



# Whole-Genome Sequence of *Paenibacillus marchantiae* Isolated from the Liverwort *Marchantia polymorpha* subsp. *ruderalis* Ecotype BoGa

 Anja Meierhenrich,<sup>a,b</sup>  Bianca Frommer,<sup>a,b</sup>  Wiebke Halpape,<sup>a,b</sup>  Marvin Hildebrandt,<sup>a</sup>  Isabell E. Bleile,<sup>c</sup>  Judith Helmig,<sup>d</sup>  
 Sabine Zachgo,<sup>d</sup>  Andrea Bräutigam,<sup>a,b</sup>  Bart Verwaaijen<sup>a,b,e</sup>

<sup>a</sup>Computational Biology, Faculty of Biology, Bielefeld University, Bielefeld, Germany

<sup>b</sup>Computational Biology, Centre of Biotechnology (CeBiTec), Bielefeld University, Bielefeld, Germany

<sup>c</sup>Molecular Biotechnology, Bielefeld University, Bielefeld, Germany

<sup>d</sup>Botany, School of Biology and Chemistry, Osnabrück University, Osnabrück, Germany

<sup>e</sup>Department of Genetics, Martin-Luther-University-Halle-Wittenberg, Halle (Saale), Germany

**ABSTRACT** The bacterium *Paenibacillus marchantiae* was isolated from male plants of the liverwort *Marchantia polymorpha* subsp. *ruderalis* ecotype BoGa. Here, we report on the complete genome sequence generated from long Nanopore reads. The genome sequence comprises 6,983,959 bp with a GC content of 46.02% and 6,195 predicted protein-coding genes.

Species belonging to the genus *Paenibacillus* are rod-shaped and Gram-positive, Gram-variable, or Gram-negative bacteria with an aerobic or facultative anaerobic lifestyle (1, 2). Until today, 285 species are known (3). Back in 1993, the genus was formed from a subgroup of bacilli using 16S rRNA (2). Representatives of *Paenibacillus* have been found in various places, such as in animals (4), rhizospheres (5), leaves (6), and soil (7). Some are known to support plant growth, for example by nitrogen fixation (8), enabling iron uptake by siderophores (9) or through their antimicrobial resistance (10).

Male *Marchantia polymorpha* subsp. *ruderalis* ecotype BoGa plants (11) were grown in petri dishes on half-strength Gamborg's medium (Gamborg B5; Duchefa Biochemie B.V., Netherlands) at room temperature under 16-h/8-h day-night conditions. The cetyltrimethylammonium bromide (CTAB) method was applied for DNA extraction using whole plants (12). DNA quality was checked with the Invitrogen Qubit 4 fluorometer (Thermo Fisher Scientific Inc., USA). To prepare DNA for sequencing, the short read eliminator kit (PacBio, USA) and the ligation sequencing genomic DNA (gDNA) kit (SQK-LSK109-XL; Oxford Nanopore Technologies [ONT], Oxford, UK) were used. For sequencing, one R9.4.1 and one R10.0 flow cell were run on a GridION platform, and base calling was performed with the high-accuracy model (MinKNOW v1.4.3; all from ONT). All programs were run with default parameters unless otherwise specified. Genomic reads were checked for contaminations with BLASTN searches (13) against the NCBI nucleotide collection database, and reads matching *Paenibacillus* genome assemblies were filtered (NCBI; Organism "Paenibacillus"; Database "Assembly"; BLAST 2.8.1+; E value < 0.001) (<https://www.ncbi.nlm.nih.gov/assembly>). *Paenibacillus marchantiae* genome assembly was performed with Canu (v2.2) (14) assuming a genome size of 7.0 Mbp. Racon (v1.4.20) (15), Minimap2 (v2.22-r1101; parameter "--ax map-ont") (16), and Medaka (v1.4.3; parameter "-m r941\_min\_high\_g360"; ONT) were used for polishing, and Berokka (v0.2.3) (<https://github.com/tseemann/berokka>) was run for overlap trimming. The assembly resulted in one circular contig (6,983,959 bp; GC content of 46.02%) (Table 1). Benchmarking Universal Single-Copy Orthologs (BUSCO) (v5.4.3; database "bacillales\_odb10") (17) and CheckM (v1.2.2) (18) were applied to check assembly quality resulting in 0.097% contamination and 99.84% assembly completeness. A total of 6,195 protein-coding genes

