Near-Complete Whole-Genome Sequence of *Paenibacillus* sp. nov. Strain J5C2022, a Sucretolerant and Endospore-Forming Bacterium Isolated from Highly Concentrated Sugar Brine

Siti Munirah Musa,[a] Shing Wei Siew,[a] Darren Dean Tay,[a] ⓘ Hajar Fauzan Ahmad[a]

[a]Faculty of Industrial Sciences and Technology, Universiti Malaysia Pahang, Gambang, Pahang, Malaysia

Shing Wei Siew and Darren Dean Tay contributed equally to this article.

ABSTRACT   Here, we present a 7.62-Mbp genome sequence of *Paenibacillus* sp. nov. strain J5C2022, a Gram-positive facultatively anaerobic bacterium that was isolated from 4-month-old fruit pickle brine and sequenced using the Illumina platform.

Gram-positive *Paenibacillus* spp., in the phylum *Firmicutes*, are endospore-forming bacteria that can switch between metabolically active or dormant states in response to stress or favorable conditions (1). They are facultatively anaerobic bacteria that grow optimally at temperatures between 28°C and 40°C under neutral pH conditions (2, 3). Strain J5C2022 was isolated from an overnight culture (on nutrient agar) of 4-month-old fruit pickle brine collected from a manufacturer in Penang, Malaysia. Briefly, 500 mL of brine was centrifuged at high speed, and the supernatant was discarded. The concentrated brine (100 μL) was spread onto a nutrient agar plate at 33°C for overnight cultivation. A single colony was picked randomly and identified using Sanger sequencing. Upon screening, strain J5C2022 was transferred to nutrient broth at 33°C for overnight cultivation to obtain a pure culture. Genomic DNA was extracted from the strain using the GenElute bacterial genomic DNA kit (Sigma-Aldrich Co., St. Louis, USA) as previously described (4). For the library preparation, 100 ng of DNA was measured and fragmented to 350 bp using a Bioruptor sonicator, which is compatible with the NEBNext Ultra II DNA library prep kit for Illumina (New England BioLabs Inc., Ipswich, USA). The sequencing was performed on the NovaSeq 6000 sequencing system (Illumina Inc., San Diego, USA) and produced 789,064,800 bp of paired-end reads (2 × 150 bp). The raw Illumina paired-end reads were trimmed using fastp v0.21 (5) to remove low-quality bases and Illumina adapter sequences. The trimmed reads were subsequently used for *de novo* assembly using Unicycler v0.4.8 (6). Contigs smaller than 500 bp, representing mostly sequencing artefacts, were removed, and the filtered assembly was used for subsequent analysis. Genome assembly statistics were generated using QUAST (7). The corresponding rRNA genes (5S, 16S, and 23S rRNA) were extracted using Barrnap (https://github.com/tseemann/barrnap) into a single FASTA file that was subjected to a BLAST search of the GTDB 16S rRNA database. The completeness of the genome was then analyzed using BUSCO v5 (https://gitlab.com/ezlab/busco), based on the Bacillales_odb10 lineage database (8). Unless otherwise noted, default parameters were used for all software tools. Strain J5C2022 has a total length of 7,616,378 bp, an N_{50} length of 137,040 bp, an average GC content of 51.6%, and 200× genome coverage. Genome annotation was carried out using the Rapid Annotations using Subsystems Technology (RAST) server (9) and the NCBI Prokaryotic Genome Annotation Pipeline (PGAP) with the GeneMarkS2 v1.14_1.25 method (10), resulting in 138 contigs with 6,340 coding sequences and 70 RNAs. A total of 856 genes were screened to be potentially involved in carbohydrate metabolism, covering

the xylose and L-rhamnose utilization pathways. In addition to this strain's ability to form endospores and its resistance to high salt and sugar conditions, its potential for use as a next-generation probiotic for application in the food industry is worth exploring.

Data availability. The complete genome sequence of *Paenibacillus* sp. nov. strain J5C2022 is available at NCBI under SRA accession number SRR19238564 and BioProject accession number PRJNA83215. The BioSample accession number is SAMN28409047. The GenBank accession number is JAOQJC000000000.1, with assembly accession number GCF_025567625.1.

ACKNOWLEDGMENTS

REFERENCES

1. Mahdi OS, Fisher NA. 2018. Sporulation and germination of Paenibacillus larvae cells. Curr Protoc Microbiol 48:9E.2.1–9E.2.10. https://doi.org/10.1002/cpmc.46.
2. Ash C, Priest FG, Collins MD. 1993. Molecular identification of rRNA group 3 bacilli (Ash, Farrow, Wallbanks and Collins) using a PCR probe test. Proposal for the creation of a new genus Paenibacillus. Antonie Van Leeuwenhoek 64:253–260. https://doi.org/10.1007/BF00873085.
3. Yoon J-H, Kang S-J, Yeo S-H, Oh T-K. 2005. Paenibacillus alkaliterrae sp. nov., isolated from an alkaline soil in Korea. Int J Syst Evol Microbiol 55:2339–2344. https://doi.org/10.1099/ijs.0.63771-0.
4. Ahmad HF, Schreiber L, Marshall IPG, Andersen PJ, Castro-Mejía JL, Nielsen DS. 2019. Draft genome sequence of Streptococcus anginosus strain CALM001, isolated from the gut of an elderly Dane. Microbiology Resource Announcements 8:e00379-19. https://doi.org/10.1128/MRA.00379-19.
5. Chen S, Zhou Y, Chen Y, Gu J. 2018. Fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 34:i884–i890. https://doi.org/10.1093/bioinformatics/bty560.
6. Wick RR, Judd LM, Gorrie CL, Holt KE. 2017. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. PLoS Comput Biol 13:e1005595. https://doi.org/10.1371/journal.pcbi.1005595.
7. Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. Bioinformatics 29:1072–1075. https://doi.org/10.1093/bioinformatics/btt086.
8. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31:3210–3212. https://doi.org/10.1093/bioinformatics/btv351.
9. Aziz RK, Bartels D, Best A, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O. 2008. The RAST server: Rapid Annotations using Subsystems Technology. BMC Genomics 9:75. https://doi.org/10.1186/1471-2164-9-75.
10. Lomsadze A, Gemayel K, Tang S, Borodovsky M. 2018. Modeling leaderless transcription and atypical genes results in more accurate gene prediction in prokaryotes. Genome Res 28:1079–1089. https://doi.org/10.1101/gr.230615.117.