JB Accepts, published online ahead of print on 23 December 2010
J. Bacteriol. doi:10.1128/JB.01351-10

Genome sequence of *Bacillus subtilis* subsp. *spizizenii* gtP20b isolated from the Indian Ocean

Longjiang Fan[1], Shiping Bo[1], Huan Chen[1], Wanzhi Ye[2], Katrin Kleinschmidt[3], Heike I. Baumann[3], Johannes F. Imhoff[3], Michael Kleine[4] and Daguang Cai[2]*

1. James D. Watson Institute of Genome Sciences & Institute of Crop Science, Zhejiang University, Hangzhou 310029, China
2. Molecular Phytopathology, Christian-Albrechts-Universität zu Kiel, D-24118 Kiel, Germany
3. Marine Microbiology, Leibniz Institute of Marine Science (IFM-GEOMAR), Düsternbrooker Weg 20, D-24105 Kiel, Germany
4. Planton GmbH, Am Kiel-Kanal 44, D-24106 Kiel, Germany

Correspondence to Prof. Dr. Daguang Cai
Phone: +49-431-8803215
Fax: +49-431-8801583
E-mail: dcai@phytomed.uni-kiel.de
Address: Department of Molecular Phytopathology, Christian-Albrechts-University of Kiel, Hermann, Rodewald Str. 9, D-24118 Kiel, Germany

Key words: Genome sequencing, *Bacillus subtilis*, marine, antimicrobial peptides, marine isolate

Abstract

Bacillus subtilis is a model organism of aerobic spore-forming Gram-positive bacteria and is of great industrial significance as the source of diverse novel functional molecules. Here we present to our knowledge the first genome sequence of a *Bacillus subtilis* strain gtP20b isolated from the marine environment. A subset of candidate genes and gene clusters were identified, which are potentially involved in production of diverse functional molecules like novel ribosomal and non-ribosomal antimicrobial peptides. The genome sequence described in this paper is due to its high strain-specificity of great importance for basic as well as applied researches on marine organisms.

Results and discussion

Bacillus subtilis is a member of the Gram-positive bacteria of the genus *Bacillus* and has been used as a model organism to investigate differentiation, gene/protein regulation and cell cycle events in bacteria for more than a century (1, 2). *B. subtilis* has industrial importance e.g. as a source for diverse novel functional molecules like antimicrobial peptides (2, 6). Members of the genus are ubiquitous in nature. Various strains of *B. subtilis* have been isolated from diverse habitats including seawater (4). The first *B. subtilis* genome was sequenced a decade ago (6) and updated recently (2, 9). Although draft genomes of 4 further strains were also released (8, 9), no genome of *B. subtilis* strain from a marine habitat has been decoded.

The *B. subtilis* subsp. *spizizenii* strain gtP20b was isolated from the sediment in 608 m water depth in the Indian Ocean and from the layer close to the ocean bottom surface. The sampling was taken through a multi corer during the cruise of the research ship Sonne at the expedition 130 in 1998 and stored at -20℃. Raw reads of the strain genome were generated by using Illumina GA (Solexa) and assembled with Velvet program (14). Based on the reference genome of *B. subtilis* strain168 (6, 7) a draft genome of gtP20b was completed. By subsequent PCR and re-sequencing 100 genome gaps were closed, but remaining four gaps and 23 unmatched short contigs (> 200 bp) with an accumulative length of 151.5 Kb, which are believed to be genome-specific and distributed in the remaining gaps.

The genome sequence of gtP20b comprises 4,247,908 bases with a G+C content of 44.8%, and covers more than 99% of the whole genome (2, 8). It contains 4.331 open reading frames (ORFs), 77 tRNAs including one pseudogene and 30 rRNAs (3). Phylogenetic analysis revealed that gtP20b is closely clustered with *B. subtilis*168 and *B. natto*, but phylogenetically apart from *B. amyloliquefaciens* and *B. licheniformis* (10). Furthermore, 81.7% of the ORFs have orthologs in the strain168 (BLASTP <1e-5), but 444 ORFs were not found in the released genomes of *Bacillus* genus, of these, 392 ORFs did not give hits in current public databases.

At least 59 genes were found to be potentially involved in secondary metabolism. They form diverse gene clusters with varied degree of synteny to other *B. subtilis* strains. A set of hits was retrieved from antimicrobial peptides (AMPs) databases (5, 11, 12, 13) including subtilisin A (*sboA*), surfactin (*sfp*), beta-lactamase precursor (*penP*) and replicative DNA helicase (*dnaC*) etc. However, they showed strong variations at both DNA- and amino acid level when compared with those of other *B. subtilis* strains, suggesting the potential of the strain gtP20b as a unique source for novel AMPs. This genome sequence is due to its high strain-specificity of great importance for both of basic and applied researches.

