




Complete Genome Sequences of 13 *Bacillus subtilis* Soil Isolates for Studying Secondary Metabolite Diversity

 Heiko T. Kiesewalter,^a  Carlos N. Lozano-Andrade,^a Gergely Maróti,^b Dan Snyder,^c  Vaughn S. Cooper,^{c,d} Tue Sparholt Jørgensen,^e Tilmann Weber,^e  Ákos T. Kovács^a

^aBacterial Interactions and Evolution Group, DTU Bioengineering, Technical University of Denmark, Kongens Lyngby, Denmark

^bInstitute of Plant Biology, Biological Research Center of the Hungarian Academy of Sciences, Szeged, Hungary

^cMicrobial Genome Sequencing Center, Pittsburgh, Pennsylvania, USA

^dDepartment of Microbiology and Molecular Genetics, University of Pittsburgh, Pittsburgh, Pennsylvania, USA

^eThe Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Kongens Lyngby, Denmark

ABSTRACT *Bacillus subtilis* is a plant-benefiting soil-dwelling Gram-positive bacterium with secondary metabolite production potential. Here, we report the complete genome sequences of 13 *B. subtilis* strains isolated from different soil samples in Germany and Denmark.

Various species of the *Bacillus* genus have been exploited for biocontrol of crop plants. *Bacillus subtilis* is the most studied bacterium of the bacilli due to its high potential for industrial production of proteins, its utilization as a plant biological, and its easy genetic modification (1). In addition, the biofilm development of *B. subtilis* has been intensely investigated under laboratory settings (2–5) and during colonization of plant root (6–8) and fungal mycelia (9). The biocontrol potential of *B. subtilis* is determined by its ability to produce a variety of secondary metabolites, including surfactin, plipastatin (or fengycin), and bacillaene (10). Here, we performed complete genome sequencing of 13 *B. subtilis* strains in order to facilitate a detailed investigation of genes involved in secondary metabolite production.

B. subtilis strains were isolated from various sampling sites in Germany and Denmark (see details under BioProject accession number [PRJNA587401](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA587401)) by using spore selection and specifically isolating architecturally complex colonies reminiscent of colony biofilm formation of *B. subtilis* (1, 5, 8, 11). Strains 73 and 75 were isolated specifically by inserting a constitutively produced green fluorescent protein (GFP) into the *amyE* gene, as described earlier (11). After biochemical assays, biofilm tests, and chemical analyses of natural products of the isolated strains, 13 *B. subtilis* strains were scrutinized with genome sequencing.

For Illumina sequencing, genomic DNA of the *B. subtilis* strains was isolated with the GeneMatrix bacterial and yeast genomic DNA purification kit according to the manufacturer's instructions (EURx, Gdansk, Poland). Paired-end libraries were prepared for the strains, except MB9_B4, using the NEBNext Ultra II DNA library prep kit for Illumina (catalog number E7645L). Paired-end reads were generated on an Illumina NextSeq sequencer using a TG NextSeq 500/550 high-output kit v. 2 (300 cycles). In the case of MB9_B4, a mate pair library was generated using an Illumina Nextera mate pair kit (catalog number FC-132-1001) with insert sizes ranging from 6 to 15 kb. DNA sequencing was carried out on an Illumina MiSeq machine using V2 sequencing chemistry, resulting in 2 × 250-bp reads.

For Nanopore sequencing, genomic DNA was extracted using Qiagen blood and tissue kits (catalog number 69506), following the manufacturer's protocol, using lysozyme digestion prior to extraction. This extra treatment was performed by resus-

Citation Kiesewalter HT, Lozano-Andrade CN, Maróti G, Snyder D, Cooper VS, Jørgensen TS, Weber T, Kovács ÁT. 2020. Complete genome sequences of 13 *Bacillus subtilis* soil isolates for studying secondary metabolite diversity. *Microbiol Resour Annu* 9:e01406-19. <https://doi.org/10.1128/MRA.01406-19>.

Editor Irene L. G. Newton, Indiana University, Bloomington

Copyright © 2020 Kiesewalter 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 Ákos T. Kovács, atkovacs@dtu.dk.

