Complete Genome Sequence of *Bacillus subtilis* Strain PY79

Jeremy W. Schroeder, Lyle A. Simmons

Department of Molecular, Cellular, and Developmental Biology, University of Michigan, Ann Arbor, Michigan, USA

Bacillus subtilis **is a Gram-positive soil-dwelling and endospore-forming bacterium in the phylum** *Firmicutes. B. subtilis* **strain PY79 is a prototrophic laboratory strain that has been highly used for studying a wide variety of cellular pathways. Here, we announce the complete whole-genome sequence of** *B. subtilis* **PY79.**

Bacillus subtilis has been studied under laboratory conditions for >100 years, yielding tremendous insight into the biology of Gram-positive bacteria. Laboratory studies have primarily used the strains *B. subtilis* PY79 and JH642 (1). JH642 is auxotrophic and contains a number of phage and integrative conjugative elements (2–4). PY79 is a prototroph lacking many of the mobile genetic elements studied in JH642 (4, 5). The whole-genome shotgun sequence for JH642 is available, but it contains 286 nucleotides located in regions of ambiguous sequence. The JH642 genome sequence and those of other *B. subtilis* strains facilitate in-depth studies of biological mechanisms (3, 6). It is surprising, then, that although PY79 has been one of the most widely used laboratory strains, its genome sequence has been unavailable. Here, we report the complete genome sequence of *B. subtilis* PY79, generated using two sequencing platforms, PacBio RS II and HiSeq 2000 (Illumina). The PY79 genome is 154,156 nucleotides shorter than that of JH642, and, using the script run-mummer3, we found there are 3,641 single-nucleotide polymorphisms (SNPs) between JH642 and PY79 (7). Our results provide the first publicly available complete reference genome for this highly studied *B. subtilis* strain.

PY79 genomic DNA was isolated by phenol-chloroform extraction (8), and a 15-kb insert library was prepared for sequencing using two single-molecule real-time (SMRT) cells on a Pacific Biosciences RS II sequencer. The resulting mean subread length was 3.57 kb. The HGAP protocol implemented in smrtanalysis version 2.0.1 was used to assemble the PY79 genome (9). This resulted in two contigs in the HGAP output, the first of which is short, at exactly 13,000 bases in length, with $5.3\times$ mean coverage. The second contig is 4,060,232 bases long, with $156\times$ mean coverage. Because of its short length and low coverage, we eliminated the first contig from further analysis. The long contig was circularized, and overlaps from the ends were removed using the minimus2 script in the AMOS package (10). The circularized genome was then used as a reference for realignment twice using our original PacBio data and additionally corrected by realignment with high-coverage ($464\times$) 50-base paired-end reads from the HiSeq 2000 platform. This correction step resulted in a complete *B. subtilis* PY79 chromosome sequence that is 4,033,459 bases long.

Genes were predicted using the RAST server (11). RAST located 4,278 features, including 4,140 coding sequences, 30 rRNA genes, and 86 tRNA genes.

Nucleotide sequence accession number. The whole-genome sequence of PY79 is available from DDBJ/EMBL/GenBank databases with accession no. CP006881. PY79 and many derivatives are available from the Bacillus Genetic Stock Center (http://www.bgsc.org/).

ACKNOWLEDGMENTS

REFERENCES

1. **Sonenshein AL, Hoch JA, Losick R.** 1993. *Bacillus subtilis* and other Gram-positive bacteria: biochemistry, physiology and molecular genetics. American Society for Microbiology, Washington, DC.
2. **Auchtung JM, Lee CA, Monson RE, Lehman AP, Grossman AD.** 2005. Regulation of a *Bacillus subtilis* mobile genetic element by intercellular signaling and the global DNA damage response. Proc. Natl. Acad. Sci. U. S. A. **102**:12554–12559.
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. doi:10.1371/journal.pgen.1000139.
4. **Youngman P, Perkins JB, Losick R.** 1984. Construction of a cloning site near one end of Tn917 into which foreign DNA may be inserted without affecting transposition in *Bacillus subtilis* or expression of the transposon-borne *erm* gene. Plasmid **12**:1–9.
5. **Zeigler DR, Prágai Z, Rodriguez S, Chevreux B, Muffler A, Albert T, Bai R, Wyss M, Perkins JB.** 2008. The origins of 168, W23, and other *Bacillus subtilis* legacy strains. J. Bacteriol. **190**:6983–6995.
6. **Earl AM, Eppinger M, Fricke WF, Rosovitz MJ, Rasko DA, Daugherty S, Losick R, Kolter R, Ravel J.** 2012. Whole-genome sequences of *Bacillus subtilis* and close relatives. J. Bacteriol. **194**:2378–2379.
7. **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. doi:10.1186/gb-2004-5-2-r12.
8. **Harwood CR, Cutting SM.** 1990. Molecular biological methods for *Bacillus*. John Wiley & Sons, Chichester, England.
9. **Chin CS, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, Clum A, Copeland A, Huddleston J, Eichler EE, Turner SW, Korlach J.** 2013. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. Nat. Methods **10**:563–569.
10. **Treangen TJ, Sommer DD, Angly FE, Koren S, Pop M.** 2011. Next-generation sequence assembly with AMOS. Curr. Protoc. Bioinformatics **11**:11.8. doi:10.1002/0471250953.bi1108s33.
11. **Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O.** 2008. The RAST server: Rapid Annotations using Subsystems Technology. BMC Genomics **9**:75. doi:10.1186/1471-2164-9-75.