**Editor** Simon Roux, DOE Joint Genome Institute

**Copyright** © 2023 Meierhenrich 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 Anja Meierhenrich, [anja.meierhenrich@uni-bielefeld.de](mailto:anja.meierhenrich@uni-bielefeld.de).

The authors declare no conflict of interest.

**Received** 10 May 2023

**Accepted** 9 June 2023

**Published** 22 June 2023

**TABLE 1** Sequencing and assembly statistics for *Paenibacillus marchantiae*

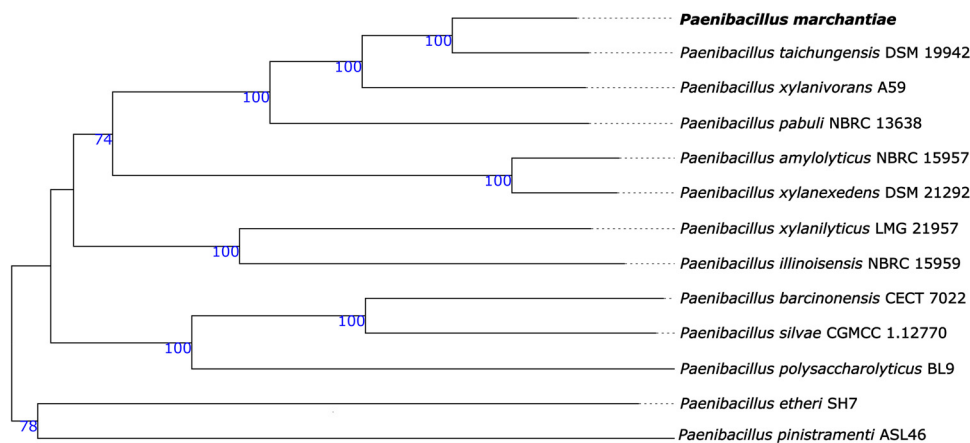
Parameter	Data
Raw genomic sequencing reads	
No. of reads	404,722
Total length (bp)	3,971,654,859
$N_{50}$ (bp)	16,110
Genome sequence	
No. of sequences	1
Length (bp)	6,983,959
GC content (%)	46.02
Genome coverage (×)	236.73
Gene annotation	
Total no. of genes	6,331
No. of protein-coding genes	6,195
No. of rRNAs	33
No. of tRNAs	102
No. of tmRNAs <sup>a</sup>	1
BUSCO results (%)	
Complete	98.4
Single copy	98.2
Duplicated	0.2
Fragmented	1.1
Missing	0.5

<sup>a</sup> tmRNAs, transfer-messenger RNAs.

were predicted with Prokka (v1.14.5) (19). Relevant statistics for the assembly and raw reads are listed in Table 1. Metabolic pathways were predicted with the KEGG Automatic Annotation Server (KAAS) (20), GO terms were calculated with eggno-mapper (v2) (21), and potential secondary metabolite biosynthesis gene clusters were identified with antiSMASH (v6.1.1) (22).

Phylogenetic relatedness to other strains was determined using the assembly and Type (Strain) Genome Server (TYGS) (23) (Fig. 1). The closest species were *Paenibacillus taichungensis* DSM 19942 (GenBank accession number [GCA\\_013359905.1](#)) and *Paenibacillus xylanivorans* A59 (GenBank accession number [GCA\\_001280595.1](#)), which were both isolated from soil (24, 25).

