



# Complete Genome Sequences of the Human Pathogen *Paenibacillus thiaminolyticus* Mbale and Type Strain *P. thiaminolyticus* NRRL B-4156

Christine Hehnly,<sup>a,b</sup> Lijun Zhang,<sup>a,b</sup> Joseph N. Paulson,<sup>c</sup> Mathieu Almeida,<sup>d</sup>  Benjamin von Bredow,<sup>e</sup> Dona S. S. Wijetunge,<sup>e</sup>  Michael Y. Galperin,<sup>f</sup> Kathryn Sheldon,<sup>a,b</sup> Steven J. Schiff,<sup>g,h,i,j</sup>  James R. Broach<sup>a,b</sup>

<sup>a</sup>Penn State Institute for Personalized Medicine, The Pennsylvania State University College of Medicine, Hershey, Pennsylvania, USA

<sup>b</sup>Department of Biochemistry and Molecular Biology, The Pennsylvania State University College of Medicine, Hershey, Pennsylvania, USA

<sup>c</sup>Department of Biostatistics, Product Development, Genentech, Inc., South San Francisco, California, USA

<sup>d</sup>MetaGénoPolis, INRA, Université Paris-Saclay, Jouy-en-Josas, France

<sup>e</sup>Department of Pathology, The Pennsylvania State University College of Medicine, Hershey, Pennsylvania, USA

<sup>f</sup>National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, Maryland, USA

<sup>g</sup>Department of Neurosurgery, The Pennsylvania State University College of Medicine, Hershey, Pennsylvania, USA

<sup>h</sup>Penn State Center for Neural Engineering, The Pennsylvania State University College of Engineering, University Park, Pennsylvania, USA

<sup>i</sup>Department of Engineering Science and Mechanics, The Pennsylvania State University College of Engineering, University Park, Pennsylvania, USA

<sup>j</sup>Department of Physics, The Pennsylvania State University Eberly College of Science, University Park, Pennsylvania, USA

**ABSTRACT** We have isolated a likely bacterial pathogen from cerebrospinal fluid from a Ugandan infant suffering from hydrocephalus. Whole-genome sequencing and assembly of the genome of the clinical isolate, as well as that of a previously deposited reference strain, identified the isolate as *Paenibacillus thiaminolyticus*, which has not been associated with widespread human infections.

The *Paenibacillus* genus has been studied widely for its industrial and agricultural associations (1) and as the pathogen causing American foulbrood disease among honey bee larvae (2). Although they cause occasional opportunistic infections in humans (3), *Paenibacillus* species have not been associated with widespread human infections. However, our recent report associated a *Paenibacillus* sp. with postinfectious hydrocephalus in infants in sub-Saharan Africa (J. N. Paulson, B. L. Williams, C. Hehnly, N. Mishra, S. Sinnar, L. Zhang, P. Ssentongo, E. Mbabazi-Kabachelor, D. S. S. Wijetunge, B. von Bredow, R. Mulondo, J. Kiwanuka, F. Bajunirwe, J. Bazira, L. Bebell, K. Burgoine, M. Couto-Rodriguez, J. E. Ericson, T. Erickson, M. Ferrari, M. Gladstone, C. Guo, M. Haran, M. Hornig, A. M. Isaacs, B. N. Kaaya, A. V. KulKarni, E. Kumbakumba, X. Li, D. D. Limbrick, Jr, J. Magombe, S. U. Morton, J. Mugamba, J. Ng, P. Olupot-Olupot, J. Onen, F. Roy, K. Sheldon, R. Townsend, A. D. Weeks, A. J. Whalen, J. Quackenbush, P. Ssenyonga, M. Y. Galperin, M. Almeida, H. Atkins, B. C. Warf, W. I. Lipkin, J. Broach, and S. J. Schiff, unpublished data). Presented here is the clinical isolate *Paenibacillus thiaminolyticus* Mbale, which was recovered from a patient in Mbale, Uganda. We assembled a draft genome using optical mapping and error-corrected long-read sequencing of the clinical isolate and the type strain *P. thiaminolyticus* NRRL B-4156, which was isolated in 1950 from a thiamine-deficient patient (4).

*Paenibacillus thiaminolyticus* was recovered from previously fresh-frozen cerebrospinal fluid, from an infant in Uganda with postinfectious hydrocephalus, by growth in a lytic anaerobic blood culture bottle (Bactec, BD) at 37°C for 14 days. High-molecular-weight DNA was prepared according to the blood and cell culture DNA isolation protocol with lysozyme digestion (Bionano Genomics, CA, USA).

A PCR-free short-read sequencing library for the clinical isolate was prepared from

**Citation** Hehnly C, Zhang L, Paulson JN, Almeida M, von Bredow B, Wijetunge DSS, Galperin MY, Sheldon K, Schiff SJ, Broach JR. 2020. Complete genome sequences of the human pathogen *Paenibacillus thiaminolyticus* Mbale and type strain *P. thiaminolyticus* NRRL B-4156. Microbiol Resour Announc 9:e00181-20. <https://doi.org/10.1128/MRA.00181-20>.

**Editor** John J. Dennehy, Queens College

**Copyright** © 2020 Hehnly et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to James R. Broach, [jbroach@pennstatehealth.psu.edu](mailto:jbroach@pennstatehealth.psu.edu).

**Received** 20 February 2020

**Accepted** 16 March 2020

**Published** 9 April 2020

1  $\mu$ g of DNA fragmented with an E220 focused ultrasonicator (Covaris), with an average insert size of 400 bp, using a HyperPrep kit (KAPA Biosystems). The final library (10 pM) was sequenced to 20 million reads using MiSeq 600-cycle v3 reagents (Illumina). Previously reported short-read sequencing data were used for the type strain NRRL B-4156 (5).

