

# Whole-Genome Sequences of *Bacillus subtilis* and Close Relatives

Ashlee M. Earl,<sup>a\*</sup> Mark Eppinger,<sup>b</sup> W. Florian Fricke,<sup>b</sup> M. J. Rosovitz,<sup>c\*</sup> David A. Rasko,<sup>b</sup> Sean Daugherty,<sup>b</sup> Richard Losick,<sup>d</sup> Roberto Kolter,<sup>a</sup> and Jacques Ravel<sup>b</sup>

Harvard Medical School, Department of Microbiology & Molecular Genetics, Boston, Massachusetts, USA<sup>a</sup>; Institute for Genome Sciences (IGS), University of Maryland, School of Medicine, Department of Microbiology and Immunology, Baltimore, Maryland, USA<sup>b</sup>; J. Craig Venter Institute, Rockville, Maryland, USA<sup>c</sup>; and Harvard University, Department of Molecular & Cellular Biology, Cambridge, Massachusetts, USA<sup>d</sup>

**We sequenced four strains of *Bacillus subtilis* and the type strains for two closely related species, *Bacillus vallismortis* and *Bacillus mojavensis*. We report the high-quality Sanger genome sequences of *B. subtilis* subspecies *subtilis* RO-NN-1 and AUSI98, *B. subtilis* subspecies *spizizenii* TU-B-10<sup>T</sup> and DV1-B-1, *Bacillus mojavensis* RO-H-1<sup>T</sup>, and *Bacillus vallismortis* DV1-F-3<sup>T</sup>.**

*Bacillus subtilis* is a model Gram-positive ubiquitous soil bacterium capable of forming endospores. Since its discovery in Marburg, Germany, in the late 1800s, there have been thousands of publications addressing various aspects of *B. subtilis* biology, leading to a level of understanding of this microbe that is virtually unparalleled in the life sciences. The genomes presented here will enable in-depth investigation of *B. subtilis* sequence variation and detailed whole-genome comparisons of this important model species.

Strains were selected for sequencing to represent phylogenetically diverse groups within and closely related to the *B. subtilis* species, including the type strains for these groups. *B. subtilis* subspecies *subtilis* strains RO-NN-1 (4,011,949 bp) (GenBank accession no. CP002906) and AUSI98 (128 contigs) (GenBank accession no. AFSF00000000) were isolated from soils originating from the Mojave Desert in Rosamond, CA, and Salzburg, Austria, respectively (10, 14). These isolates are closely related to the well-known laboratory strain *B. subtilis* 168<sup>T</sup>, the type strain for the *B. subtilis* subspecies (1, 17), as well as the recently sequenced strains BSn5 and BEST195 (3, 11). TU-B-10<sup>T</sup> (4,207,222 bp) (GenBank accession no. CP002905), the type strain for *Bacillus subtilis* subspecies *spizizenii* (10, 14), and DV1-B-1 (20 contigs) (GenBank accession no. AFSG00000000) (15) are closely related to the recently sequenced laboratory strain W23 and marine strain gTP20b (7, 18). TU-B-10<sup>T</sup> and DV1-B-1 were isolated from soils collected near Nefta, Tunisia, and Death Valley National Monument, CA (10, 14, 15), respectively. *Bacillus vallismortis* DV1-F-3<sup>T</sup> (94 contigs) (GenBank accession no. AFSH00000000) is the type strain for this *B. subtilis* close relative and was also isolated from Death Valley National Monument, CA (14, 16). RO-H-1<sup>T</sup> (45 contigs) (GenBank accession no. AFSI00000000) represents the type strain for *Bacillus mojavensis*, another *B. subtilis* close relative that was isolated from Mojave Desert soil in Rosamond, CA (15).

Genomic DNA was subjected to random whole-genome Sanger or hybrid Sanger-454 shotgun sequencing and closure strategies as previously described (6). Plasmid (pHOS2) and fosmid (pCC1fos) libraries were constructed with target insert sizes of 3 to 5, 10 to 12, and 30 to 39 kb. Sequences were assembled using the Celera assembler (8). All genomes were manually annotated using the Manatee system (<http://manatee.sourceforge.net/>). These genomes exhibit a high degree of proteome conservation and synteny genome architectures (2, 12, 13). The synteny appears to be well conserved even in more distantly related bacilli (5, 7, 9). However, each strain harbors numerous strain-specific re-

gions that are found interspersed throughout the conserved genomic backbone. Such strain-specific regions vary enormously in size (1 kb to 100 kb), the majority being smaller than 5 kb. Genomic islands harbor secondary metabolism and developmental genes (e.g., sporulation and competence), suggesting that these pathways are plastic, subject to environmental selection, which may explain how *B. subtilis* has become so broadly adapted to different ecological niches (4).

Access to these high-quality genome sequences and their comparative analyses with multiple genomes from *B. subtilis* and close relatives will expand our understanding of gene flow and speciation among sympatric soil bacilli. Understanding the genetic diversity and genome dynamics in the bacilli will further aid experimental analyses of this important model species and allow insights into the physiology, ecology, and evolution of the group.

