average nucleotide identity (ANI), DNA-DNA hybridization (DDH), and phylogenetic tree analyses. Meanwhile, we analyzed the characteristics of the pangenome, core genes, and accessory genes of *Rummeliibacillus* sp. and evaluated the phylogenetic relationship of *Rummeliibacillus* sp. at the genome level. To study the potential functional characteristics of *Rummeliibacillus* sp., the metabolic capacity of the core genome and the accessory genome were analyzed. We evaluated genomic plasticity and genome evolution by the analysis of mobile genetic elements (MGEs)[18].

1 Materials and Methods

1.1 Pangenome analysis

The 12 genomes of *Rummeliibacillus* sp. used for the pangenome analysis were downloaded from the FTP site of the Reference Sequence (RefSeq) database at NCBI (ftp://ftp.ncbi.nih.gov/genomes/, accessed on March 4, 2023). Once the download was complete, we assessed of the completeness and contamination of these 12 genomes using CheckM (https://github. com/Ecogenomics/CheckM; Table S1, the data has been submitted to the National Microbiology Data Center, with the registration number: NMDCX0001747). The detailed genomic information is shown in Table 1. Bacterial pan-genome analysis (BPGA), an ultra-fast software package that provides detailed comprehensive information about microorganismal genomes, was used. Usearch was chosen as the clustering tool with a 50% sequence identity cutoff[19]. Gene clusters found in the genomes of all the analyzed strains were classified into the core genome. The accessory genome is composed of genes shared by 2−11 strains. The specific genome includes genes that exist only in one single strain of the species[20]. The pangenome and core genome profiles were evaluated with PanGP[21], using the gene presence-absence binary matrix (pan-matrix) obtained from BPGA as an input. The calculation of this matrix is based on similarity or dissimilarity among orthologous gene clusters[22].

1.2 Phylogenetic analysis

For the phylogenomic analysis, whole-genome sequences of *Rummeliibacillus* sp. strains were compared by computing the ANI (two strains are considered to belong to the same species if the ANI is greater than 95%) [23] and DDH (DDH>70% reflects the consistency of the genome type) [24]. Origin (https://www. originlab. com) and Adobe Illustrator 2020 (https://www. adobe. com/cn/products/illustrator) were used to draw related graphs. To construct the phylogenetic tree based on the core genome, Molecular evolutionary genetics analysis version 11 (MEGA 11) [25] software was used to align the concatenated amino acid sequences of the core genome, which were obtained from the BPGA pipeline analysis.

1.3 Functional annotation of the pangenome

Clusters of orthologous groups (COGs) of annotated proteins were generated using eggNOG-mapper to assign genes to COG categories[26]. Kyoto encyclopedia of genes and genomes (KEGG) orthology (KO) annotation of the pangenome genes was carried out *via* the KEGG automatic annotation server (KAAS) pipeline[27]. Metabolic pathways of *Rummeliibacillus* sp. were constructed *via* KEGG Mapper based on the assigned KO numbers[28]. To predict the MGEs in the *Rummeliibacillus* sp. genomes, the genomic islands (GIs) and prophage sequences were predicted. GIs were predicted using IslandViewer 4, which involved three methods: SIGI-HMM, IslandPath-DIMOB, and IslandPick[29]. Prophage sequences were annotated using PHASTER[30], and all instant prophage sequences were considered intact, questionable, or incomplete. The sequences of the core genome, the accessory genome, and the specific genome were compared with the virulence factor database (VFDB) to predict virulence factors of 12 strains of *Rummeliibacillus* sp.

2 Results and Discussion

2.1 Pangenome and core genome of *Rummeliibacillus* sp.

A total of 12 *Rummeliibacillus* sp. genomes were used for the pangenome analysis. Genome sizes ranged from 3.24 to 4.17 Mb. The average number of protein-coding genes was 3 404, and the G+C content ranged from 34.40% to 37.70% (Table 1). All protein-coding genes in the 12 genomes of *Rummeliibacillus* sp. were grouped into 8 024 gene clusters. Among them, 1 550 gene clusters were found in the genomes of all 12 strains, which constituted the core genome of *Rummeliibacillus* sp. (Figure 1). These genes may represent the common metabolic and physiological characteristics of *Rummeliibacillus* sp. The accessory genome includes 3 941 gene clusters, made up of genes present in two or more genomes, but not in all the genomes studied. The number of strain-specific genes in each genome ranged from one to 600 (Figure 1). *R. pycnus* and *Rummeliibacillus* sp. SL167 had the largest number of strain-specific genes (600 and 419, respectively). *R. stabekisii* DSM 25578 and *R. stabekisii* NBRC 104870 had the lowest number of strain-specific genes (one and six, respectively). *R. suwonensis* 3B-1 and *R. suwonensis* G20 contained 183 and 201 specific genes, respectively.

Analysis of the existence of open and closed

Table 1 *Rummeliibacillus* sp. isolates were used for pangenome analysis in this study

Strain	Country	Source	Size (Mb)	G+C content (%)	Proteins	Scaffolds	Accession number
DSM 15030	Germany	Soil	3.85	34.60	3 554	1	GCA_002884495.1
G20	Korea	Soil	4.11	35.90	3 638	16	GCA_007896435.1
3B-1	China	Pit mud	4.12	35.90	3 530	83	GCA_017578305.1
G93	China	Soil	3.24	37.70	3 121	1	GCA_023515935.1
POC4	Poland	Sewage sludge	3.69	34.50	3 570	163	GCA_003576525.1
SL167	Korea	Soil	4.17	35.80	3 757	14	GCA_007896505.1
TYF-LIM-RU47	China	Vinegar	3.30	37.40	3 273	14	GCA_008638315.1
TYF005	China	Vinegar	3.70	34.40	3 564	117	GCA_003844195.1
NBRC 104870	USA	Missing	3.26	37.30	3 150	23	GCA_007988965.1
PP9	Antarctica	Soil	3.42	37.69	3 334	2	GCA_001617605.1
DSM 25578	Missing	Missing	3.28	37.30	3 105	12	GCA_014202625.1
MER TA 13	USA	Missing	3.35	37.40	3 257	68	GCA_023713565.1