Long-read sequencing libraries for the type strain and the clinical isolate were prepared using a HyperPrep kit and 1D sequencing kit and were sequenced on the MinION platform with an SQK-LSK108 SpotON flow cell (Oxford Nanopore Technologies, UK). Base calling and trimming using Albacore v2.3.1 (<https://github.com/dvera/albacore>) and Porechop v0.2.3 (<https://github.com/rwick/Porechop>) resulted in 130,472 and 559,846 reads for the type strain and clinical isolate, respectively. We assembled long-read genomes using Canu v1.7 (6) and then corrected them with short-read sequences using Pilon v1.22 (7) and Bowtie 2 v2.3.4.3 (8). All software programs were run using default settings. We obtained 6 contigs, with an  $N_{50}$  of 2.26 Mbp, for the clinical isolate and 1 continuous 6,612,442-bp contig for the type strain. The type strain assembly is likely the entire chromosome.

Optical mapping of the clinical isolate DNA on a Bionano Genomics Saphyr system, following direct labeling and staining protocol revision A (Bionano Genomics), resulted in a continuous 7.6-Mbp linear chromosome with overlapping ends, with 70 $\times$  effective coverage. Three fragmented contigs from the corrected long-read assembly were aligned to the optical map, yielding a 6,744,575-bp hybrid scaffold, using Bionano Access v1.3.0, Bionano Tools v1.3.8041.8044, and Hybrid Scaffold v10252018.

Three phage insert regions were found in the type strain genome and 11 phage regions were found in the clinical isolate genome using PHASTER (9). Two of the 11 regions were identified on 99,275-bp and 35,613-bp contigs, which identified them as likely extrachromosomal phages of unknown origin. The remaining 52,271-bp contig had 37,930 bp that mapped back to the hybrid scaffold and another 14,393 bp, both with 99.8% similarity. From optical mapping, a 40-kbp insertion lies at position 4098996 in the hybrid scaffold, which correlates with the remaining contig. Further investigations would verify whether this is a technical or biological event.

The 16S rRNA genes of the two genomes had 99.2 to 99.4% sequence similarity, with a genomic average nucleotide identity (ANI) of 97.06%, as determined by the ANI calculator (10). Phylogenetic analysis placed the clinical isolate between two *P. thiaminolyticus* species, using 40 conserved markers that were extracted and aligned from the annotated genome (11) using fetchMG v1.0 (12) and MUSCLE v3.8.425 (13).

**Data availability.** The whole-genome assemblies, shotgun short reads, and MinION reads were deposited in GenBank (accession numbers [PRJNA552221](#), [PRJNA552222](#), [SRR11015844](#), and [SRR11031176](#)).

## ACKNOWLEDGMENTS

We thank David Craft, Penn State College of Medicine, for his assistance with microbiological culturing and testing and Nathan Olson, National Institute of Standards and Technology, for his assistance with sequence analysis.

This work was supported by an NIH Director's Pioneer Award (grant 1DP1HD086071) to S.J.S.

## REFERENCES

- Grady EN, MacDonald J, Liu L, Richman A, Yuan ZC. 2016. Current knowledge and perspectives of *Paenibacillus*: a review. Microb Cell Fact 15:203. <https://doi.org/10.1186/s12934-016-0603-7>.
- Genersch E. 2010. American foulbrood in honeybees and its causative agent, *Paenibacillus* larvae. J Invertebr Pathol 103(Suppl 1):S10–S19. <https://doi.org/10.1016/j.jip.2009.06.015>.
- Sáez-Nieto JA, Medina-Pascual MJ, Carrasco G, Garrido N, Fernandez-Torres MA, Villalón P, Valdezate S. 2017. *Paenibacillus* spp. isolated from human and environmental samples in Spain: detection of 11 new species. New Microbes New Infect 19:19–27. <https://doi.org/10.1016/j.nmni.2017.05.006>.
- Fujita A. 1954. Thiaminase. Adv Enzymol Relat Subj Biochem 15: 389–421. <https://doi.org/10.1002/9780470122600.ch9>.
- Sannino D, Angert ER. 2017. Genomic insights into the thiamin metabolism of *Paenibacillus thiaminolyticus* NRRL B-4156 and *P. apiarius* NRRL B-23460. Stand Genomic Sci 12:59. <https://doi.org/10.1186/s40793-017-0276-9>.
- Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. 2017.

- Canu: scalable and accurate long-read assembly via adaptive *k*-mer weighting and repeat separation. *Genome Res* 27:722–736. <https://doi.org/10.1101/gr.215087.116>.
7. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, Earl AM. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 9:e112963. <https://doi.org/10.1371/journal.pone.0112963>.
  8. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9:357–359. <https://doi.org/10.1038/nmeth.1923>.
  9. Arndt D, Grant JR, Marcu A, Sajed T, Pon A, Liang Y, Wishart DS. 2016. PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Res* 44:W16–W21. <https://doi.org/10.1093/nar/gkw387>.
  10. Rodriguez-R LM, Konstantinidis KT. 2016. The enveomics collection: a toolbox for specialized analyses of microbial genomes and metagenomes. *PeerJ Prepr* 4:e1900v1. <https://doi.org/10.7287/peerj.preprints.1900v1>.
  11. Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11:119. <https://doi.org/10.1186/1471-2105-11-119>.
  12. Kultima JR, Sunagawa S, Li J, Chen W, Chen H, Mende DR, Arumugam M, Pan Q, Liu B, Qin J, Wang J, Bork P. 2012. MOCAT: a metagenomics assembly and gene prediction toolkit. *PLoS One* 7:e47656. <https://doi.org/10.1371/journal.pone.0047656>.
  13. Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32:1792–1797. <https://doi.org/10.1093/nar/gkh340>.