AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Resource Announcements

| Plant Microbiology | Announcement

Whole-genome sequence of *Paenibacillus amylolyticus* strain W018, isolated from *Triticum aestivum* L. seeds, obtained using nanopore sequencing

Vladimir K. Chebotar,[1] Maria S. Gancheva,[1,2] Elena P. Chizhevskaya,[1] Oksana V. Keleinikova,[1] Maria E. Baganova,[1] Alexander N. Zaplatkin,[1] Kharon A. Husainov[3]

AUTHOR AFFILIATIONS See affiliation list on p. 2.

ABSTRACT In this study, we performed nanopore sequencing of the genome of *Paenibacillus amylolyticus* strain W018, isolated from the seeds of winter wheat, cv. Bezostaya 100. The genome size is 7.07 Mb, with a GC content of 45.8%, and contains 8,190 genes.

KEYWORDS endophytes, wheat seeds, *Paenibacillus amylolyticus*, whole-genome sequence

Several species from the genus *Paenibacillus* are known to be involved in atmospheric nitrogen fixation, solubilization of phosphate and potassium, and the production of phytohormones and antibiotics [reviewed in reference (1)]. Here, we report the complete genome sequence of *Paenibacillus amylolyticus* strain W018, isolated from the seeds of winter wheat (*Triticum aestivum* L.), cv. Bezostaya 100, grown on typical medium-sized low-humus chernozem soil in the mountain and forest zone, Vedeno region, Chechen Republic, Russia. Seeds were disinfected with 70% ethanol for 30 s, rinsed with sterile water two times, then sterilized with bleach solution (15% sodium hypochlorite) for 30 min with shaking on a rotary shaker (PSU 20i, Biosan, Latvia) at 180 rpm, washed seven times with sterile water, and placed on GMF medium (g/L: beef enzymatic hydrolysate—15.0; NaCl—9.0; agar—15.0, LLC NICF, Saint-Petersburg, Russia) plates at 28°C for 48 h. If no bacteria or fungi growth was observed, seeds in 5 mL of sterile water were crushed with a mortar and pestle under sterile conditions. Aliquots of 40 mL of the resulting suspension were plated onto GMF medium. Plates were incubated at 28°C for 2 days. Bacterial isolation was performed as previously described (2). Strain W018 was grown in Luria-Bertani (LB) medium at 30°C overnight. DNA was extracted from a single colony with the cetyltrimethylammonium bromide-NaCl method (3).

The library was prepared with a ligation sequencing kit (catalog number SQK-LSK109) and barcoded using the EXP-NBD104 and EXP-NBD114 kits (Oxford Nanopore, UK) according to the manufacturer's instructions. Sequencing was performed using a MinION device with a 9.4.1 flow cell (Oxford Nanopore, UK) in the Core Centrum "Genomic Technologies, Proteomics and Cell Biology" at the All-Russia Research Institute for Agricultural Microbiology. Base calling was performed with Guppy v3.3.0 (4) in the high-accuracy mode, and read data was quality assessed with FastQC v0.11.9 (5). Oxford Nanopore Technologies generated a total of 13,081 reads, accounting for 139,549,745 bases with a mean length of 10,668 bp and 16.5× coverage. The genome coverage was counted by SAMtools (6), and the assembly was computed with Flye v2.9 (7). The resulted contigs were polished using Racon v. 1.3.2 (8) (four iterations with parameters -m 8 -x -6 -g -8 -w 500) in combination with Medaka v. 1.4.3 (https://github.com/nanoporetech/medaka). The software default parameters were used except where

Editor David A. Baltrus, University of Arizona, Tucson, Arizona, USA

Address correspondence to Vladimir K. Chebotar, vladchebotar@rambler.ru.

The authors declare no conflict of interest.

Received 1 August 2023
Accepted 22 August 2023
Published 25 September 2023

otherwise noted. The assembly was circularized upon submission to the NCBI database. The quality of the genome assembly was assessed using QUAST v 5.1.0 (9). The genome sequence of *P. amylolyticus* W018 was assembled into a single circular chromosomal contig of 7,070,395 bp with an average GC content of 45.79%. The assembly was annotated using Prokaryotic Genome Annotation Pipeline (PGAP) v6.5 (10). The PGAP annotation identified 5,616 coding DNA sequences and 142 RNA sequences in the assembly (102 tRNAs, 36 complete rRNAs, and 4 noncoding RNAs). Notably, PGAP reports that the number of pseudogenes is about 30% of the gene features (2,432 pseudogenes out of 8,048 coding DNA sequences). These are probably caused by indels, which are typical for nanopore sequencing-only assemblies (11).

Genome similarity metrics (ANIb, FastANI, orthoANIb, and orthoANIu) were calculated using the gcType portal (12). The results of the genome analysis showed that the genome of strain W018 had high similarity with GCA_004001025.1 *P. amylolyticus*. Biosynthetic gene clusters for antimicrobial secondary metabolites were identified using antiSMASH 7.0 (13). The genome of strain W018 was predicted to encode antibiotics (paenilarvins A, B, and C; polymyxin; zwittermicin A), opine-type metallophores (bacillopaline), and lasso peptides (paeninodin).