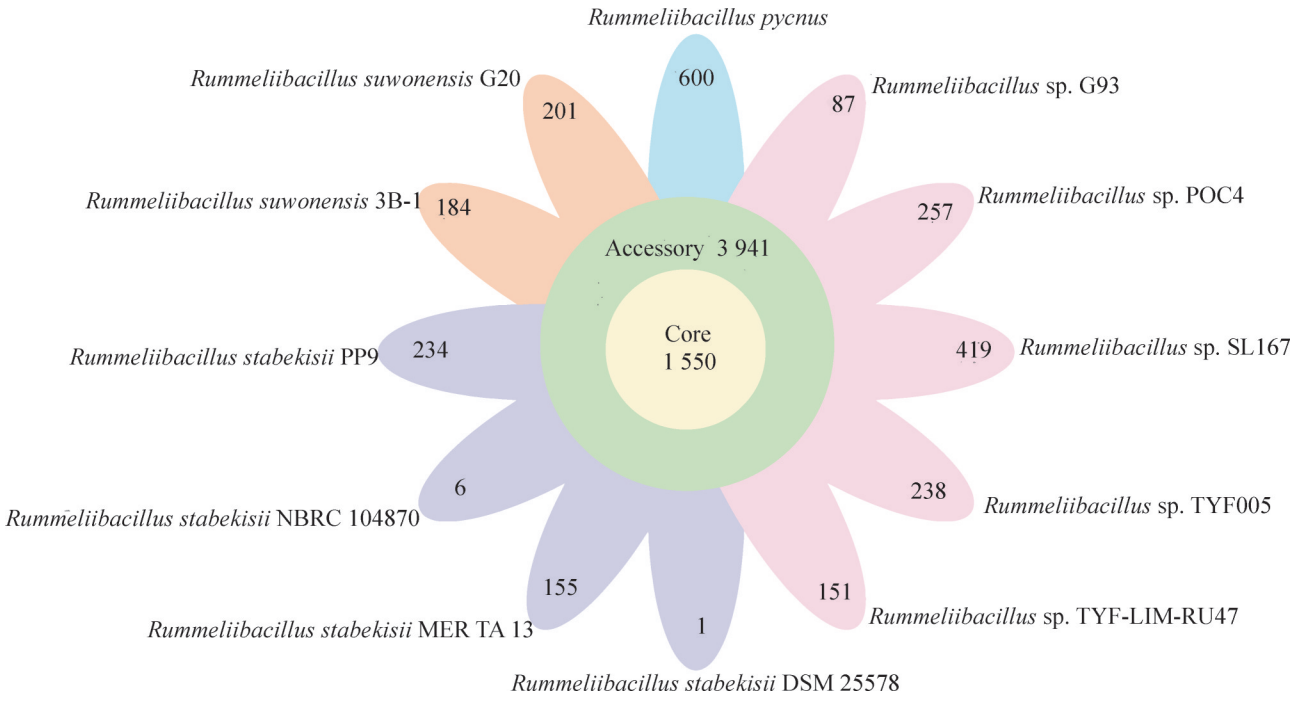

Figure 1 Pangenome of *Rummeliibacillus* sp. core and accessory genome sizes and the number of specific genes in each *Rummeliibacillus* sp. strain. The same color for a specific gene represents the same species of bacteria.

genomes can now be performed in many genera as the result of the burgeoning increase in microbial genome sequences from different strains within the same genus[31]. First, cumulative curves were generated by PanGP. The mathematical formula for pangenome size fitting is a power-law regression based on Heaps' law ($y=Ax^B+C$, where y denotes the number of genes of the pangenome, x denotes the analyzed genome number, and A, B, and C are fitting parameters). When $0<B<1$, the number of

genes of the pangenome increases when newly analyzed genomes are added, and the pangenome is considered as open. When B>1, the number of genes of the pangenome does not increase when newly analyzed genomes are added and the pangenome can be considered as closed. The mathematical formula for the number of genes of the core genome fitting is an exponential regression model ($y=Ae^{Bx}+C$, where y denotes the number of genes of the core genome, x denotes the number of analyzed genomes, and A, B, and C are fitting parameters) [32]. The fitted curves for the pangenome profile analysis of 12 strains of *Rummeliibacillus* sp. showed that the fitted exponent of the curve was positive ($0 \leqslant 0.2 < 1$), indicating that the *Rummeliibacillus* pangenome is open, which suggested that each added genome will contribute new genes and increase the number of gene clusters in the pangenome (Figure 2). Although this pangenome is obviously a mere mathematical extrapolation from the available sequenced strains, it makes clear the fact that some species exhibit extreme versatility in gene content.

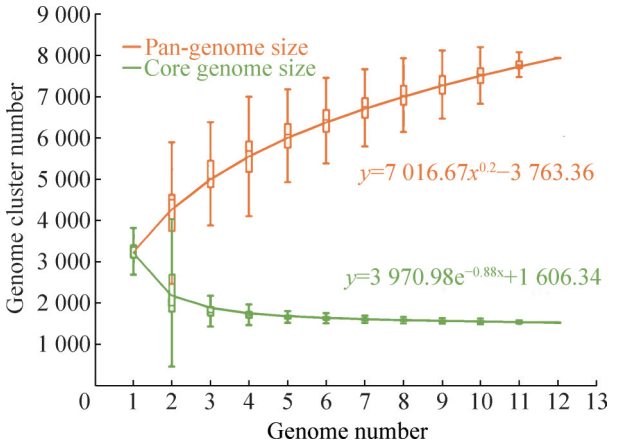

Figure 2 Development of pangenome and core genome sizes for *Rummeliibacillus* sp. strains with genome number varying from 1 to 12. The cumulative curve (in orange) supports an open pangenome.

2.2 Phylogenetic analysis of *Rummeliibacillus* sp.

A comparison of the ANI between unknown genera and known species was performed using JSpeciesWS[23] (https://jspecies. ribohost. com/jspeciesws/). JSpeciesWS calculated ANI between the genomes for a pairwise comparison using BLAST, and the results are shown in Figure 3A. The DDH values between the genomes were calculated using genome-to-genome distance calculator 3.0[24] (https://ggdc. dsmz. de/ggdc. php). The ANI and DDH values of *Rummeliibacillus* sp. G93 and *R. stabekisii* NBRC 104870 were 98.75% and 90.2%, respectively. The ANI and DDH values of *Rummeliibacillus* sp. G93 and *R. stabekisii* sp. DSM 25578 also were 98.75% and 90.30%, respectively. Therefore, *Rummeliibacillus* sp. G93 belongs to *R. stabekisii*. Moreover, to analyze the phylogenetic relationship of the 12 *Rummeliibacillus* sp. strains, phylogenetic trees were constructed based on the concatenated core gene alignments. In the phylogenetic tree, 12 strains were grouped into two main clades (Figure 3B). The tree is divided into two large branches, with *R. stabekisii* clustered on one branch and *R. suwonensis* and *R. pycnus* clustered on the other branch. Meanwhile, in the phylogenetic tree, *R. stabekisii* NBRC 104870 and *R. stabekisii* DSM 25578 were in the same branch together with *Rummeliibacillus* sp. G93, indicating that *Rummeliibacillus* sp. G93 belongs to *R. stabekisii*. The ANI and DDH values of *Rummeliibacillus* sp. TYF-LIM-RU47 and *R. stabekisii* MER TA 13 were 98.14% and 87.70%,

