

Complete Genome Sequences of *Bacillus subtilis* subsp. *subtilis* Laboratory Strains JH642 (AG174) and AG1839

Janet L. Smith,^a Jonathan M. Goldberg,^{b*} Alan D. Grossman^a

Department of Biology, MIT, Cambridge, Massachusetts, USA^a; Broad Institute, Cambridge, Massachusetts, USA^b

* Present address: Jonathan M. Goldberg, Department of Immunology and Infectious Disease, Harvard School of Public Health, Boston, Massachusetts, USA.

The Gram-positive bacterium *Bacillus subtilis* is widely used for studies of cellular and molecular processes. We announce the complete genomic sequences of strain AG174, our stock of the commonly used strain JH642, and strain AG1839, a derivative that contains a mutation in the replication initiation gene *dnaB* and a linked Tn917.

Received 10 June 2014 Accepted 16 June 2014 Published 3 July 2014

Citation Smith JL, Goldberg JM, Grossman AD. 2014. Complete genome sequences of *Bacillus subtilis* subsp. *subtilis* laboratory strains JH642 (AG174) and AG1839. *Genome Announc.* 2(4):e00663-14. doi:10.1128/genomeA.00663-14.

Copyright © 2014 Smith et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 3.0 Unported license](https://creativecommons.org/licenses/by/3.0/).

Address correspondence to Alan D. Grossman, adg@mit.edu.

Bacillus subtilis is a soil bacterium studied in many laboratories and used commercially. It serves as a model for related Gram-positive pathogens. Many cellular processes are studied in *B. subtilis*, including transcription, translation, replication, metabolism, biofilm formation, sporulation, and horizontal gene transfer.

We report the genome sequence of *B. subtilis* subsp. *subtilis* strain AG174, our stock of strain JH642 (1). This widely used strain contains known point mutations conferring auxotrophy for tryptophan and phenylalanine and cold sensitivity (2, 3). We also report the sequence of AG1839 (also known as KPL69), a derivative of AG174 used in studies of replication initiation (4). It contains a mutation in the replication initiation gene *dnaB*, conferring a temperature-sensitive phenotype, and a linked Tn917 insertion.

For AG174, 7.9 million 36-bp Illumina HiSeq reads and 0.88 million 150-bp paired-end reads were aligned to the AG1839 sequence (see below). The final circular sequence was 4,188,369 nucleotides (nt), with 98× mean coverage. Comparison of the finished AG174 genome to the previously reported JH642 genome (GenBank accession number CM000489.1 [3]) using Mauve (5) indicated that the genomes were 99.6% identical, with ~400 indels and ~30 point mutations. The vast majority of these differences appeared to be due to sequencing or alignment issues with the previously reported sequence. In addition, 269 unknown nucleotides (Ns) in the CM000489.1 sequence were resolved.

For AG1839, we used Bowtie (6) to align 15.8 million 40-bp reads to the previously sequenced version of JH642 (3). We used three rounds of alignment and refinement, followed by *de novo* assembly (7) to place the Tn917 insertion in *ydbT*. PCR and conventional sequencing were used to determine the junctions of a 10.8-kb deletion that fuses *ppsC* and *ppsD* (encoding plipistatin synthetase). AG1839 differed from the parental AG174 by the expected point mutation in *dnaB* and the linked Tn917. Two additional point mutations were also present, in *thrS* (T85A) (4 kb upstream of *dnaB* linked to Tn917) and in *ytnA* (M105I) (60 kb downstream of *dnaB*). The final circular sequence was 4,193,640 nt, with 151× mean coverage.

Initial gene sets were predicted using Prodigal (8). The genome sequence of *B. subtilis* 168 (GenBank accession no. NC000964 [9, 10]) was aligned with those of AG174 and AG1839 using NUCMER (11). Coordinates from the alignments were used to add gene models missed by Prodigal and to map gene names and symbols to the new annotations. Gene models with potential problems were corrected or dropped if necessary, resulting in 4,227 and 4,231 predicted protein-coding genes for AG174 and AG1839, respectively. For both genomes, 86 tRNA genes were predicted using TRNASCAN (12) and 10 rRNA operons were predicted using RNAmmer (13).

