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

Kiesewalter, Heiko T.; Lozano-Andrade, Carlos N.; Maróti, Gergely; Snyder, Dan; Snyder, Dan; Jørgensen, Tue Sparholt; Weber, Tilmann; Kovács, Ákos T.

Published in:
Microbiology Resource Announcements

Link to article, DOI:
10.1128/mra.01406-19

Publication date:
2020

Document Version
Publisher's PDF, also known as Version of record

Link back to DTU Orbit

Citation (APA):

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.
Complete Genome Sequences of 13 \textit{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

ABSTRACT \textit{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 \textit{B. subtilis} strains isolated from different soil samples in Germany and Denmark.

Various species of the \textit{Bacillus} genus have been exploited for biocontrol of crop plants. \textit{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 \textit{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 \textit{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 \textit{B. subtilis} strains in order to facilitate a detailed investigation of genes involved in secondary metabolite production.

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

For Illumina sequencing, genomic DNA of the \textit{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-
pending the cell pellet in 200 µl 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), bacillicin (pks), and bacillibactin (dbh) biosynthesis in all isolates except strain P5_B2, which lacks the majority of the bacillicin (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. The complete genome sequence accession numbers are listed in Table 1.

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

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

* CDS, coding sequences.

**ACKNOWLEDGMENT**

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