respectively. Meanwhile, *Rummeliibacillus* sp. TYF-LIM-RU47 and *R. stabekisii* MER TA 13 were in the same branch of the evolutionary tree, so it can be concluded that *Rummeliibacillus* sp. TYF-LIM-RU47 also belongs to *R. stabekisii*. In the other branch, *Rummeliibacillus* sp. POC4 and *Rummeliibacillus* sp. TYF005 had an ANI value of 98.41% and a DDH value of 86.50%, confirming that they are the same species. The ANI values of *Rummeliibacillus* sp. POC4 with *R. pycnus*, *R. suwonensis* 3B-1, *R. suwonensis* G20, and *Rummeliibacillus* sp. SL167 were 80.17%, 81.02%, 80.97%, and 81.09%, respectively. The DDH values of *Rummeliibacillus* sp. POC4 with *R. pycnus*, *R. suwonensis* 3B-1, and *R. suwonensis* G20 were only 25.50%, 22.40%, and 22.60%, respectively. This indicates that *Rummeliibacillus* sp. POC4 and *Rummeliibacillus* sp. TYF005 are neither *R. pycnus* nor *R. suwonensis* and that they may belong to a new species in this genus.

2.3 COG functional annotation of *Rummeliibacillus* sp. genes

COG analysis of pan-genomic gene clusters was conducted. Unknown function (S) was the largest category of the core genome, accessory genome, and strain-specific genes, accounting for 26.7%, 22.3%, and 29.8%, respectively (Figure 4). In terms of COG categories, most of the genes in the core genome are essential for life activities, such as transcription (K) (6.2%), translation, nucleosome structure, and biogenesis (J) (10.3%), amino acid transport and metabolism (E) (7.6%), energy production and conversion (C) (5.2%), replication, recombination, and repair (L) (6.5%), and cell wall/membrane/envelope biogenesis (M) (4.8%) (Figure 4). For the accessory genome, COG annotation showed that the largest categories were nucleotide transport and metabolism (F) (22.0%), transcription (K) (9.5%), and Inorganic ion transport and metabolism (P) (5.8%) (Figure 4).

2.4 Genomic metabolic pathway analysis of *Rummeliibacillus* sp.

The KAAS annotation showed that all the 1 550 core gene clusters (19.3%) were assigned with KO numbers. The most annotated gene families in the core genome belong to carbohydrate metabolism. For substrate transport, ATP-binding cassette (ABC) transporters and phosphotransferase systems (PTSs) were the main transporting systems annotated by KAAS. The number of genes distributed in metabolic pathways is shown in Table S2 (The data has been submitted to the National Microbiology Data Center, with the registration number: NMDCX0001748). In the carbohydrate metabolism pathway, 138 genes are annotated in the core genome, 195 genes are annotated in the accessory genome, and 54 genes are annotated in the specific genome.

2.4.1 Caproic acid and acetoin metabolism

In this study, the pathways involved in the KEGG pathway caproic acid and acetoin metabolism were constructed in *Rummeliibacillus* sp. Some glycolysis genes can be found in the core genome, and the missing genes *ptsG* and *pgi* exist in the accessory genome (*R. pycnus*, *Rummeliibacillus* sp. POC4, *Rummeliibacillus* sp. SL167, *Rummeliibacillus* sp. TYF005, *R. suwonensis* 3B-1, and *R. suwonensis* G20). Figure 5B shows the metabolic process of caproic acid and acetoin of *Rummeliibacillus* sp. strains. *R. pycnus*,

