

Draft Genome Sequence of *Bacillus simplex* DSM 1321 for Setting Up Phylogenomics in Genomic Taxonomy of the *Bacillus*-Like Bacteria

Guo-hong Liu, Bo Liu, Jie-ping Wang, Jian-mei Che, Qian-qian Chen, Zheng Chen

Agricultural Bio-Resources Research Institute, Fujian Academy of Agricultural Sciences, Fuzhou, China

***Bacillus simplex* DSM 1321 is a Gram-positive, spore-forming, and aerobic bacterium. Here, we report the draft genome sequence of *B. simplex* DSM 1321, with 6,494,937 bp, which will provide useful information for setting up phylogenomics in genomic taxonomy of the *Bacillus*-like bacteria as well as for the functional gene mining and application of *B. simplex* DSM 1321.**

Received 4 May 2016 Accepted 6 May 2016 Published 23 June 2016

Citation Liu G-H, Liu B, Wang J-P, Che J-M, Chen Q-Q, Chen Z. 2016. Draft genome sequence of *Bacillus simplex* DSM 1321 for setting up phylogenomics in genomic taxonomy of the *Bacillus*-like bacteria. *Genome Announc* 4(3):e00574-16. doi:10.1128/genomeA.00574-16.

Copyright © 2016 Liu 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 Bo Liu, fzliubo@163.com

Bacillus simplex DSM 1321 is widely spread in the soil and has high similarity with *Bacillus muralis* DSM 16288^T. With decreases in the cost of genomic sequencing, it has been proposed that whole-genome sequencing information be combined with the main phenotypic characteristics as a polyphasic approach strategy (taxonogenomics) to describe new bacterial taxa (1–4). In this study, a high-quality genome sequence of *B. simplex* DSM 1321 was sequenced, which would promote research on the genomic taxonomy of the *Bacillus*-like bacteria.

The genome of *B. simplex* DSM 1321 was sequenced with massively parallel sequencing (MPS) Illumina technology. Two DNA libraries were constructed: a paired-end library with an insert size of 500 bp and a mate-pair library with an insert size of 5 kb. The 500-bp library and the 5-kb library were sequenced using an Illumina HiSeq 2500 using a PE125 strategy. Library construction and sequencing were performed at the Beijing Novogene Bioinformatics Technology Co., Ltd. Quality control of both paired-end and mate-pair reads was performed using an in-house program. After this step, Illumina PCR adapter reads and low-quality reads were filtered. The filtered reads were assembled by SOAPdenovo (5, 6) to generate scaffolds. All reads were used for further gap closure. Through the data assembly, 6,494,937 bp within 23 scaffolds were obtained, and the scaffold N_{50} was 1,760,672 bp. The average length of the scaffolds was 1,585,992 bp, and the longest and shortest scaffolds were 3,171,336 bp and 647 bp, respectively.

Gene prediction was performed on the *B. simplex* DSM 1321 genome assembly by GeneMarkS (7). Transfer RNA (tRNA) genes were predicted with tRNAscan-SE (8), ribosomal RNA (rRNA) genes were predicted with RNAmmer (9), and small RNAs (sRNAs) were predicted by BLAST against the Rfam (10) database. PHAST (11) is used for prophage prediction, and CRISPR-Finder (12) is used for clustered regularly interspaced short palindromic repeat (CRISPR) identification. A total of 6,932 genes were predicted, including 6,813 coding sequences (CDSs), 5 sRNAs, 81 tRNAs, and 33 rRNA genes (14 5S rRNAs, 8 16S rRNAs, and 11 23S rRNAs). Also, 9 prophage and 3 CRISPR arrays were

found in the draft genome. The average DNA G+C content was 40.11%.

Nucleotide sequence accession numbers. This whole-genome shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession no. [LWJJ000000000](https://www.ncbi.nlm.nih.gov/nuccore/LWJJ000000000). The version described in this paper is version LWJJ000000000.1.