Nucleotide sequence accession numbers

This Whole Genome Shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession AEHM00000000. The version described here is the first version under the accession number: AEHM01000000.

Acknowledgements

This project was supported by the Bundesministerium für Bildung und Forschung (BMBF), Germany (grant number 0315231A, B) and the Minsterium für Wissenschaft, Wirtschaft und Verkehr des Landes Schleswig-Holstein (grant number 122-08-002). Authors thanks DAAD (grant number D/08/01773, 4) and China Scholarship Council (grant number A/10/00701) for providing the scholarship reward as well as international exchange grants. Authors thank Jun Wang for his help in Solexa sequencing and Ms. Katharina Peetz for her technical support.

Reference

1. **Alcaraz, L. D., G. Moreno-Hagelsieb, L. Eguiarte, V. Souza, L. Herrera-Estrella, and G. Olmedo.** 2010. Understanding the evolutionary relationships and major traits of *Bacillus* through comparative. BMC Genomics **11:**332.
2. **Barbe, V., S. Cruveiller, F. Kunst, P. Lenoble, G. Meurice, A. Sekowska, D. Vallenet, T. Z. Wang, I. Moszer, C. Medigue, and A. Danchin.** 2009. From a consortium sequence to a unified sequence: the *Bacillus subtilis* 168 reference genome a decade later. Microbiol. **155:**1758-1775.
3. **Delcher, A.L., K. A. Bratke, E. C. Powers, and S. L. Salzberg.** 2007. Identifying bacterial genes and endosymbiont DNA with Glimmer. Bioinformatics **23:**673-679.
4. **Ettoumi, B., N. Raddadi, S. Borin, D. Daffonchio, A. Boudabous, and A. Cherif.** 2009. Diversity and phylogeny of culturable spore-forming *Bacilli* isolated from marine sediments. J. Basic Microbiol. **49:**S13-S23.
5. **Gueguen, Y., J. Garnier, L. Robert, M. P. Lefranc, I. Mougenot, J. de Lorgeril, M. Janech, P. S. Gross, G. W. Warr, B. Cuthbertson, M. A. Barracco, P. Bulet, A. Aumelas, Y. Yang, D. Bo, J. Xiang, A. Tassanakajon, D. Piquemal, and E. Bachère.** 2006. PenBase, the shrimp antimicrobial peptide penaeidin database: Sequence-based classification and recommended nomenclature. Dev. Comp. Immunol. **30:**283-288.
6. **Kunst, F., N. Ogasawara, I. Moszer, A. M. Albertini et al.** 1997. The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*. Nature **390:**249-256.
7. **Kurtz, S., A. Phillippy, A. L. Delcher, M. Smoot, M. Shumway, C. Antonescu, and S. L. Salzberg.** 2004. Versatile and open software for comparing large genomes. Genome Biol. **5**:9.
8. **Nishito, Y., Y. Osana, T. Hachiya, K. Popendorf, A. Toyoda, A. Fujiyama, M. Itaya, and Y. Sakakibara.** 2010. Whole genome assembly of a natto production strain *Bacillus subtilis* natto from very short read data. BMC Genomics **11:**12.
9. **Srivatsan, A., Y. Han, J. Peng, A. K. Tehranchi, R. Gibbs, J. D. Wang, and R. Chen.** 2008. High-precision, whole-genome sequencing of laboratory strains facilitates genetic studies. PLoS Gentics. **4(8):** e1000139. doi:10.1371/journal.pgen.1000139.
10. **Tamura, K., J. Dudley, M. Nei, and S. Kumar.** 2007. MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0. Mol. Biol. Evol. **24:**1596-1599.

11. Thomas, S., S. Karnik, R. S. Barai, V. K. Jayaraman, and S. Idicula-Thomas. 2010. CAMP: a useful resource for research on antimicrobial peptides. Nucleic Acids Res. **38:**D774-D780.

12. Wang, G. S., X. Li, and Z. Wang. 2009. APD2: the updated antimicrobial peptide database and its application in peptide design. Nucleic Acids Res. **37:**D933-D937.

13. Wang, C. K.L., Q. Kaas, L. Chiche, and D. J. Craik. 2008. CyBase: a database of cyclic protein sequences and structures, with applications in protein discovery and engineering. Nucleic Acids Res. **36:**D206-D210.

14. Zerbino, D. R., and E. Birney. 2008. Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. **18:**821-829.