ACKNOWLEDGMENTS

This work was funded by the Ministry of Science and Higher Education of the Russian Federation in accordance with agreement no. 075-15-2021-1055, 28 September 2021, on providing a grant in the form of subsidies from the Federal budget of the Russian Federation. The grant was provided for the implementation of the project Mobilization of the Genetic Resources of Microorganisms on the Basis of the Russian Collection of Agricultural Microorganisms (RCAM) at the All-Russia Research Institute for Agricultural Microbiology (ARRIAM) according to the Network Principle of Organization.

AUTHOR AFFILIATIONS

[1]Laboratory of Microbial Technology, All-Russian Research Institute for Agricultural Microbiology, St. Petersburg, Pushkin, Russia
[2]Department of Genetics and Biotechnology, Faculty of Biology, Saint Petersburg State University, St. Petersburg, Russia
[3]Laboratory of Generic and Selection of Microorganisms, Chechen Research Institute of Agriculture, Chechen Republic, Russia

AUTHOR ORCIDs

Vladimir K. Chebotar http://orcid.org/0000-0001-9762-989X
Maria S. Gancheva http://orcid.org/0000-0002-9631-6143

AUTHOR CONTRIBUTIONS

Vladimir K. Chebotar, Conceptualization, Funding acquisition, Project administration, Resources, Supervision, Writing – review and editing | Maria S. Gancheva, Methodology, Software, Validation, Visualization, Writing – original draft | Elena P. Chizhevskaya, Data curation, Investigation, Methodology | Oksana V. Keleinikova, Formal analysis, Resources | Maria E. Baganova, Investigation | Alexander N. Zaplatkin, Data curation, Formal analysis | Kharon A. Husainov, Resources

DATA AVAILABILITY

This project has been deposited at GenBank under accession no. PRJNA991090. The raw sequencing reads were deposited in the Sequence Read Archive under accession no. SRR25131389. The draft genome sequence accession number is CP130152.

REFERENCES

1. Patowary R, Deka H. 2020. Paenibacillus. Beneficial Microbes in Agro-Ecology:339–361. https://doi.org/10.1016/b978-0-12-823414-3.00017-4
2. Chebotar VK, Gancheva MS, Chizhevskaya EP, Keleinikova OV, Baganova ME, Zaplatkin AN, Pishchik VN, BaltrusDA. 2022. Draft genome sequence of *Bacillus vallismortis* strain BL01, isolated from *Artemisia lerchiana* Web. Microbiol Resour Announc 11:e0064722. https://doi.org/10.1128/mra.00647-22
3. Wilson K. 2001. Preparation of genomic DNA from bacteria. Curr Protoc Mol Biol. https://doi.org/10.1002/0471142727.mb0204s56
4. Wick RR, Judd LM, Holt KE. 2019. Performance of neural network basecalling tools for Oxford Nanopore sequencing. Genome Biol 20:129. https://doi.org/10.1186/s13059-019-1727-y
5. Andrews S. 2010. Fastqc: a quality control tool for high throughput sequence data. Available from: http://www.bioinformatics.babraham.ac.uk/projects/fastqc
6. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The sequence alignment/map format and Samtools. Bioinformatics 25:2078–2079. https://doi.org/10.1093/bioinformatics/btp352
7. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. 2019. Assembly of long, error-prone reads using repeat graphs. Nat Biotechnol 37:540–546. https://doi.org/10.1038/s41587-019-0072-8
8. Vaser R, Sović I, Nagarajan N, Šikić M. 2017. Fast and accurate de novo genome assembly from long uncorrected reads. Genome Res 27:737–746. https://doi.org/10.1101/gr.214270.116
9. 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
10. 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
11. Leidenfrost RM, Wappler N, Wünschiers R. 2020. Draft genome assembly of Rhodobacter sphaeroides 2.4.1 substrain H2 from nanopore data. Microbiol Resour Announc 9:e00414-20. https://doi.org/10.1128/MRA.00414-20
12. Shi W, Sun Q, Fan G, Hideaki S, Moriya O, Itoh T, Zhou Y, Cai M, Kim SG, Lee JS, Sedlacek I, Arahal DR, Lucena T, Kawasaki H, Evtushenko L, Weir BS, Alexander S, Dénes D, Tanasupawat S, Eurwilaichitr L, Ingsriswang S, Gomez-Gil B, Hazbón MH, Riojas MA, Suwannachart C, Yao S, Vandamme P, Peng F, Chen Z, Liu D, Sun X, Zhang X, Zhou Y, Meng Z, Wu L, Ma J. 2021. gcType: a high-quality type strain genome database for microbial phylogenetic and functional research. Nucleic Acids Res 49:D694–D705. https://doi.org/10.1093/nar/gkaa957
13. Blin K, Shaw S, Augustijn HE, Reitz ZL, Biermann F, Alanjary M, Fetter A, Terlouw BR, Metcalf WW, Helfrich EJN, van Wezel GP, Medema MH, Weber T. 2023. antiSMASH 7.0: new and improved predictions for detection, regulation, chemical structures and visualisation. Nucleic Acids Res 51:W46–W50. https://doi.org/10.1093/nar/gkad344