| Bacteriology | Announcement

Draft genome sequence of multidrug-resistant *Bacillus subtilis* strain Hakim RU_LJ isolated from lacchi juice in Bangladesh

Muhib Ullah Khan,[1] Md. Arif-Uz-Zaman Polash,[1] Nusrat Zahan,[1] Md. Mottaleb Ali,[1] Nilima Rubaba Azad,[1] Subir Sarker,[2,3] Md. Hakimul Haque[1,2]

AUTHOR AFFILIATIONS See affiliation list on p. 3.

ABSTRACT   This report details the draft genome sequence of the multidrug-resistant *Bacillus subtilis* strain Hakim RU_LJ derived from lacchi juice. The assembled genome spans 4.03 Mb, with 50.33× coverage and a GC content of 43.5%. Analysis revealed six CRISPR arrays, 19 prophages, 17 antibiotic resistance genes, and 10 virulence factor genes.

*B*acillus subtilis, an aerobic, endospore-forming, Gram-positive bacterium with Generally Recognized as Safe (GRAS) status, can contaminate fermented beverages, such as lacchi juice, through soil, water, or raw ingredients (1–3). Its ability to thrive during fermentation is attributed to its spore-forming nature and adaptability. The bacterium's multidrug resistance is likely driven by environmental exposure to antibiotics, which facilitates resistance development through mutations or horizontal gene transfer (4). Continuous antimicrobial resistance (AMR) surveillance is crucial for assessing potential risks, enhancing food safety measures, and unlocking biotechnological applications (5, 6).

In May 2024, with authorization from the Institute of Biological Sciences (IBScs) at the University of Rajshahi, Bangladesh (Memo No. 56/321/IAMEBBC/IBScs), we collected samples of lacchi juice from street vendors near the university (24.3733°N, 88.6049°E) using standard protocols. The samples were thoroughly mixed, transferred into sterile containers, and promptly delivered to the laboratory (7). *Bacillus subtilis* was isolated by culturing the samples on nutrient agar (HiMedia, India) and blood agar (HiMedia, India), with aerobic incubation at 37°C for 18–24 hours, followed by microscopic staining and biochemical analyses (8). The antimicrobial resistance patterns of the isolates were evaluated against 10 widely used antibiotics through the disk diffusion technique (9), adhering to CLSI standards (10). A multidrug-resistant strain, showing resistance to tetracycline, doxycycline, cefixime, ceftazidime, and nalidixic acid, was selected for genomic analysis.

The bacterial strain was grown overnight in nutrient broth (HiMedia, India) at 37°C, and its genomic DNA was isolated using the Qiagen DNA Mini Kit (QIAGEN, Hilden, Germany). The extracted DNA was fragmented enzymatically with the NEBNext dsDNA Fragmentase Kit (NEB, MA, USA), and fragments were size-selected using SPRI beads (11)). A sequencing library was constructed with the Nextera DNA Flex Library Preparation Kit (Illumina, San Diego, CA, USA) and sequenced on the Illumina NextSeq2000 platform, generating 2 × 150 paired-end reads. Sequence quality was assessed with FastQC v0.11.7 (12), and the raw reads ($n = 1,382,869$) were processed using Trimmomatic v0.39 (13) to remove low-quality bases and adapters. Genome assembly was carried out with Unicycler v0.4.9 (14), followed by gene annotation using PGAP v3.0 (15).

TABLE 1 Genomic elements of the *Bacillus subtilis* strain Hakim RU_LJ

Elements	Values
Genome size	4,033,419 bp
Genome coverage	50.33×
G + C content	43.5%
Number of contigs	104
Contig L50	12
Contig N50	110,116 bp
Total genes	4,321
Coding sequences	4,293
Coding genes	4,117
RNA genes	28
tRNAs genes	18
rRNAs genes	4
tmRNAs gene	1
ncRNAs genes	5
Pseudo genes	176
Genes assigned to SEED subsystems	4,428
Number of subsystems	326

Additional analyses involved screening for antibiotic resistance genes with CARD v3.2.4 (16) and RGI v6.0.2 (17), identifying virulence factors using VFDB and VFanalyzer v4.0 (18), estimating pathogenicity with PathogenFinder v1.1 (19), determining sequence type via MLST v2.0 (20), detecting CRISPR arrays with CRISPRimmunity (21), and evaluating metabolic capabilities with RAST v2.0 (22). Default settings were applied for all software unless specified otherwise.

