



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

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

^aUniversity of Hohenheim, Institute of Food Science and Biotechnology, Department of Bioprocess Engineering (150k), Stuttgart, Germany

^bBrandenburg University of Technology, FG Synthetic Microbiology, Cottbus-Senftenberg, Germany

Lars Lilge and Robert Hertel contributed equally to this work. Author order was determined by the funding, which was provided by a working group of University Hohenheim, for which Lars Lilge is the principal investigator.

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^T 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^T strain, making it a promising bacterial system for basic research and industrial strain engineering. According to the DSMZ, *B. subtilis* DSM10^T corresponds to *B. subtilis* strain NCIB3610 (8). Whole-genome sequencing of *B. subtilis* DSM10^T 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 × 101-bp paired-end read files. The paired-end reads obtained (2 × 35.9 million) were quality analyzed with FastQC v0.11.9 (9). A subset of 2 × 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 N_{50} value of 1,014,761 bp, and an N_{90} value of 240,612 bp. The genomes of the sibling strain *B. subtilis* NCIB3610 (GenBank accession number [CP020102](https://doi.org/10.1128/MRA.00158-21)) and its plasmid pBS32 (GenBank accession number [CP020103](https://doi.org/10.1128/MRA.00158-21)) were used as references to sort and correctly orient the acquired draft genome of *B. subtilis* DSM10^T with the program Mauve v2015-02-13 (<http://www.darlinglab.org/mauve>). The alignment obtained revealed an almost perfect match of the DSM10^T draft genome to the chromosome of NCIB3610. However, contigs resembling pBS32 were not observed, confirming its absence in DSM10^T. The DSM10^T 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](https://creativecommons.org/licenses/by/4.0/).

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

Received 10 February 2021

Accepted 11 February 2021

Published 11 March 2021

TABLE 1 Sequence variations between *B. subtilis* DSM10^T and NCIB3610

Contig accession no.	Nucleotide position	Mutation	Annotation ^a	Gene	Description ^b
JAEPVU010000002.1	26036	C→T	P307L (CCG→CTG)	<i>mtfD</i> →	Transcription repair coupling factor
JAEPVU010000003.1	11953	—G	Gene-pseudogene	<i>cysE</i> →	Serine O-acetyltransferase
JAEPVU010000006.1	105978	(ATGATAGT)1→2	Intergenic	H9S96_03325 ← / → H9S96_03330	Catalase/tRNA-Arg
JAEPVU010000007.1	83281	(G)7→6	Gene-pseudogene	<i>yerH</i> →	Hypothetical protein
JAEPVU010000007.1	85980	(C)6→5	Gene-pseudogene	<i>sapB</i> ←	Methyltransferase
JAEPVU010000007.1	141088	T→C	V142A (GCG→GTG)	<i>lplB</i> →	Protein lplB
JAEPVU010000007.1	147278	(G)6→5	Intergenic	H9S96_04045 → / → <i>yezD</i>	Sulfate transporter/DUF2292 domain-containing protein
JAEPVU010000007.1	241945	(G)6→5	Gene-pseudogene	<i>acoL</i> →	Dihydrolipoyl dehydrogenase
JAEPVU010000009.1	94306	(A)7→6	Gene-pseudogene	<i>bmrD (yheH)</i> →	Multidrug ABC transporter permease
JAEPVU010000009.1	234838	—ATGTAC	Coding	<i>yitS</i> ←	Fatty acid-binding protein DegV
JAEPVU010000009.1	320518	(G)6→5	Gene-pseudogene	<i>manP</i> →	PTS mannose transporter subunit IIABC
JAEPVU010000009.1	357081	—G	Gene-pseudogene	<i>uxaB</i> →	Altronate oxidoreductase
JAEPVU010000009.1	358313	(G)6→5	Gene-pseudogene	<i>uxaB</i> →	Altronate oxidoreductase
JAEPVU010000009.1	375395	(C)6→5	Gene-pseudogene	<i>xkdE</i> →	Phage portal protein
JAEPVU010000009.1	471575	(TAAT)4→3	Intergenic	<i>mtmK</i> ← / ← <i>mtmU</i>	S-Methyl-5-thioribose kinase/hydrolase MtmU
JAEPVU010000009.1	514727	(T)8→7	Intergenic	<i>ykwD</i> ← / → <i>pbpH</i>	Hypothetical protein/penicillin-binding protein
JAEPVU010000009.1	737764	A→G	I176V (ATC→GTC)	<i>cody</i> →	GTP-sensing pleiotropic transcriptional regulator CodY
JAEPVU010000009.1	956741	A→G	S179G (AGT→GGT)	<i>yndE</i> →	Germination protein
JAEPVU010000011.1	175058	—C	Coding (131/138 nt)	<i>yotE</i> ←	Hypothetical protein
JAEPVU010000011.1	222781	C→A	V91V (GTG→GTT)	<i>yopW</i> ←	Hypothetical protein
JAEPVU010000011.1	379673	(C)6→5	Intergenic	<i>mgsA</i> ← / ← <i>dapB</i>	Methylglyoxal synthase/4-hydroxy-tetrahydrolipicolinate reductase
JAEPVU010000011.1	396124	(C)7→6	Gene-pseudogene	<i>trpD</i> ←	Anthranilate phosphoribosyltransferase
JAEPVU010000011.1	534030	(C)7→6	Gene-pseudogene	<i>mrgA</i> ←	Acetyl-CoA acetyltransferase
JAEPVU010000011.1	638280	—C	Gene-pseudogene	<i>yqfA (floA)</i> ←	Hypothetical protein
JAEPVU010000011.1	719953	(T)8→7	Gene-pseudogene	<i>yqadD</i> ←	Hypothetical protein
JAEPVU010000011.1	728398	(C)8→7	Intergenic	<i>yrkN</i> → / ← <i>yrkL</i>	N-Acetyltransferase/general stress protein
JAEPVU010000011.1	749018	(C)6→5	Gene-pseudogene	<i>azlD</i> ←	Branched-chain amino acid transporter AzlD
JAEPVU010000011.1	752943	(C)6→5	Intergenic	<i>cypA</i> ← / ← <i>yrdC</i>	Cytochrome P450/cysteine hydrolase
JAEPVU010000011.1	875960	A→G	G66G (GGT→GGC)	<i>rplU</i> ←	50S ribosomal protein L21
JAEPVU010000011.1	900227	(T)7→6	Coding	<i>engB</i> ←	YihA family ribosome biogenesis GTP-binding protein
JAEPVU010000011.1	982769	(C)6→5	Intergenic	<i>ytxC</i> ← / ← <i>ytxB</i>	Hypothetical protein/TVP38/TMEM64 family membrane protein YtxB
JAEPVU010000011.1	1079814	(C)8→7	Intergenic	<i>ytnP</i> ← / ← <i>trmB</i>	MBL fold metallo-hydrolase/tRNA (guanosine-46-N7)-methyltransferase TrmB
JAEPVU010000012.1	88	A→G	Noncoding	H9S96_16680 ←	tRNA-Ala
JAEPVU010000013.1	61274	(G)6→5	Intergenic	<i>malR (yurfM)</i> → / → <i>yurfN (nupN)</i>	Two-component system response regulator DcurR/BMP family ABC transporter substrate-binding protein
JAEPVU010000013.1	227225	(T)6→5	Gene-pseudogene	<i>yvrE</i> ←	Sugar lactone lactonase YvrE
JAEPVU010000013.1	317921	—C	Gene-pseudogene	<i>yvrT</i> ←	Two-component sensor histidine kinase
JAEPVU010000013.1	401206	(T)8→7	Gene-pseudogene	<i>ywzA</i> ←	Hypothetical protein
JAEPVU010000013.1	414816	G→A	L256F (CTC→TTC)	<i>lgt</i> ←	Prolipoprotein diacylglycerol transferase
JAEPVU010000014.1	21905	G→T	F118L (TTC→TTA)	<i>rocB</i> ←	Peptidase M20
JAEPVU010000014.1	220111	A→G	L25L (TTG→CTG)	<i>ioIG</i> ←	Inositol 2-dehydrogenase

