Draft Genome Sequence of the Type Strain *Bacillus subtilis* subsp. *subtilis* DSM10

Lars Lilge, Robert Hertel, Kambiz Morabbi Heravi, Marius Henkel, Fabian M. Commichau, Rudolf Hausmann

ABSTRACT  The *Bacillus subtilis* subsp. *subtilis* type strain DSM10 has been used as a reference in various studies. However, detailed information about the genome has not been available. Therefore, whole-genome sequencing was performed, and the sequence was compared with that of the related *B. subtilis* strain NCIB3610.

The *Bacillus subtilis* subsp. *subtilis* type strain DSM10 is a generally accessible *Bacillus* strain from the German Collection of Microorganisms and Cultures GmbH (DSMZ). It has been used as a reference strain for applied biotechnological research (1–4). For instance, the DSM10 strain produces notable amounts of surfactin (2, 5, 6) and secretes proteases (data not shown). Due to its descent from the *B. subtilis* Marburg strain (7), it is feasible to genetically engineer the DSM10 strain, making it a promising bacterial system for basic research and industrial strain engineering. According to the DSMZ, *B. subtilis* DSM10 corresponds to *B. subtilis* strain NCIB3610 (8). Whole-genome sequencing of *B. subtilis* DSM10 was performed to verify this assumption.

A single colony was inoculated in LB medium and cultivated overnight at 37°C and 120 rpm. Subsequently, chromosomal DNA was extracted with an innuPrep bacterial DNA kit (Analytik Jena, Jena, Germany). Library preparation and whole-genome sequencing were performed by Eurofins Genomics (Ebersberg, Germany). An Illumina HiSeq 4000 system was employed for sequencing, resulting in 2 x 101-bp paired-end read files. The paired-end reads obtained (2 x 35.9 million) were quality analyzed with FastQC v0.11.9 (9). A subset of 2 x 5 million reads, randomly extracted with seqtk v1.3-r106 (10), were used for short-read assembly with SPAdes v3.14.0 (11) with the option “careful.” This resulted in 26 contigs of >0.2 kb with a total size of 4,166,758 bp, an N50 value of 1,014,761 bp, and an N90 value of 240,612 bp. The genomes of the sibling strain *B. subtilis* NCIB3610 (GenBank accession number CP020102) and its plasmid pBS32 (GenBank accession number CP020103) were used as references to sort and correctly orient the acquired draft genome of *B. subtilis* DSM10 with the program Mauve v2015-02-13 (http://www.darlinglab.org/mauve). The alignment obtained revealed an almost perfect match of the DSM10 draft genome to the chromosome of NCIB3610. However, contigs resembling pBS32 were not observed, confirming its absence in DSM10. The DSM10 genome underwent automated gene annotation by the Prokaryotic Genome Annotation Pipeline (PGAP) during uploading to GenBank (12). This process led to the identification of 4,289 genes, of which 4,252 are protein-coding genes and 37 are pseudogenes. Moreover, 42 tRNAs and 5 noncoding RNAs were identified and annotated. The program breseq v0.35.1 (13) was used to identify specific differences by applying DSM10 reads to the NCIB3610 genome. In all, 39 sequence variations could be identified (Table 1). Seventeen single-nucleotide polymorphisms restored a corresponding NCIB3610 pseudogene.

Citation Lilge L, Hertel R, Morabbi Heravi K, Henkel M, Commichau FM, Hausmann R. 2021. Draft genome sequence of the type strain *Bacillus subtilis* subsp. *subtilis* DSM10. Microbiol Resour Announc 10:e00158-21. https://doi.org/10.1128/MRA.00158-21.

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

Copyright © 2021 Lilge et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license.

Address correspondence to Lars Lilge, lars.lilge@uni-hohenheim.de.