AG174 differs from PY79, another commonly used laboratory strain (14). AG174 contains ICEBs1 and SP beta. There are 1,734 single-nucleotide polymorphisms (SNPs) in a 71-kb hypervariable region from *panB* to *hepT* and 35 SNPs distributed over the rest of the genome.

Nucleotide sequence accession numbers. The complete genome sequences are in GenBank, accession numbers CP007800 (AG174) and CP008698 (AG1839). AG174 and AG1839 are available from the Bacillus Genetic Stock Center.

ACKNOWLEDGMENTS

Research reported in this publication was supported, in part, by the National Institute of General Medical Sciences of the National Institutes of Health under award number GM41934 to A.D.G.

The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

REFERENCES

1. Brehm SP, Staal SP, Hoch JA. 1973. Phenotypes of pleiotropic-negative sporulation mutants of *Bacillus subtilis*. *J. Bacteriol.* 115:1063–1070.
2. Wiegeshoff F, Marahiel MA. 2007. Characterization of a mutation in the acetolactate synthase of *Bacillus subtilis* that causes a cold-sensitive phenotype. *FEMS Microbiol. Lett.* 272:30–34. <http://dx.doi.org/10.1111/j.1574-6968.2007.00739.x>.
3. Srivatsan A, Han Y, Peng J, Tehranchi AK, Gibbs R, Wang JD, Chen R. 2008. High-precision, whole-genome sequencing of laboratory strains facilitates genetic studies. *PLoS Genet.* 4:e1000139. <http://dx.doi.org/10.1371/journal.pgen.1000139>.

4. Rokop ME, Auchtung JM, Grossman AD. 2004. Control of DNA replication initiation by recruitment of an essential initiation protein to the membrane of *Bacillus subtilis*. *Mol. Microbiol.* 52:1757–1767. <http://dx.doi.org/10.1111/j.1365-2958.2004.04091.x>.
5. Darling AC, Mau B, Blattner FR, Perna NT. 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res.* 14:1394–1403. <http://dx.doi.org/10.1101/gr.2289704>.
6. Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 10:R25. <http://dx.doi.org/10.1186/gb-2009-10-3-r25>.
7. Zerbino DR, Birney E. 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.* 18:821–829. <http://dx.doi.org/10.1101/gr.074492.107>.
8. Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11:119. <http://dx.doi.org/10.1186/1471-2105-11-119>.
9. Kunst F, Ogasawara N, Moszer I, Albertini AM, Alloni G, Azevedo V, Bertero MG, Bessières P, Bolotin A, Borchert S, Borriss R, Boursier L, Brans A, Braun M, Brignell SC, Bron S, Brouillet S, Bruschi CV, Caldwell B, Capuano V, Carter NM, Choi SK, Codani JJ, Connerton IF, Cummings NJ, Daniel RA, Denizot F, Devine KM, Dusterhöft A, Ehrlich SD, Emmerson PT, Entian KD, Errington J, Fabret C, Ferrari E, Foulger D, Fritz C, Fujita M, Fujita Y, Fuma S, Danchin A, et al. 1997. The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*. *Nature* 390:249–256. <http://dx.doi.org/10.1038/36786>.
10. Barbe V, Cruveiller S, Kunst F, Lenoble P, Meurice G, Sekowska A, Vallenet D, Wang T, Moszer I, Médigue C, Danchin A. 2009. From a consortium sequence to a unified sequence: the *Bacillus subtilis* 168 reference genome a decade later. *Microbiology* 155:1758–1775. <http://dx.doi.org/10.1099/mic.0.027839-0>.
11. Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, Salzberg SL. 2004. Versatile and open software for comparing large genomes. *Genome Biol.* 5:R12. <http://dx.doi.org/10.1186/gb-2004-5-6-p12>.
12. Schattner P, Brooks AN, Lowe TM. 2005. The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Res.* 33:W686–W689. <http://dx.doi.org/10.1093/nar/gki366>.
13. Lagesen K, Hallin P, Rødland EA, Staerfeldt HH, Rognes T, Ussery DW. 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res.* 35:3100–3108. <http://dx.doi.org/10.1093/nar/gkm160>.
14. Schroeder JW, Simmons LA. 2013. Complete genome sequence of *Bacillus subtilis* strain PY79. *Genome Announc.* 1(6):e01085-13. <http://dx.doi.org/10.1128/genomeA.01085-13>.