The elements of the draft genomes are documented in Table 1. Notably, 17 ARGs, including putative aadK, mph(K), tet(L), acquired resistance genes and aminoglycosides, macrolide, tetracycline antibiotic classes, and 10 virulence genes comprising ClpP, cap8D, cpsJ, LPG_RS03745, hasC, cap8P, cps4I, wbpM, fepC, and galF were predicted. MLST classified the genome as an unknown sequence type, and the PathogenFinder tool specified a pathogenicity index of 0.236. The genome exhibited five CRISPR arrays with six signature genes (cas3, DEDDh, DinG, TnsC, csa3, and WYL) and 19 prophages. RAST analysis revealed 326 subsystems comprising 4,428 genes with 27% coverage (Fig. 1).
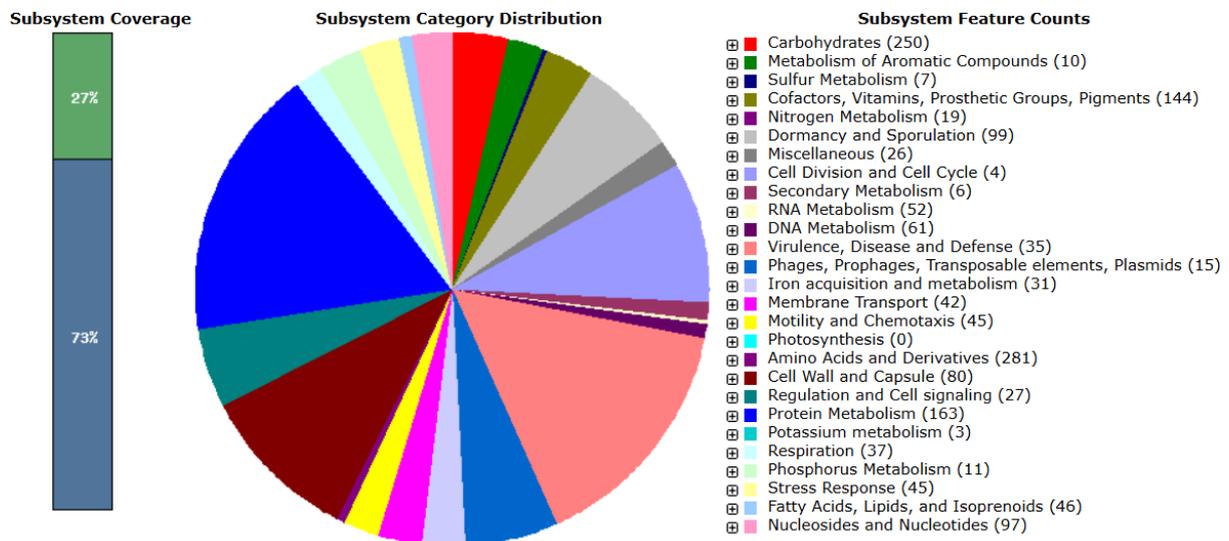

FIG 1 Metabolic functional features in the *Bacillus subtilis* Hakim RU_LJ assembled genome in SEED viewer. The 27% coverage indicates the completeness of functional roles within a specific subsystem across different genomes.

AUTHOR AFFILIATIONS

[1]Department of Veterinary and Animal Sciences, University of Rajshahi, Rajshahi, Rajshahi Division, Bangladesh

[2]Biomedical Sciences & Molecular Biology, College of Medicine and Dentistry, James Cook University, Townsville, Queensland, Australia

[3]Australian Institute of Tropical Health and Medicine, James Cook University, Townsville, Queensland, Australia

AUTHOR ORCIDs

Md. Hakimul Haque http://orcid.org/0000-0002-3856-6478

AUTHOR CONTRIBUTIONS

Muhib Ullah Khan, Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Writing – original draft | Md. Arif-Uz-Zaman Polash, Conceptualization, Data curation, Methodology, Software, Writing – original draft | Nusrat Zahan, Conceptualization, Data curation, Investigation, Methodology | Md. Mottaleb Ali, Software, Writing – review and editing | Nilima Rubaba Azad, Writing – review and editing | Subir Sarker, Writing – review and editing | Md. Hakimul Haque, Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Resources, Supervision, Validation, Writing – original draft, Writing – review and editing

DATA AVAILABILITY

The study on *Bacillus subtilis* strain Hakim RU_LJ, conducted using the WGS shotgun approach, was submitted to NCBI/GenBank, and the assembly was deposited under the accession number JBLYQJ000000000. The pertinent data, including the original readings, were stored with BioProject accession number PRJNA1222092, BioSample accession number SAMN46761393, and SRA accession number SRR32305481. The specific version mentioned in this document is labeled as JBLYQJ000000000.1.