^a Variants identified between *B. subtilis* strain DSM10^T 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](https://ncbi.nlm.nih.gov/GenBank/entry/JAEPVU000000000). The raw sequence reads have been submitted to the NCBI Sequence Read Archive (SRA) (15) under the accession number [SRR12632401](https://ncbi.nlm.nih.gov/SRA/entry/SRR12632401). The BioProject accession number is [PRJNA659394](https://ncbi.nlm.nih.gov/BioProject/entry/PRJNA659394), and the BioSample accession number is [SAMN15904628](https://ncbi.nlm.nih.gov/BioSample/entry/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

- 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>.
- Willenbacher J, Yeremchuk W, Mohr T, Sylatk 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/s13568-015-0145-0>.
- Kuephadungphan W, Helaly SE, Daengrot C, Phongpaichit S, Luangsa-Ard JJ, Rukachaisirikul V, Stadler M. 2017. Akanthopyrones A–D, α -pyrones bearing a 4-O-methyl- β -D-glucopyranose moiety from the spider-associated ascomycete *Akanthomyces novoguineensis*. *Molecules* 22:1202. <https://doi.org/10.3390/molecules22071202>.
- Sabdaningsih A, Liu Y, Mettal U, Heep J, Riyanti Wang L, Cristianawati O, Nuryadi H, Sibero MT, Marnier M, Radjasa OK, Sabdono A, Trianto A, Schäberle TF. 2020. A new citrinin derivative from the Indonesian marine sponge-associated fungus *Penicillium citrinum*. *Mar Drugs* 18:227. <https://doi.org/10.3390/md18040227>.
- 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.jchromb.2016.11.013>.
- 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/s13568-019-0806-5>.
- Harwood 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](https://doi.org/10.1016/0014-5793(96)00524-8).
- Nye TM, Schroeder JW, Kearns 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>.
- Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data. <https://www.bioinformatics.babraham.ac.uk/projects/fastqc>.
- Li H. 2012. Seqtk: a toolkit for processing sequences in FASTA/Q formats. <https://github.com/lh3/seqtk>.
- Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski 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>.
- Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, Lomsadze A, Pruitt KD, Borodovsky M, Ostell J. 2016. NCBI Prokaryotic Genome Annotation Pipeline. *Nucleic Acids Res* 44:6614–6624. <https://doi.org/10.1093/nar/gkw569>.
- 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.
- 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>.
- Leinonen R, Sugawara H, Shumway M. 2011. The Sequence Read Archive. *Nucleic Acids Res* 39:D19–D21. <https://doi.org/10.1093/nar/gkq1019>.