ACKNOWLEDGMENTS

This work was financially supported by the National Natural Science Foundation of China (grant no. 31370059), the Scientific Research Foundation for Returned Scholars, Fujian Academy of Agricultural Sciences (grant no. YJRC2014-1), the Fujian key science and technology special projects-key agricultural science and technology special project (grant 2015NZ0003-1), and Seed industry innovation project of Fujian Province—Fujian Resource Preservation Center of the *Bacillus*-like Bacteria in the Seed industry innovation and industrialization of project of Fujian Province (FJZZZY-1544).

FUNDING INFORMATION

This work, including the efforts of Bo Liu, was funded by National Natural Science Foundation of China (NSFC) (31370059).

REFERENCES

- Ramasamy D, Mishra AK, Lagier JC, Padhmanabhan R, Rossi M, Sentausta E, Raoult D, Fournier PE. 2014. A polyphasic strategy incorporating genomic data for the taxonomic description of novel bacterial species. *Int J Syst Evol Microbiol* 64:384–391. <http://dx.doi.org/10.1099/ijs.0.057091-0>.
- Keita MB, Diene SM, Robert C, Raoult D, Fournier PE, Bittar F. 2013. Non-contiguous finished genome sequence and description of *Bacillus massiliogorillae* sp. nov. *Stand Genomic Sci* 9:93–105. <http://dx.doi.org/10.4056/sigs.4388124>.
- Mishra AK, Pfeleiderer A, Lagier JC, Robert C, Raoult D, Fournier PE. 2013. Non-contiguous finished genome sequence and description of *Bacillus massilianorexius* sp. nov. *Stand Genomic Sci* 8:465–479. <http://dx.doi.org/10.4056/sigs.4087826>.
- Mishra AK, Lagier JC, Rivet R, Raoult D, Fournier PE. 2012. Non-contiguous finished genome sequence and description of *Paenibacillus senegalensis* sp. nov. *Stand Genomic Sci* 7:70–81. <http://dx.doi.org/10.4056/sigs.3056450>.
- Li R, Zhu H, Ruan J, Qian W, Fang X, Shi Z, Li Y, Li S, Shan G,

- Kristiansen K, Li S, Yang H, Wang J, Wang J. 2010. *De novo* assembly of human genomes with massively parallel short read sequencing. *Genome Res* 20:265–272. <http://dx.doi.org/10.1101/gr.097261.109>.
6. Li R, Li Y, Kristiansen K, Wang J. 2008. SOAP: short oligonucleotide alignment program. *Bioinformatics* 24:713–714. <http://dx.doi.org/10.1093/bioinformatics/btn025>.
 7. Besemer J, Lomsadze A, Borodovsky M. 2001. GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. *Nucleic Acids Res* 29:2607–2618. <http://dx.doi.org/10.1093/nar/29.12.2607>.
 8. Lowe TM, Eddy SR. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* 25: 0955–0964. <http://dx.doi.org/10.1093/nar/25.5.0955>.
 9. Lagesen K, Hallin P, Rødland EA, Staerfeldt H-H, 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>.
 10. Gardner PP, Daub J, Tate JG, Nawrocki EP, Kolbe DL, Lindgreen S, Wilkinson AC, Finn RD, Griffiths-Jones S, Eddy SR, Bateman A. 2009. Rfam: updates to the RNA families database. *Nucleic Acids Res* 37(Suppl 1):D136–D140. <http://dx.doi.org/10.1093/nar/gkn766>.
 11. Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS. 2011. PHAST: a fast phage search tool[J]. *Nucleic Acids Res* 39(Suppl 2):W347–W352. <http://dx.doi.org/10.1093/nar/gkr485>.
 12. Grissa I, Vergnaud G, Pourcel C. 2007. CRISPRFinder: a Web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Res* 35(Suppl 2):W52–W57. <http://dx.doi.org/10.1093/nar/gkm360>.