The KAAS analysis indicates that *Paenibacillus marchantiae* is unable to produce galactose and sucrose but can take up and degrade them. The gene cluster of paenibactin was identified and comprises all genes for bacillibactin synthesis. These siderophores are responsible for the extracellular conversion of  $Fe^{2+}$  and  $Fe^{3+}$  and thus enable the provision of iron for plants (9). The metabolic capabilities indicate a potential beneficial partnership between *Paenibacillus marchantiae* and BoGa.



**FIG 1** A phylogenetic tree was generated for *Paenibacillus marchantiae* from the genome sequence with the Type (Strain) Genome Server (TYGS) (22). The lengths of the branches are based on the determination of the GBDP distance formula d5. Above the branches are the GBDP pseudobootstrap support values of >60% from 100 replicates. The mean branch support is 89.9% (22). The phylogenetic tree was aligned at the midpoint. The *Paenibacillus marchantiae* presented here is highlighted in bold type.

**Data availability.** The genome assembly of *Paenibacillus marchantiae* has been deposited under the NCBI GenBank accession number [CP118270.1](https://www.ncbi.nlm.nih.gov/nuclseq/CP118270.1). The SRA, BioProject, and BioSample accession numbers are [SRP421713](https://www.ncbi.nlm.nih.gov/sra/SRP421713), [PRJNA933122](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA933122), and [SAMN33224160](https://www.ncbi.nlm.nih.gov/biosample/SAMN33224160), respectively.

## ACKNOWLEDGMENTS

We thank Jörn Kalinowski and Tobias Busche for providing the flow cells and the sequencing platform.

We acknowledge support for the publication costs by the Open Access Publication Fund of Bielefeld University and the Deutsche Forschungsgemeinschaft (DFG). This work was supported by the BMBF-funded de. NBI Cloud within the German Network for Bioinformatics Infrastructure (de.NBI).