**Nucleotide sequence accession numbers.** Strains have been deposited in the Bacillus Genetic Stock Center and ARS (NRRL) Culture Collection under accession numbers BGSCID 3A27 and NRRL B-14823 (RO-NN-1), BGSCID 3A26 (AUSI98), BGSCID 2A11<sup>T</sup> and NRRL B-23049<sup>T</sup> (TU-B-10<sup>T</sup>), BGSCID 2A12 and NRRL B-23054 (DV1-B-1), BGSCID 28A5 and NRRL B-14698 (RO-H-1<sup>T</sup>), and BGSCID 28A4 and NRRL B-14890 (DV1-F-3<sup>T</sup>).

## ACKNOWLEDGMENTS

The genome sequencing projects were supported with funds from National Science Foundation grant NSF-EF-0523471 and the University of Maryland School of Medicine.

We thank John Perkins, Patrick Eichenberger, Dan Kearns, John D. Helmann, Tina Henkin, Jeff Meisner, Matthias Schmalish, Yunrong (Win) Chai, Michael D'Elia, Jenny Auchtung, Adam Breier, Heiko Liesegang, Evert-Jan Blom, Michiel de Hoon, and Claudio Aguilar for their contributions to the annotation of the genomes.

Received 30 June 2011 Accepted 13 February 2012

Address correspondence to Jacques Ravel, [javel@som.umaryland.edu](mailto:javel@som.umaryland.edu).

\* Present address: Ashlee M. Earl, The Broad Institute of MIT & Harvard, Genome Sequencing & Analysis Program, Cambridge, Massachusetts, USA; M. J. Rosovitz, National Biodefense Analysis and Countermeasures Center (NBACC), Frederick, Maryland, USA.

Copyright © 2012, American Society for Microbiology. All Rights Reserved.

doi:10.1128/JB.05675-11

## REFERENCES

1. Barbe V, et al. 2009. From a consortium sequence to a unified sequence: the *Bacillus subtilis* 168 reference genome a decade later. *Microbiology* 155:1758–1775.
2. Darling AC, Mau B, Blattner FR, Perna NT. 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res.* 14:1394–1403.
3. Deng Y, et al. 2011. Complete genome sequence of *Bacillus subtilis* BSn5, an endophytic bacterium of *Amorphophallus konjac* with antimicrobial activity for the plant pathogen *Erwinia carotovora* subsp. *carotovora*. *J. Bacteriol.* 193:2070–2071.
4. Earl AM, Losick R, Kolter R. 2008. Ecology and genomics of *Bacillus subtilis*. *Trends Microbiol.* 16:269–275.
5. Eppinger M, et al. 2011. Genome sequences of the biotechnologically important *Bacillus megaterium* strains QM B1551 and DSM319. *J. Bacteriol.* 193:4199–4213.
6. Eppinger M, et al. 2010. Genome sequence of the deep-rooted *Yersinia pestis* strain Angola reveals new insights into the evolution and pangenome of the plague bacterium. *J. Bacteriol.* 192:1685–1699.
7. Fan L, et al. 2011. Genome sequence of *Bacillus subtilis* subsp. *spizizenii* gtP20b, isolated from the Indian ocean. *J. Bacteriol.* 193:1276–1277.
8. Huson DH, et al. 2001. Design of a compartmentalized shotgun assembler for the human genome. *Bioinformatics* 17(Suppl 1):S132–S139.
9. Kunst F, et al. 1997. The complete genome sequence of the gram-positive bacterium *Bacillus subtilis*. *Nature* 390:249–256.
10. Nakamura LK, Roberts MS, Cohan FM. 1999. Relationship of *Bacillus subtilis* clades associated with strains 168 and W23: a proposal for *Bacillus subtilis* subsp. *subtilis* subsp. nov. and *Bacillus subtilis* subsp. *spizizenii* subsp. nov. *Int. J. Syst. Bacteriol.* 49(Pt 3):1211–1215.
11. Nishito Y, et al. 2010. Whole genome assembly of a natto production strain *Bacillus subtilis* natto from very short read data. *BMC Genomics* 11:243.
12. Rasko DA, Myers GS, Ravel J. 2005. Visualization of comparative genomic analyses by BLAST score ratio. *BMC Bioinformatics* 6:2.
13. Rissman AI, et al. 2009. Reordering contigs of draft genomes using the Mauve aligner. *Bioinformatics* 25:2071–2073.
14. Roberts MS, Cohan FM. 1995. Recombination and migration rates in natural populations of *Bacillus subtilis* and *Bacillus mojavensis*. *Evolution* 49:1081–1094.
15. Roberts MS, Nakamura LK, Cohan FM. 1994. *Bacillus mojavensis* sp. nov., distinguishable from *Bacillus subtilis* by sexual isolation, divergence in DNA sequence, and differences in fatty acid composition. *Int. J. Syst. Bacteriol.* 44:256–264.
16. Roberts MS, Nakamura LK, Cohan FM. 1996. *Bacillus vallismortis* sp. nov., a close relative of *Bacillus subtilis*, isolated from soil in Death Valley, California. *Int. J. Syst. Bacteriol.* 46:470–475.
17. Srivatsan A, et al. 2008. High-precision, whole-genome sequencing of laboratory strains facilitates genetic studies. *PLoS Genet.* 4:e1000139.
18. Zeigler DR. 2011. The genome sequence of *Bacillus subtilis* subsp. *spizizenii* W23: insights into speciation within the *B. subtilis* complex and into the history of *B. subtilis* genetics. *Microbiology* 157(Pt 7): 2033–2041.