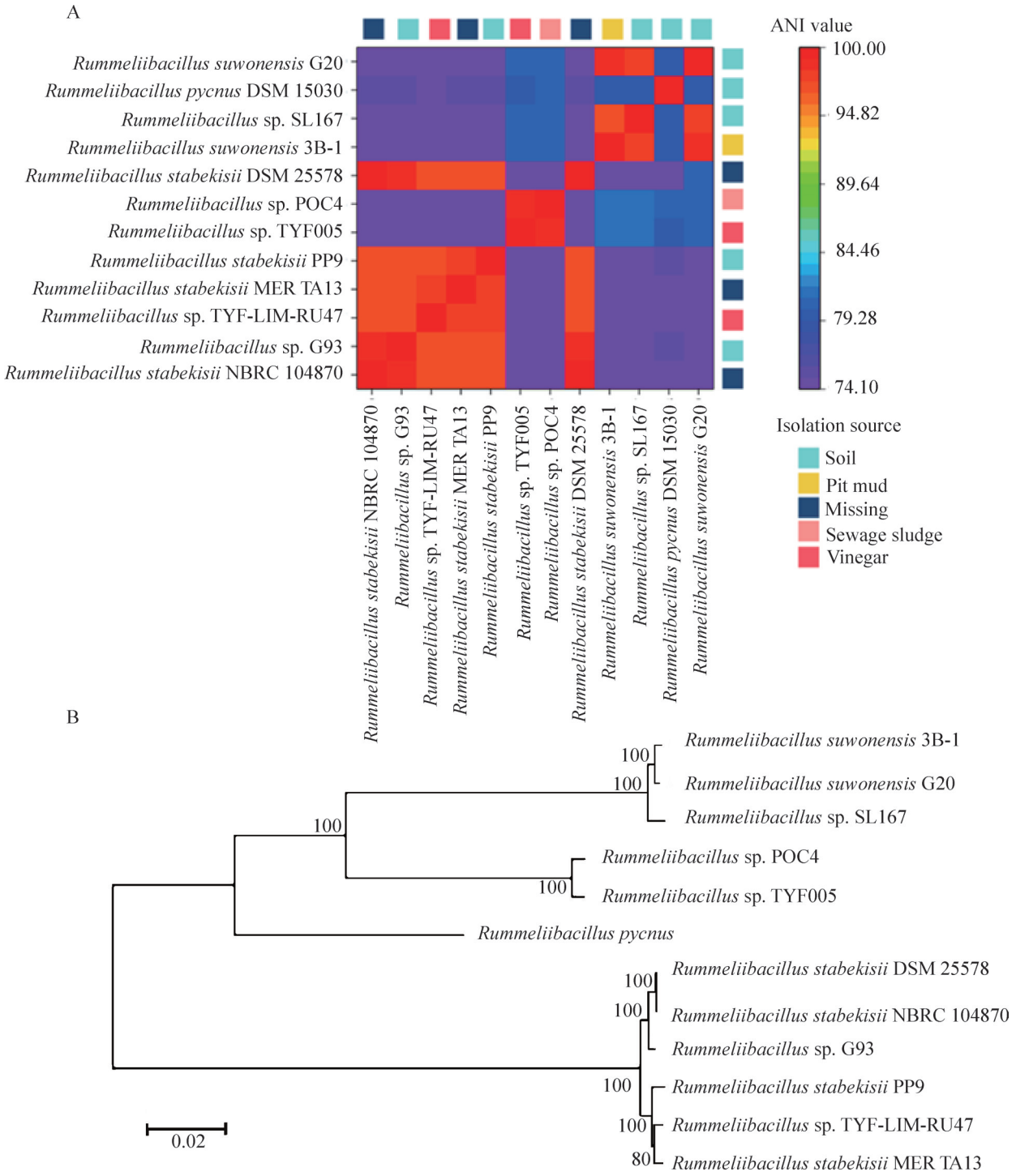

Figure 3 Comparison of the average nucleotide identity (ANI) values between the genomes of the 12 *Rummeliibacillus* sp. strains. A: Heatmaps display the ANI values between the 12 *Rummeliibacillus* sp. strains. The color represents the identity of strains, with purple indicating lower ANI values and red indicating higher ANI values. Color bars above and to the right of the heatmaps correspond to the source of each *Rummeliibacillus* sp. isolate. B: Phylogenetic trees of *Rummeliibacillus* sp. strains based on the core genome.

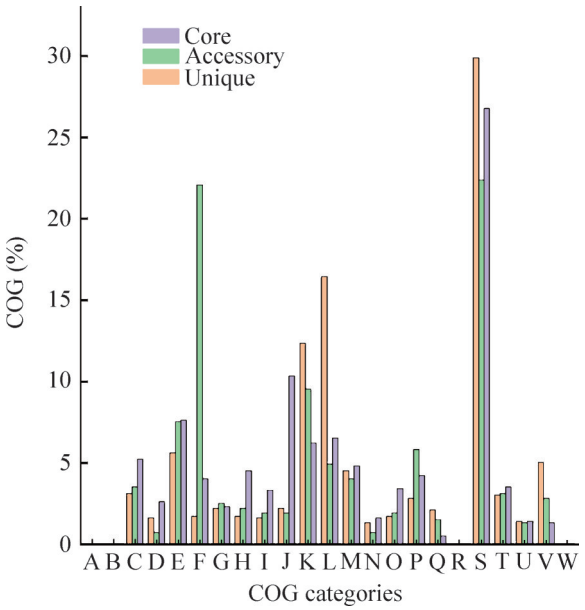

Figure 4 Distribution of COG categories between the core genome, accessory genome, and strain-specific genes of *Rummeliibacillus* sp. strains. A: RNA processing and modification; B: Chromatin structure and dynamics; C: Energy production and conversion; D: Cell cycle control, cell division, chromosome partitioning; E: Amino acid transport and metabolism; F: Nucleotide transport and metabolism; G: Carbohydrate transport and metabolism; H: Coenzyme transport and metabolism; I: Lipid transport and metabolism; J: Translation, ribosomal structure, and biogenesis; K: Transcription; L: Replication, recombination, and repair; M: Cell wall/membrane/envelope biogenesis; N: Cell motility; O: Posttranslational modification, protein turnover, chaperones; P: Inorganic ion transport and metabolism; Q: Secondary metabolite biosynthesis, transport, and catabolism; S: Function unknown; T: Signal transduction mechanisms; U: Intracellular trafficking, secretion, and vesicular transport; V: Defense mechanisms; W: Extracellular structures.

Rummeliibacillus sp. SL167, *Rummeliibacillus* sp. TYF005, *R. suwonensis* 3B-1, and *R. suwonensis* G20 could not synthesize acetoin due to the lack of the *alaS* gene, while the other seven strains had complete metabolic pathways. Based on existing reports, *Rummeliibacillus* sp. TYF-LIM-RU47 could indeed produce acetoin from various carbon sources, such as arabinose, xylose, glucose, xylan, and starch[10]. Therefore, they can synthesize acetoin.

In addition, the caproic acid biosynthesis pathway showed that most caproic acid synthesis genes were present in the core genome, but the *bcd* gene is found only in *R. pycnus* and *R. suwonensis* 3B-1 (Figure 5B). The other 10 strains did not have the ability to synthesize caproic acid. Currently, only *Rummeliibacillus* 3B-1 has been reported to synthesize caproic acid in this genus[8]. Acid production by *R. pycnus* has not been reported, so further research is needed to determine whether *R. pycnus* has the ability to synthesize caproic acid.

2.4.2 Arginine metabolism

Amino acid metabolism is another important metabolic pathway, with 126, 229 and 58 genes

annotated to the core genome, accessory genome, and specific genome, respectively, of this genus. Figure 5C shows the synthetic route of arginine. The genes in the whole arginine cycle pathway exist in the core genome. *Arc* is present in strains G93, TYF-LIM-RU47, DSM 25578, MER TA 13, NBRC 104870, and PP9, while *aspB* is present in strains SL167, G20, 3B-1, etc. (Figure 5D). One study reported that a new and heat-resistant arginase was found in *R. pycnus* SK31.001[4]. This shows that *Rummeliibacillus* sp. may have the ability to synthesize arginine. It is worth noting that the metabolic characteristics of arginine in *Rummeliibacillus* sp. are similar to those reported in *Escherichia coli*[33-34]. Glutamate is converted to ornithine, from which arginine is synthesized through the ornithine cycle.