## REFERENCES

- Shida O, Takagi H, Kadowaki K, Nakamura LK, Komagata K. 1997. Transfer of *Bacillus alginolyticus*, *Bacillus chondroitinus*, *Bacillus curdlanolyticus*, *Bacillus glucanolyticus*, *Bacillus kobensis*, and *Bacillus thiaminolyticus* to the genus *Paenibacillus* and emended description of the genus *Paenibacillus*. *Int J Syst Bacteriol* 47:289–298. <https://doi.org/10.1099/00207713-47-2-289>.
- 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>.
- List of Prokaryotic Names with Standing in Nomenclature. 2022. Genus *Paenibacillus*. <https://lpsn.dsmz.de/genus/paenibacillus>.
- Horn MA, Ihssen J, Matthies C, Schramm A, Acker G, Drake HL. 2005. *Dechloromonas denitrificans* sp. nov., *Flavobacterium denitrificans* sp. nov., *Paenibacillus anaericanus* sp. nov. and *Paenibacillus terrae* strain MH72, N<sub>2</sub>O-producing bacteria isolated from the gut of the earthworm *Aporrectodea caliginosa*. *Int J Syst Evol Microbiol* 55:1255–1265. <https://doi.org/10.1099/ijs.0.63484-0>.
- Validov S, Kamilova F, Qi S, Stephan D, Wang JJ, Makarova N, Lugtenberg B. 2007. Selection of bacteria able to control *Fusarium oxysporum* f. sp. *radicislycopersici* in stonewool substrate. *J Appl Microbiol* 102:461–471. <https://doi.org/10.1111/j.1365-2672.2006.03083.x>.
- Madhaiyan M, Poonguzhali S, Saravanan VS, Duraipandiyar V, Al-Dhabi NA, Kwon S-W, Whitman WB. 2017. *Paenibacillus polysaccharolyticus* sp. nov., a xylanolytic and cellulolytic bacteria isolated from leaves of Bamboo *Phyllostachys aureosulcata*. *Int J Syst Evol Microbiol* 67:2127–2133. <https://doi.org/10.1099/ijsem.0.001901>.
- Shagal CC, Krishnamoorthy R, Kim K, Sundaram S, Sa T. 2014. Arsenic-tolerant plant-growth-promoting bacteria isolated from arsenic-polluted soils in South Korea. *Environ Sci Pollut Res Int* 21:9356–9365. <https://doi.org/10.1007/s11356-014-2852-5>.
- Xie J-B, Du Z, Bai L, Tian C, Zhang Y, Xie J-Y, Wang T, Liu X, Chen X, Cheng Q, Chen S, Li J. 2014. Comparative genomic analysis of N<sub>2</sub>-fixing and non-N<sub>2</sub>-fixing *Paenibacillus* spp.: organization, evolution and expression of the nitrogen fixation genes. *PLoS Genet* 10:e1004231. <https://doi.org/10.1371/journal.pgen.1004231>.
- Raza W, Shen Q. 2010. Growth, Fe<sup>3+</sup> reductase activity, and siderophore production by *Paenibacillus polymyxa* SQR-21 under differential iron conditions. *Curr Microbiol* 61:390–395. <https://doi.org/10.1007/s00284-010-9624-3>.
- Lorentz RH, Artico S, Da Silveira AB, Einsfeld A, Corção G. 2006. Evaluation of antimicrobial activity in *Paenibacillus* spp. strains isolated from natural environment. *Lett Appl Microbiol* 43:541–547. <https://doi.org/10.1111/j.1472-765X.2006.01995.x>.
- Busch A, Deckena M, Almeida-Trapp M, Kopischke S, Kock C, Schüssler E, Tsiantis M, Mithöfer A, Zachgo S. 2019. MpTCP1 controls cell proliferation and redox processes in *Marchantia polymorpha*. *New Phytol* 224:1627–1641. <https://doi.org/10.1111/nph.16132>.
- Pucker B. 2020. Plant DNA extraction and preparation for ONT sequencing. <https://doi.org/10.17504/protocols.io/bcviyw7w>.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol* 215:403–410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2).
- 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>.
- Koren S, 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>.
- Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics (Oxford, England)* 34:3094–3100. <https://doi.org/10.1093/bioinformatics/bty191>.
- Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EM. 2021. BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol Biol Evol* 38:4647–4654. <https://doi.org/10.1093/molbev/msab199>.
- Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* 25:1043–1055. <https://doi.org/10.1101/gr.186072.114>.
- Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics (Oxford, England)* 30:2068–2069. <https://doi.org/10.1093/bioinformatics/btu153>.
- Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M. 2007. KAA3: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res* 35: W182–W185. <https://doi.org/10.1093/nar/gkm321>.
- Cantalapiedra CP, Hernández-Plaza A, Letunic I, Bork P, Huerta-Cepas J. 2021. eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. *Mol Biol Evol* 38: 5825–5829. <https://doi.org/10.1093/molbev/msab293>.
- Blin K, Shaw S, Kloosterman AM, Charlop-Powers Z, van Wezel GP, Medema MH, Weber T. 2021. antiSMASH 6.0: improving cluster detection and comparison capabilities. *Nucleic Acids Res* 49:W29–W35. <https://doi.org/10.1093/nar/gkab335>.
- Meier-Kolthoff JP, Göker M. 2019. TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy. *Nat Commun* 10: 2182. <https://doi.org/10.1038/s41467-019-10210-3>.
- Lee F-L, Tien C-J, Tai C-J, Wang L-T, Liu Y-C, Chen L-L. 2008. *Paenibacillus tai-chungensis* sp. nov., from soil in Taiwan. *Int J Syst Evol Microbiol* 58:2640–2645. <https://doi.org/10.1099/ijs.0.65776-0>.
- Ghio S, Sauka DH, Ferrari AE, Piccini FE, Ontañón OM, Campos E. 2019. *Paenibacillus xylanivorans* sp. nov., a xylan-degrading bacterium isolated from decaying forest soil. *Int J Syst Evol Microbiol* 69:3818–3823. <https://doi.org/10.1099/ijsem.0.003686>.