Received 10 February 2021
Accepted 11 February 2021
Published 11 March 2021
| Contig accession no. | Nucleotide position | Mutation | Annotation | Gene | Description |
|---------------------|---------------------|----------|------------|------|-------------|
| JAEPVU010000002.1   | 26036               | C→T     | P307L (CCG→CTG) | nmd  | —           |
| JAEPVU010000003.1   | 11953               | —G      | Gene-pseudogene | cysE | Serine O-acetyltransferase |
| JAEPVU010000006.1   | 105978              | (ATGATG)→T | Intergenic | H9596_03325 ←→ H9596_03330 | Catalase/IRNA-Arg |
| JAEPVU010000007.1   | 83281               | (G)→→   | Gene-pseudogene | yerH | Hypothetical protein |
| JAEPVU010000007.1   | 85980               | (C)→→   | Gene-pseudogene | sapB | Methyltransferase |
| JAEPVU010000007.1   | 141088              | T→G     | V142A (GCC→GTG) | lipB | Protein lipB |
| JAEPVU010000007.1   | 147278              | (G)→→   | Intergenic | H9596_04045 ←→ yezD | Sulfate transporter/DUF2292 domain-containing protein |
| JAEPVU010000009.1   | 241945              | (G)→→   | Gene-pseudogene | acol | Dihydroxypropionate dehydratase |
| JAEPVU010000009.1   | 94306               | (A)→→   | Gene-pseudogene | bmrD (yheH) | Multidrug ABC transporter permease |
| JAEPVU010000009.1   | 234838              | —ATGAC  | Coding | yitS | Fatty acid-binding protein DegV |
| JAEPVU010000009.1   | 320518              | (G)→→   | Gene-pseudogene | manP | PTS mannose transporter subunit IIABC |
| JAEPVU010000009.1   | 357081              | —G      | Gene-pseudogene | uxaB | Alloctonate oxidoreductase |
| JAEPVU010000009.1   | 358313              | (G)→→   | Gene-pseudogene | uxaB | Alloctonate oxidoreductase |
| JAEPVU010000009.1   | 375395              | (C)→→   | Gene-pseudogene | xkel | Phage portal protein |
| JAEPVU010000009.1   | 471575              | (TAAT)→→ | Intergenic | mtnK ←→ mtnU | S-Methyl-S-thioribose kinase/hydrolase MtnU |
| JAEPVU010000009.1   | 514726              | (T)→→   | Intergenic | ykwD ←→ pbpH | Hypothetical protein/penicillin-binding protein |
| JAEPVU010000009.1   | 737764              | A→→     | II76V (ATC→GTC) | codY | GTP-sensing pleiotropic transcriptional regulator CodY |
| JAEPVU010000009.1   | 956741              | A→→     | S179G (AGT→GTT) | yneD | Germination |
| JAEPVU010000011.1   | 175058              | C→→     | Coding (131/138 nt) | yaeE | Hypothetical protein |
| JAEPVU010000011.1   | 222781              | C→→     | V91V (GTC→GGT) | yopW | Hypothetical protein |
| JAEPVU010000011.1   | 379673              | (C)→→   | Intergenic | msgA ←→ dapB | Methyglyoxal synthase/4-hydroxy-tetrahydrodipicolinate reductase |
| JAEPVU010000011.1   | 396124              | (C)→→   | Gene-pseudogene | trpD | Anthranilate phosphoribosyltransferase |
| JAEPVU010000011.1   | 534030              | (C)→→   | Gene-pseudogene | mmaA | Acetyl-CoA acetyltransferase |
| JAEPVU010000011.1   | 638800              | —C      | Gene-pseudogene | yufA (flaA)← | Hypothetical protein |
| JAEPVU010000011.1   | 719953              | (T)→→   | Gene-pseudogene | yopD ← | Hypothetical protein |
| JAEPVU010000011.1   | 728398              | (C)→→   | Intergenic | yklK ←→ yklL | N-Acetyltransferase/general stress protein |
| JAEPVU010000011.1   | 749018              | (C)→→   | Gene-pseudogene | azID | Branched-chain amino acid transferase AazD |
| JAEPVU010000011.1   | 752943              | (C)→→   | Intergenic | cyPA ←→ yrdC | Cytochrome P450/cysteine hydrolase |
| JAEPVU010000011.1   | 875960              | A→→     | G66G (GTT→GGC) | rplU | 50S ribosomal protein L21 |
| JAEPVU010000011.1   | 900227              | (T)→→   | Coding | engB | YihA family ribosome biogenesis GTP-binding protein |
| JAEPVU010000011.1   | 982769              | (C)→→   | Intergenic | ytxC ←→ ytbB | Hypothetical protein/TVP38/TMEM64 family membrane protein YtxB |
| JAEPVU010000011.1   | 1079814             | (C)→→   | Intergenic | ytmP ←→ trmB | MBL fold metallo-hydrolase/tRNA (guanosine-46-N7)-methyltransferase TrmB |
| JAEPVU010000012.1   | 88                  | A→→     | Noncoding | H9596_16680 ← | tRNA-Aa |
| JAEPVU010000013.1   | 61274               | (G)→→   | Intergenic | maIR (yuFM) ←→ yufN (nuPK) | Two-component system response regulator DcuR/BMP family ABC transporter substrate-binding protein |
| JAEPVU010000013.1   | 222225              | (T)→→   | Gene-pseudogene | yveE | Two-component sensor histidine kinase |
| JAEPVU010000013.1   | 317921              | —C      | Gene-pseudogene | yveT | — |
| JAEPVU010000013.1   | 401206              | (T)→→   | Gene-pseudogene | yveA | Hypothetical protein |
| JAEPVU010000013.1   | 414816              | G→→     | L256F (CCT→TTC) | lgt | Prolipoprotein diacylglycerol transferase |
| JAEPVU010000014.1   | 21905               | G→→     | F118L (TTC→TTA) | rocB | Peptidase M20 |
| JAEPVU010000014.1   | 220111              | A→→     | L25L (TTG→CTG) | iolG | Inositol 2-dehydrogenase |

*a Variations identified between *B. subtilis* DSM10*®* and strain NCIB3610, nt, nucleotides.