2.4.3 Assimilatory sulfate reduction

In the accessory genome, the gene set for assimilatory sulfate reduction is complete in strains such as 3B-1, POC4, and G20 (Figure 5A, 5D). Sulfate can be absorbed and assimilated by bacteria and degraded to cysteine, and assimilative sulfur reduction by microorganisms provides a large amount of organic sulfur source for growth[35]. The related genes (*sat*, *cysC*, *cysH*, etc.) mainly exist in *R. pycnus* and the strains POC4, SL167, TYF005, 3B-1, and G20, which indicates that the strains may have sulfate reducing effects. This suggests that *R. stabekisii* is capable of biomineralization[7]. Therefore, it can be inferred that *Rummeliibacillus* may play an important role in soil maintenance.

2.5 Virulence factor analysis

The sequences of the core genome, the accessory genome, and the specific genome were compared with the VFDB database. In the *Rummeliibacillus* sp. pangenome, 38 virulence genes were identified in total. Of these, 13 core virulence genes were shared by all strains and four unique virulence factors were present in one strain each (Table S3, the data has been submitted to the National Microbiology Data Center, with the registration number: NMDCX0001749). *Rummeliibacillus* sp. SL167 (from soil) had the highest number of virulence genes (32), and *R. pycnus* (from soil) had the lowest number of virulence genes (19). The genomes of all 12 *Rummeliibacillus* sp. strains harbor genes encoding virulence factors, which are involved in processes including adhesion (*flmH* and *slrA*), secretion (*clpB* and *cdsN*), regulation (*cheY* and *lisR*), and motility (*fliQ*). Adhesion-related genes can promote adhesion and biofilm formation, which is an important factor in the pathogenesis of *Streptococcus*[36]. The adhesion gene *slrA* encodes many surface proteins[37]. These surface proteins have been identified as important virulence factors, involving the adhesion of bacteria to the epithelial cells of host cells, mediated by microbial surface components that recognize adhesion matrix molecules, thus contributing to host cell adhesion and tissue colonization[38]. However, studies have shown that although SlrA is involved in colonization, it does not contribute significantly to invasive pneumococcal disease[39]. Moreover, *R. pycnus*, *Rummeliibacillus* sp. POC4, *R. suwonensis* 3B-1, and *R. suwonensis* G20 carried three toxic genes (*cylR2*, *cysC1*, and *hlyIII*). *Rummeliibacillus* sp. TYF-LIM-RU47, *R. stabekisii* DSM 25578, *R. stabekisii* MER TA 13, and *R. stabekisii* NBRC

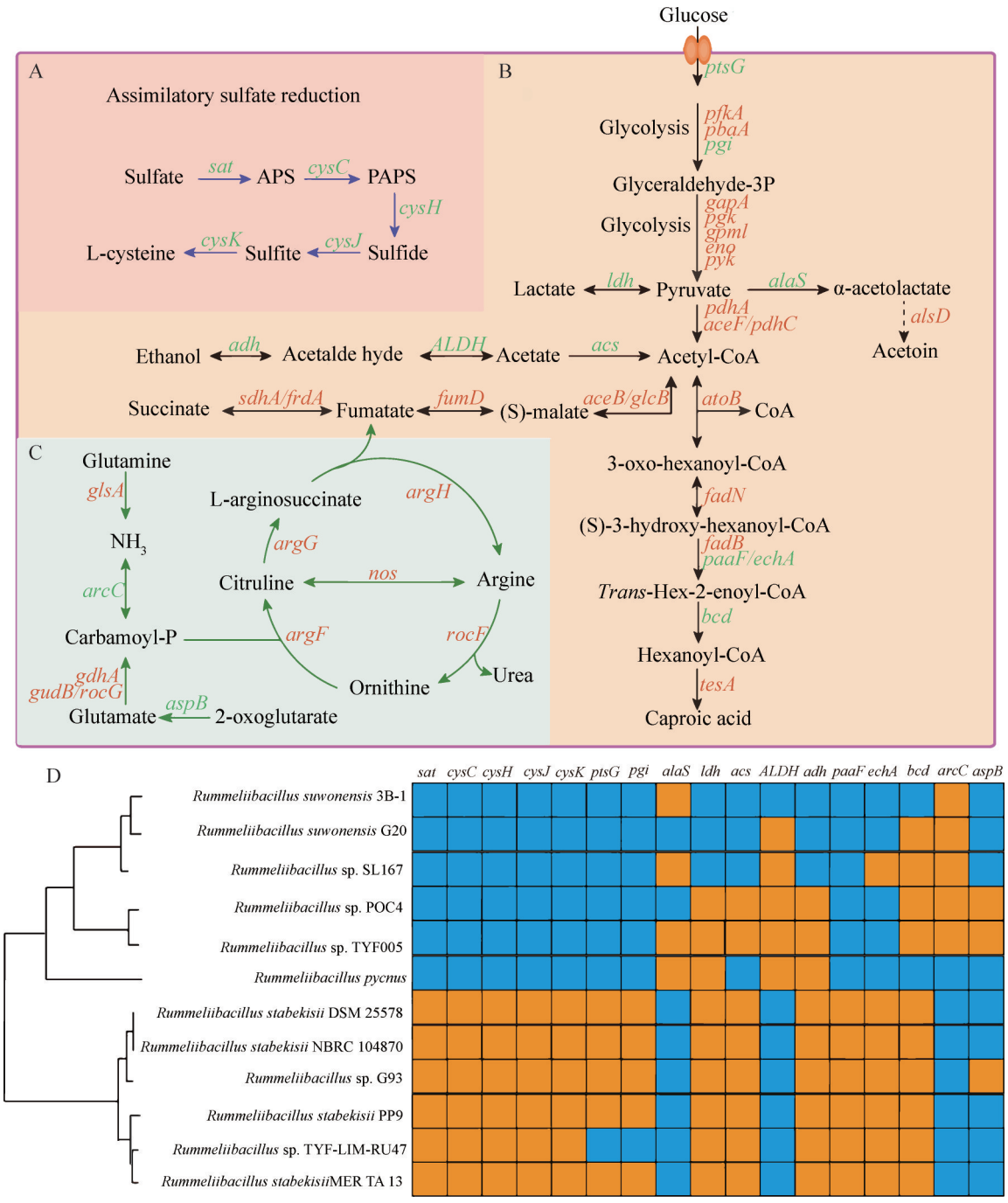

Figure 5 Overview of caproic acid, arginine, and sulfur metabolism in *Rummeliibacillus* sp. strains and distribution of accessory genome genes. A: Assimilatory sulfate reduction and main carbon metabolism annotated in the pangenome of *Rummeliibacillus* sp. strains; B: Genes annotated in caproic acid fermentation and metabolism of acetoin. Genes shared by all the 12 genomes (core genome) are shown in red, and genes in the accessory genome are shown in green; C: Distribution of genes involved in arginine metabolism; D: The distribution of genes in the accessory genome among different strains (blue represents the presence of the gene, while orange represents the absence of the gene).