Received 14 November 2019

Accepted 10 December 2019

Published 9 January 2020

TABLE 1 Accession numbers, assembly metrics, and annotated features of the sequenced *B. subtilis* strains isolated

Strain or plasmid	GenBank accession no.	Avg coverage (×) for:		Genome assembly size (bp)	G+C content (%)	No. of rRNAs	No. of tRNAs	No. of CDS ^a	Topology
		Illumina reads	Nanopore reads						
73	CP045826	595	263	4,166,516	43.7	30	86	4,203	Circular
75	CP045825	651	289	4,156,459	43.9	30	86	4,169	Circular
MB8_B1	CP045823	691	420	4,221,278	43.5	30	86	4,275	Circular
MB8_B7	CP045821	667	320	4,191,568	43.4	30	88	4,362	Circular
Plasmid pBs001	CP045822	519	393	84,033					Circular
MB8_B10	CP045824	602	550	4,225,362	43.5	30	86	4,289	Circular
MB9_B1	CP045820	636	517	4,263,919	43.5	30	86	4,320	Circular
MB9_B4	CP045819	138	480	4,105,407	43.8	30	86	4,108	Circular
MB9_B6	CP045818	671	440	4,087,720	43.8	30	86	4,089	Circular
P5_B1	CP045817	512	395	4,083,248	43.8	30	88	4,053	Circular
P5_B2	CP045816	586	353	4,103,324	43.6	30	86	4,245	Circular
P8_B1	CP045922	585	358	4,215,512	43.4	30	88	4,386	Circular
Plasmid pBs003	CP045923	510	260	84,215					Circular
P8_B3	CP045812	498	319	4,215,511	43.4	30	88	4,386	Circular
Plasmid pBs005	CP045813	386	490	84,215					Circular
P9_B1	CP045811	658	430	4,063,468	43.8	30	86	4,075	Circular

^aCDS, coding sequences.

pending the cell pellet in 200 μ l of 20 mg/ml lysozyme and incubating the samples for 20 minutes at 37°C. Before sequencing on the Nanopore instrument, a ligation sequencing kit (catalog number SQK-LSK109) was used with native barcoding expansion 1-12 (catalog number EXP-NBD104) following the manufacturer's protocol. Libraries were sequenced using an R9.4.1 flow cell and a MinION device running a 48-h sequencing cycle without base calling. The reads were base called and demultiplexed using Guppy v. 3.1.5 on an Amazon Web Service (AWS) GPU instance with quality control, as described before (12).

For *de novo* assembly, Illumina reads were adapter and quality trimmed using AdapterRemoval v. 2.1.7 (13) with the switches `-trimns` and `-trimqualities`. Nanopore reads were adapter and quality trimmed using Porechop v. 0.2.4 (14) and assembled with the Flye assembler v. 2.6 (15) with the switches `-g 5m` and `-plasmids` as suggested in a recent benchmark (16). Then, the Flye assembly graph, the trimmed Nanopore reads, and the trimmed Illumina reads from each sample were used as input for Unicycler v. 0.4.8-beta assembly with the switches `-existing_long_read_assembly` and `-no_correct`. Unicycler builds on several existing tools based around SPAdes assembly v. 3.13.0 (17), Pilon v. 1.22 (18), and SAMtools v. 1.9 (19). Assemblies were evaluated using the graph visualization software Bandage v. 0.8.1 (20) and BUSCO v. 3 (21) with the *Bacillales* ODB9 database to evaluate the core gene content of each genome.

The assembly produced 13 circularized chromosomes comprising 4,063,468 to 4,263,919 bases with a G+C content of 43.4 to 43.9%. Three isolates contained circular plasmids, each 84 kb in size. Automated annotation was performed using the NCBI Prokaryotic Genome Annotation Pipeline (Table 1).

Genes coding for proteins possibly involved in secondary metabolite production were identified using antiSMASH v. 5 (22), which revealed the presence of gene clusters encoding surfactin (*srf*), plipastatin (*pps*), bacillaene (*pks*), and bacillibactin (*dhb*) biosynthesis in all isolates except strain P5_B2, which lacks the majority of the bacillaene (*pks*) biosynthetic gene cluster. Future detailed analysis of these biosynthetic gene clusters will be performed to reveal differences in secondary metabolite profiles.

Data availability. The raw data and assemblies have been deposited in GenBank under the BioProject accession number [PRJNA587401](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA587401). The complete genome sequence accession numbers are listed in Table 1.

ACKNOWLEDGMENT

This project was supported by the Danish National Research Foundation (DNRF137) for the Center for Microbial Secondary Metabolites.