*b PTS, phosphotransferase system; CoA, coenzyme A.
In conclusion, *B. subtilis* DSM10\(^T\) is genomically very similar to NCIB3610 but is not identical. The absence of pBS32 could explain the ability of DSM10\(^T\) to develop competence because no plasmid-borne single-pass transmembrane protein ComI is present to downregulate it, as it is for NCIB3610 (14).

**Data availability.** The genome sequence of *B. subtilis* subsp. *subtilis* strain DSM10\(^T\) has been deposited in GenBank under the accession number JAEPVU000000000. The raw sequence reads have been submitted to the NCBI Sequence Read Archive (SRA) (15) under the accession number SRR12632401. The BioProject accession number is PRJNA659394, and the BioSample accession number is SAMN15904628.

**ACKNOWLEDGMENTS**

This project was funded by the German Research Foundation (DFG) (project 398354917). The funders had no role in study design, data collection, interpretation, or the decision to submit the work for publication.

We thank Eike Grunwaldt and Maliheh Vahidinasab for their technical assistance.

**REFERENCES**

1. O'Mahony MM, Henneberger R, Selvin J, Kennedy J, Doohan F, Marchesi JR, Dobson ADW. 2015. Inhibition of the growth of *Bacillus subtilis* DSM10 by a newly discovered antibacterial protein from the soil metagenome. Bioengineered 6:89–98. https://doi.org/10.1080/21655979.2015.1018493.

2. Willenbacher J, Yeremchuk W, Mohr T, Syldatk C, Hausmann R. 2015. Enhancement of surfactin yield by improving the medium composition and fermentation process. AMB Express 5:145. https://doi.org/10.1186/s13668-015-0145-0.

3. Kuephadungphan W, Helaly SE, Daengrot C, Phongpaichit S, Luangsa-Ard JJ, Rukachaisirikul V, Studier M. 2017. Akanthopyrones A–D, a-pyrones bearing a 4-O-methyl-D-glucopyranose moiety from the spider-associated ascomycete *Akanthomyces novoguineensis*. Molecules 22:1202. https://doi.org/10.3390/molecules22071202.

4. Sabdaningsih A, Liu Y, Mettal U, Heep J, Riyanti Wang L, Cristianawati O, Nuryadi H, Sibero MT, Marner M, Radjasa OK, Sabdono A, Trianto A, Schäberle TF. 2020. A new citrinin derivative from the Indonesian marine sponge-associated fungus *Penicillium citrinum*. Molecules 25:1202. https://doi.org/10.3390/molecules25071202.

5. Geissler M, Oellig C, Moss K, Schwack W, Henkel M, Hausmann R. 2017. High-performance thin-layer chromatography (HPTLC) for the simultaneous quantification of the cyclic lipopeptides surfactin, iturin A and fengycin in culture samples of *Bacillus* species. J Chromatogr B Analyt Technol Biomed Life Sci 1044–1045:214–224. https://doi.org/10.1016/j.chroma.2016.11.013.

6. Geissler M, Kühle I, Heravi KM, Altenbuchner J, Henkel M, Hausmann R. 2019. Evaluation of surfactin synthesis in a genome reduced *Bacillus subtilis* strain. AMB Express 9:84. https://doi.org/10.1186/s13668-019-0806-5.

7. Hanwood CR, Wipat A. 1996. Sequencing and functional analysis of the genome of *Bacillus subtilis* strain 168. FEBS Lett 389:84–87. https://doi.org/10.1016/0014-5793(96)00524-8.

8. Nye TM, Schroeder JW, Kears DB, Simmons LA. 2017. Complete genome sequence of undomesticated *Bacillus subtilis* strain NCIB 3610. Genome Announc 5:e00364-17. https://doi.org/10.1128/genomeA.00364-17.

9. Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data. https://www.bioinformatics.babraham.ac.uk/projects/fastqc.

10. Li H. 2012. Seqtk: a toolkit for processing sequences in FASTA/Q formats. https://github.com/lh3/seqtk.

11. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prijibelski AD, Pyshkin AV, Sidorenkov 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.

12. Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, Lomsadze A, Prut KD, Borodovsky M, Ostell J. 2016. NCBI Prokaryotic Genome Annotation Pipeline. Nucleic Acids Res 44:6614–6624. https://doi.org/10.1093/nar/gkw569.

13. Deatherage DE, Barrick JE. 2014. Identification of mutations in laboratory-evolved microbes from next-generation sequencing data using breseq. Methods Mol Biol 1151:165–188. https://doi.org/10.1007/978-1-4939-0554-6_12.

14. Konkol MA, Blair KM, Kearns DB. 2013. Plasmid-encoded ComI inhibits competence in the ancestral 3610 strain of *Bacillus subtilis*. J Bacteriol 195:4085–4093. https://doi.org/10.1128/JB.00696-13.

15. Leinonen R, Sugawara H, Shumway M. 2011. The Sequence Read Archive. Nucleic Acids Res 39:D19–D21. https://doi.org/10.1093/nar/gkq1019.