104870 contain only one toxic gene, *hlyIII*. The rest contain two toxic genes (*cylR2* and *hlyIII*). They all have a virulence gene, *hlyIII*, which encodes an integral outer membrane protein with hemolytic activity that forms pores[40-41]. However, the expression of enterococcal hemolysin requires the complete set of CylR2, CylA, CylB, and eight other proteins[41]. Similarly, enterotoxin is not toxic when it is expressed alone[42]. Therefore, it can be said that *Rummeliibacillus* are non-pathogenic bacteria.

2.6 Prediction of MGEs

To study the MGEs in *Rummeliibacillus* sp., we used IslandViewer 4 (an integrated interface for computational identification and visualization of GIs). MGEs can mediate the acquisition of DNA and promote the expansion of the gene pool of bacterial groups[43]. The number of GIs in the genome of *Rummeliibacillus* sp. ranged from four (*R. stabekisii* DSM 25578 and *R. stabekisii* NBRC 104870) to 14 (*Rummeliibacillus* sp. SL167 and *Rummeliibacillus* sp. TYF005), indicating that MGEs are widespread in *Rummeliibacillus* sp. (Table S4, the data has been submitted to the National Microbiology Data Center, with the registration number: NMDCX0001750). *R. suwonensis* sp. G20 had the largest total length of GIs, occupying 8.22% of its genome size (4.11 Mb). *Rummeliibacillus* sp. G93 had the smallest total length of GIs, occupying 3.62% of its genome size (3.24 Mb).

In addition, the genomes of *Rummeliibacillus* sp. in this study were scanned using the PHASTER online service to obtain phage sequences[29]. After searching the phages of 12 *Rummeliibacillus* sp. strains, a total of eight regions were intact, eight regions were questionable, and 26 regions were incomplete (Table 2). *R. stabekisii* PP9 was most complete, and three hypothetical phage groups were detected, including PHAGE_Paenib_Vegas (NC_028767) and PHAGE_Aeriba_AP45 (NC_048651) (Table 3). However, all regions in *R. suwonensis* 3B-1 and *R. stabekisii* DSM 25578 were incomplete. PHAGE_Aeriba_AP45 (NC_048651) was found in *Rummeliibacillus* sp. G93, *Rummeliibacillus* sp. TYF-LIM-RU47, and *R.*

Table 2 Number of prophage regions (considered as intact, questionable or incomplete)

Strain	Intact	Questionable	Incomplete
Rummeliibacillus pycnus DSM 15030	0	1	0
Rummeliibacillus suwonensis G20	0	1	2
Rummeliibacillus suwonensis 3B-1	0	0	2
Rummeliibacillus sp. G93	1	0	0
Rummeliibacillus sp. POC4	0	2	4
Rummeliibacillus sp. SL167	1	0	2
Rummeliibacillus sp. TYF-LIM-RU47	2	1	4
Rummeliibacillus sp. TYF005	1	0	1
Rummeliibacillus stabekisii NBRC 104870	0	0	4
Rummeliibacillus stabekisii PP9	3	1	1
Rummeliibacillus stabekisii DSM 25578	0	0	4
Rummeliibacillus stabekisii MER TA 13	0	2	2

Table 3 Phage information for all *Rummeliibacillus* sp. strains

Strain	Region length (kb)	Total proteins	Most common phage (NCBI accession)	G+C content (%)
R. pycnus DSM 15030	37.4	23	PHAGE_Staphy_6ec (NC_024355)	31.75
R. suwonensis G20	50.3	43	PHAGE_Lister_LP_101 (NC_024387)	31.07
Rummeliibacillus sp. G93	43.9	78	PHAGE_Aeriba_AP45 (NC_048651)	36.52
Rummeliibacillus sp. POC4	35.1	58	PHAGE_Bacill_phBC6A52 (NC_004821)	33.71
Rummeliibacillus sp. POC4	18.6	25	PHAGE_Bacill_SPP1 (NC_004166)	36.28
Rummeliibacillus sp. SL167	55.5	37	PHAGE_Geobac_GBK2 (NC_023612)	42.43
Rummeliibacillus sp. TYF-LIM-RU47	24.7	44	PHAGE_Aeriba_AP45 (NC_048651)	36.96
Rummeliibacillus sp. TYF-LIM-RU47	34.3	52	PHAGE_Lister_B054 (NC_009813)	37.10
Rummeliibacillus sp. TYF005	49.0	72	PHAGE_Bacill_1 (NC_009737)	33.43
R. stabekisii PP9	54.2	78	PHAGE_Paenib_Vegas (NC_028767)	36.81
R. stabekisii PP9	47.0	63	PHAGE_Paenib_Vegas (NC_028767)	36.85
R. stabekisii PP9	44.9	75	PHAGE_Aeriba_AP45 (NC_048651)	37.01
R. stabekisii MER TA 13	17.4	20	PHAGE_Bacill_SPP1 (NC_004166)	38.77
R. stabekisii MER TA 13	31.9	23	PHAGE_Paenib_Vegas (NC_028767)	34.24

stabekisii PP9, suggesting that the phage played an important role in the evolution and diversity of these strains.

3 Conclusions

This work represents the first characterization of *Rummeliibacillus* species using pan-genomic analysis. The pangenome of *Rummeliibacillus* sp. strains is open, and the addition of newly sequenced genomes could increase the number of genes and the size of the pangenome. The pathway for arginine metabolism was discovered in all 12 *Rummeliibacillus* sp. strains, and only two strains had the ability to completely metabolize caproic acid, while seven strains had the ability to completely biosynthesize acetoin. Additionally, a complete assimilation sulfate reduction process was found in six strains. MGEs, including bacteriophages and GIs, were detected in the pangenome of *Rummeliibacillus* sp., which might give rise to horizontal gene transfer for environmental adaptation and increased genome diversity of *Rummeliibacillus* sp. This study provides insights for future research on the genetics and practical application of *Rummeliibacillus* sp.

Acknowledgments

We thank International Science Editing (http://www.internationalscienceediting.com) for editing this manuscript.