REFERENCES

- Kovács ÁT. 2019. *Bacillus subtilis*. Trends Microbiol 27:724–725. <https://doi.org/10.1016/j.tim.2019.03.008>.
- Branda SS, González-Pastor JE, Ben-Yehuda S, Losick R, Kolter R. 2001. Fruiting body formation by *Bacillus subtilis*. Proc Natl Acad Sci U S A 98:11621–11626. <https://doi.org/10.1073/pnas.191384198>.
- Mhatre E, Monterrosa RG, Kovács ÁT. 2014. From environmental signals to regulators: modulation of biofilm development in Gram-positive bacteria. J Basic Microbiol 54:616–632. <https://doi.org/10.1002/jobm.201400175>.
- Dragoš A, Martin M, Garcia CF, Kricks L, Pausch P, Heimerl T, Bálint B, Maróti G, Bange G, López D, Lieleg O, Kovács ÁT. 2018. Collapse of genetic division of labour and evolution of autonomy in pellicle biofilms. Nat Microbiol 3:1451–1460. <https://doi.org/10.1038/s41564-018-0263-y>.
- Kovács ÁT, Dragoš A. 2019. Evolved biofilm: review on the experimental evolution studies of *Bacillus subtilis* pellicles. J Mol Biol 431:4749–4759. <https://doi.org/10.1016/j.jmb.2019.02.005>.
- Beauregard PB, Chai Y, Vlamakis H, Losick R, Kolter R. 2013. *Bacillus subtilis* biofilm induction by plant polysaccharides. Proc Natl Acad Sci U S A 110:E1621–E1630. <https://doi.org/10.1073/pnas.1218984110>.
- Dragoš A, Kiesewalter H, Martin M, Hsu C-Y, Hartmann R, Wechsler T, Eriksen C, Brix S, Drescher K, Stanley-Wall N, Kümmerli R, Kovács T. 2018. Division of labor during biofilm matrix production. Curr Biol 28:1903–1913. <https://doi.org/10.1016/j.cub.2018.04.046>.
- Gallegos-Monterrosa R, Mhatre E, Kovács ÁT. 2016. Specific *Bacillus subtilis* 168 variants form biofilms on nutrient-rich medium. Microbiology 162:1922–1932. <https://doi.org/10.1099/mic.0.000371>.
- Kjeldgaard B, Listian SA, Ramaswamhi V, Richter A, Kiesewalter H, Kovács ÁT. 2019. Fungal hyphae colonization by *Bacillus subtilis* relies on biofilm matrix components. Biofilm 1:100007. <https://doi.org/10.1016/j.biofilm.2019.100007>.
- Caulier S, Nannan C, Gillis A, Licciardi F, Bragard C, Mahillon J. 2019. Overview of the antimicrobial compounds produced by members of the *Bacillus subtilis* group. Front Microbiol 10:302. <https://doi.org/10.3389/fmicb.2019.00302>.
- van Gestel J, Weissing FJ, Kuipers OP, Kovács ÁT. 2014. Density of founder cells affects spatial pattern formation and cooperation in *Bacillus subtilis* biofilms. ISME J 8:2069–2079. <https://doi.org/10.1038/ismej.2014.52>.
- Lanfear R, Schalamun M, Kainer D, Wang W, Schwessinger B. 2019. MinIONQC: fast and simple quality control for MinION sequencing data. Bioinformatics 35:523–525. <https://doi.org/10.1093/bioinformatics/bty654>.
- Schubert M, Lindgreen S, Orlando L. 2016. AdapterRemoval v2: rapid adapter trimming, identification, and read merging. BMC Res Notes 9:1–7. <https://doi.org/10.1186/s13104-016-1900-2>.
- Wick RR. 2017. Porechop. <https://github.com/rwick/Porechop>.
- 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>.
- Wick RR, Holt KE. Long-read-assembler-comparison: initial release. 2019. <https://doi.org/10.5281/zenodo.2702443>.
- Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Pribelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19:455–477. <https://doi.org/10.1089/cmb.2012.0021>.
- Walker BJ, Abeel T, Shea T, Priest M, Boueiri 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>.
- Li H, Handsaker B, Wysoker A, Fennel 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>.
- Wick RR, Schultz MB, Zobel J, Holt KE. 2015. Bandage: interactive visualization of de novo genome assemblies. Bioinformatics 31:3350–3352. <https://doi.org/10.1093/bioinformatics/btv383>.
- Waterhouse RM, Seppey M, Simao FA, Manni M, Ioannidis P, Kliuchnikov G, Kriventseva EV, Zdobnov EM. 2018. BUSCO applications from quality assessments to gene prediction and phylogenomics. Mol Biol Evol 35:543–548. <https://doi.org/10.1093/molbev/msx319>.
- Blin K, Shaw S, Steinke K, Villebro R, Ziemert N, Lee SY, Medema MH, Weber T. 2019. antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. Nucleic Acids Res 47:W81–W87. <https://doi.org/10.1093/nar/gkz310>.