REFERENCES

1. Haque MA, Wang F, Chen Y, Hossen F, Islam MA, Hossain MA, Siddique N, He C, Ahmed F. 2021. *Bacillus* spp. contamination: a novel risk originated from animal feed to human food chains in South-Eastern Bangladesh. Front Microbiol 12:783103. https://doi.org/10.3389/fmicb.2021.783103
2. Alcaraz LD, Moreno-Hagelsieb G, Eguiarte LE, Souza V, Herrera-Estrella L, Olmedo G. 2010. Understanding the evolutionary relationships and major traits of *Bacillus* through comparative genomics. BMC Genomics 11:332. https://doi.org/10.1186/1471-2164-11-332
3. Earl AM, Losick R, Kolter R. 2008. Ecology and genomics of *Bacillus subtilis*. Trends Microbiol 16:269–275. https://doi.org/10.1016/j.tim.2008.03.004
4. Ahmed I, Rabbi MB, Sultana S. 2019. Antibiotic resistance in Bangladesh: a systematic review. Int J Infect Dis 80:54–61. https://doi.org/10.1016/j.ijid.2018.12.017
5. Schallmey M, Singh A, Ward OP. 2004. Developments in the use of *Bacillus* species for industrial production. Can J Microbiol 50:1–17. https://doi.org/10.1139/w03-076
6. Saikat TA, Sayem Khan MA, Islam MS, Tasnim Z, Ahmed S. 2024. Characterization and genome mining of *Bacillus subtilis* BDSA1 isolated from river water in Bangladesh: a promising bacterium with diverse biotechnological applications. Heliyon 10:e34369. https://doi.org/10.1016/j.heliyon.2024.e34369
7. Haque MH, Islam MZ, Miah ML, Das SK, Sarker S. 2023. Molecular characterization of multidrug-resistant *Escherichia coli* isolated from

human urine infections with their antibiogram profile. J Adv Biotechnol Exp Ther 6:172. https://doi.org/10.5455/jabet.2023.d116

8. Bergey DH, Buchanan RE, Gibbons NE. 1974. Bergey's manual of determinative bacteriology. 8th ed, p 966–1097. American Society for Microbiology, Washington, DC, USA.

9. Bauer AW, Kirby WMM, Sherris JC, Turck M. 1966. Antibiotic susceptibility testing by a standardized single disk method. Am J Clin Pathol 45:493–496. https://doi.org/10.1093/acjp/45.4_ts.493

10. M100-S32. 2022 Performance standards for antimicrobial susceptibility testing. Clinical and Laboratory Standards Institute, Wayne, PA, USA.

11. Liu D, Li Q, Luo J, Huang Q, Zhang Y. 2023. An SPRI beads-based DNA purification strategy for flexibility and cost-effectiveness. BMC Genomics 24:125. https://doi.org/10.1186/s12864-023-09211-w

12. Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data. Available from: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

13. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. https://doi.org/10.1093/bioinformatics/btu170

14. 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

15. Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, Lomsadze A, Pruitt KD, Borodovsky M, Ostell J. 2016. NCBI prokaryotic genome annotation pipeline. Nucleic Acids Res 44:6614–6624. https://doi.org/10.1093/nar/gkw569

16. Alcock BP, Raphenya AR, Lau TTY, Tsang KK, Bouchard M, Edalatmand A, Huynh W, Nguyen A-LV, Cheng AA, Liu S, et al. 2020. CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database. Nucleic Acids Res 48:D517–D525. https://doi.org/10.1093/nar/gkz935

17. Florensa AF, Kaas RS, Clausen PTLC, Aytan-Aktug D, Aarestrup FM. 2022. ResFinder - an open online resource for identification of antimicrobial resistance genes in next-generation sequencing data and prediction of phenotypes from genotypes. Microb Genom 8:000748. https://doi.org/10.1099/mgen.0.000748

18. Liu B, Zheng D, Zhou S, Chen L, Yang J. 2022. VFDB 2022: a general classification scheme for bacterial virulence factors. Nucleic Acids Res 50:D912–D917. https://doi.org/10.1093/nar/gkab1107

19. Cosentino S, Voldby Larsen M, Møller Aarestrup F, Lund O. 2013. PathogenFinder--distinguishing friend from foe using bacterial whole genome sequence data. PLoS One 8:e77302. https://doi.org/10.1371/journal.pone.0077302

20. Larsen MV, Cosentino S, Rasmussen S, Friis C, Hasman H, Marvig RL, Jelsbak L, Sicheritz-Pontén T, Ussery DW, Aarestrup FM, Lund O. 2012. Multilocus sequence typing of total-genome-sequenced bacteria. J Clin Microbiol 50:1355–1361. https://doi.org/10.1128/JCM.06094-11

21. Zhou F, Yu X, Gan R, Ren K, Chen C, Ren C, Cui M, Liu Y, Gao Y, Wang S, Yin M, Huang T, Huang Z, Zhang F. 2023. CRISPRimmunity: an interactive web server for CRISPR-associated important molecular events and modulators used in genome editing tool identifying. Nucleic Acids Res 51:W93–W107. https://doi.org/10.1093/nar/gkad425

22. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, et al. 2008. The RAST Server: rapid annotations using subsystems technology. BMC Genomics 9:75. https://doi.org/10.1186/1471-2164-9-75