Author Contributions

ZOU Wei: Conceptualization, formal analysis, supervision, methodology, writing – review & editing. YANG Lingling: Formal analysis, visualization, writing – review & editing. LIU Chaojie: Methodology, formal analysis, visualization, writing – original draft preparation, data curation. ZHENG Jia: validation. ZHANG Kaizheng: Investigation. QIAO Zongwei: Investigation.

Conflicts of Interest

The authors declare that there is no conflict of interest.

References

[1] VAISHAMPAYAN P, MIYASHITA M, OHNISHI A, SATOMI M, ROONEY A, DUC MT, VENKATESWARAN K. Description of *Rummeliibacillus stabekisii* gen. nov., sp. nov. and reclassification of *Bacillus pycnus* Nakamuraet al. 2002 as *Rummeliibacillus pycnus* comb. nov.[J]. International Journal of Systematic and Evolutionary Microbiology, 2009, 59(Pt5): 1094-1099.

[2] HER J, KIM J. *Rummeliibacillus suwonensis* sp. nov., isolated from soil collected in a mountain area of Korea[J]. Journal of Microbiology, 2013, 51(2): 268-272.

[3] LI M, LI Y, FAN XJ, QIN YH, HE YJ, LV YK. Draft genome sequence of *Rummeliibacillus* sp. strain TYF005, a physiologically recalcitrant bacterium with high ethanol and salt tolerance isolated from spoilage vinegar[J]. Microbiology Resource Announcements, 2019, 8(31): e00244-19.

[4] HUANG K, ZHANG T, JIANG B, MU WM, MIAO M. Characterization of a thermostable arginase from *Rummeliibacillus pycnus* SK31. 001[J]. Journal of Molecular Catalysis B: Enzymatic, 2016, 133: S68-S75.

[5] HUANG K, ZHANG T, JIANG B, YAN X, MU WM, MIAO M. Overproduction of *Rummeliibacillus pycnus* arginase with multi-copy insertion of the $arg^{R.pyc}$ cassette into the *Bacillus subtilis* chromosome[J]. Applied Microbiology and Biotechnology, 2017, 101(15): 6039-6048.

[6] JUNPADIT P, SUKSAROJ TT, BOONSAWANG P. Transformation of palm oil mill effluent to terpolymer polyhydroxyalkanoate and biodiesel using *Rummeliibacillus pycnus* strain TS_8[J]. Waste and Biomass Valorization, 2017, 8(4): 1247-1256.

[7] MUDGIL D, BASKAR S, BASKAR R, PAUL D, SHOUCHE YS. Biomineralization potential of *Bacillus subtilis*, *Rummeliibacillus stabekisii* and *Staphylococcus epidermidis* strains *in vitro* isolated from speleothems, Khasi Hill Caves, Meghalaya, India[J]. Geomicrobiology Journal, 2018, 35(8): 675-694.

[8] LIU CJ, DU YF, ZHENG J, QIAO ZW, LUO HB, ZOU W. Production of caproic acid by *Rummeliibacillus suwonensis* 3B-1 isolated from the pit mud of strong-flavor Baijiu[J]. Journal of Biotechnology, 2022, 358: 33-40.

[9] ZHANG YY, PAN XR, ZUO JE, HU JM. Production of n-caproate using food waste through thermophilic fermentation without addition of external electron donors[J]. Bioresource Technology, 2022, 343: 126144.

[10] FENG GY, FAN XJ, LIANG YN, LI C, XING JD, HE YJ. Genomic and transcriptional characteristics of strain *Rummeliibacillus* sp. TYF-LIM-RU47 with an aptitude of directly producing acetoin from lignocellulose[J]. Fermentation, 2022, 8(8): 414.

[11] YANG G, HU YM, WANG JL. Biohydrogen production from co-fermentation of fallen leaves and sewage sludge[J]. Bioresource Technology, 2019, 285: 121342.

[12] TINRAT S, SEDTANANUN S. Novel *Rummeliibacillus* sp. isolated from fermented vegetable products as the potential probiotics[J]. Journal of Microbiology, Biotechnology and Food Sciences, 2022, 11(5): e4194.

[13] TAN HY, CHEN SW, HU SY. Improvements in the growth performance, immunity, disease resistance, and gut microbiota by the probiotic *Rummeliibacillus stabekisii* in Nile tilapia (*Oreochromis niloticus*)[J]. Fish & Shellfish Immunology, 2019, 92: 265-275.

[14] LIVINGSTONE PG, MORPHEW RM, WHITWORTH DE. Genome sequencing and pan-genome analysis of 23 *Corallococcus* spp. strains reveal unexpected diversity, with particular plasticity of predatory gene sets[J]. Frontiers in Microbiology, 2018, 9: 3187.

[15] LAND M, HAUSER L, JUN SR, NOOKAEW I, LEUZE MR, AHN TH, KARPINETS T, LUND O, KORA G, WASSENAAR T, POUDEL S, USSERY DW. Insights from 20 years of bacterial genome sequencing[J]. Functional & Integrative Genomics, 2015, 15(2): 141-161.

[16] GOLICZ AA, BAYER PE, BHALLA PL, BATLEY J, EDWARDS D. Pangenomics comes of age: from bacteria to plant and animal applications[J]. Trends in Genetics, 2020, 36(2): 132-145.

[17] MEDINI D, DONATI C, TETTELIN H, MASIGNANI V, RAPPUOLI R. The microbial pan-genome[J]. Current Opinion in Genetics & Development, 2005, 15(6): 589-594.

[18] YIN ZQ, LIU XB, QIAN CQ, SUN L, PANG SQ, LIU JN, LI W, HUANG WW, CUI SY, ZHANG CK, SONG WX, WANG DD, XIE ZH. Pan-genome analysis of *Delftia tsuruhatensis* reveals important traits concerning the genetic diversity, pathogenicity, and biotechnological properties of the species[J]. Microbiology Spectrum, 2022, 10(2): e02072-21.

[19] CHAUDHARI NM, GUPTA VK, DUTTA C. BPGA-an ultra-fast pan-genome analysis pipeline[J]. Scientific Reports, 2016, 6: 24373.

[20] PERIWAL V, PATOWARY A, VELLARIKKAL SK, GUPTA A, SINGH M, MITTAL A, JEYAPAUL S, CHAUHAN RK, SINGH AV, SINGH PK, GARG P, KATOCH VM, KATOCH K, CHAUHAN DS, SIVASUBBU S, SCARIA V. Comparative whole-genome analysis of clinical isolates reveals characteristic architecture of *Mycobacterium tuberculosis* pangenome [J]. PLoS One, 2015, 10(4): e0122979.

[21] ZHAO YB, JIA XM, YANG JH, LING YC, ZHANG Z, YU J, WU JY, XIAO JF. PanGP: a tool for quickly analyzing bacterial pan-genome profile[J]. Bioinformatics, 2014, 30(9): 1297-1299.

[22] ZEB S, GULFAM SM, BOKHARI H. Comparative core/pan genome analysis of *Vibrio cholerae* isolates from Pakistan[J]. Infection, Genetics and Evolution, 2020, 82:

104316.
[23] RICHTER M, ROSSELLO-MORA R, OLIVER GLOCKNER FO, PEPLIES J. JSpeciesWS: a web server for prokaryotic species circumscription based on pairwise genome comparison[J]. Bioinformatics, 2016, 32(6): 929-931.
[24] AUCH AF, von JAN M, KLENK HP, GOKER M. Digital DNA-DNA hybridization for microbial species delineation by means of genome-to-genome sequence comparison[J]. Standards in Genomic Sciences, 2010, 2: 117-134.
[25] TAMURA K, STECHER G, KUMAR S. MEGA 11: molecular evolutionary genetics analysis version 11[J]. Molecular Biology and Evolution, 2021, 38(7): 3022-3027.
[26] HUERTA-CEPAS J, SZKLARCZYK D, HELLER D, HERNANDEZ-PLAZA A, FORSLUND SK, COOK H, MENDE DR, LETUNIC I, RATTEI T, JENSEN LJ, MERING CV, BORK P. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses[J]. Nucleic Acids Research, 2019, 47(D1): gD309-D314.
[27] MORIYA Y, ITOH M, OKUDA S, YOSHIZAWA AC, KANEHISA M. KAAS: an automatic genome annotation and pathway reconstruction server[J]. Nucleic Acids Research, 2007, 35(suppl_2): W182-W185.
[28] KANEHISA M, SATO Y. KEGG Mapper for inferring cellular functions from protein sequences[J]. Protein Science, 2020, 29(1): 28-35.
[29] BERTELLI C, LAIRD MR, WILLIAMS KP, Simon Fraser University Research Computing Group, LAU BY, HOAD G, WINSOR GL, BRINKMAN FSL. IslandViewer 4: expanded prediction of genomic islands for larger-scale datasets[J]. Nucleic Acids Research, 2017, 45(W1): W30-W35.
[30] ARNDT D, GRANT JR, MARCU A, SAJED T, PON A, LIANG YJ, WISHART DS. PHASTER: a better, faster version of the PHAST phage search tool[J]. Nucleic Acids Research, 2016, 44(W1): W16-W21.
[31] MIRA A, MARTIN-CUADRADO AB, D'AURIA G, RODRGUEZ-VALERA F. The bacterial pan-genome: a new paradigm in microbiology[J]. International Microbiology, 2010, 13(2): 45-57.
[32] TETTELIN H, MASIGNANI V, CIESLEWICZ MJ, DONATI C, MEDINI D, WARD NL, ANGIUOLI SV, CRABTREE J, JONES AL, DURKIN AS, DeBOY RT, DAVIDSEN TM, MORA M, SCARSELLI M, MARGARIT Y ROS I, PETERSON JD, HAUSER CR, SUNDARAM JP, NELSON WC, MADUPU R, et al. Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial "pan-genome"[J]. Proceedings of the National Academy of Sciences, 2005, 102(39): 13950-13955.
[33] CHARLIER D, BERVOETS I. Regulation of arginine biosynthesis, catabolism and transport in *Escherichia coli* [J]. Amino Acids, 2019, 51(8): 1103-1127.
[34] GINESY M, BELOTSERKOVSKY J, ENMAN J, ISAKSSON L, ROVA U. Metabolic engineering of *Escherichia coli* for enhanced arginine biosynthesis[J]. Microbial Cell Factories, 2015, 14: 29.
[35] WANG YX, WU Y, ZHANG HL, QU XH, XIN YF. Driven by microbial metabolism of sulfur and its biological ecological relationship[J]. Journal of Microbiology, Biotechnology and Food Sciences, 2022, 62(3): 930-948.
[36] WIDGREN S, FROSSLING J. Spatio-temporal evaluation of cattle trade in Sweden: description of a grid network visualization technique[J]. Geospatial Health, 2010, 5(1): 119-130.
[37] BOBER M, MORGELIN M, OLIN AI, von PAWEL-RAMMINGEN U, COLLIN M. The membrane bound LRR lipoprotein Slr, and the cell wall-anchored M1 protein from *Streptococcus pyogenes* both interact with type I collagen[J]. PLoS One, 2011, 6(5): e20345.
[38] XU SY, LIU Y, GAO J, ZHOU M, YANG JY, HE FM, KASTELIC JP, DENG ZJ, HAN B. Comparative genomic analysis of *Streptococcus dysgalactiae* subspecies *dysgalactiae* isolated from bovine mastitis in China[J]. Frontiers in Microbiology, 2021, 12: 751863.
[39] HERMANS PWM, ADRIAN PV, ALBERT C, ESTEVAO S, HOOGENBOEZEM T, LUIJENDIJK IHT, KAMPHAUSEN T, HAMMERSCHMIDT S. The streptococcal lipoprotein rotamase A (SlrA) is a functional peptidyl-prolyl isomerase involved in pneumococcal colonization[J]. Journal of Biological Chemistry, 2006, 281(2): 968-976.
[40] BAIDA GE, KUZMIN NP. Mechanism of action of hemolysin III from *Bacillus cereus*[J]. Biochimica et Biophysica Acta (BBA) -Biomembranes, 1996, 1284(2): 122-124.
[41] CHEN YC, CHANG MC, CHUANG YC, JEANG CL. Characterization and virulence of hemolysin III from *Vibrio vulnificus*[J]. Current Microbiology, 2004, 49(3): 175-179.
[42] JIA WJ, SONG LL, ZHANG LY, WANG XL. The latest research progress of *Bacillus cereus* toxin[J]. Chinese Journal of Antibiotics, 2022, 47(06):537-542.
[43] OCHMAN H, LAWRENCE JG, GROISMAN EA. Lateral gene transfer and the nature of bacterial innovation[J]. Nature, 2000, 405(6784): 299-304.