

# Draft Genome Sequence of *Bacillus* sp. FJAT-27238 for Setting up Phylogenomic Analysis of Genomic Taxonomy of *Bacillus*-Like Bacteria

Guo-hong Liu, Bo Liu, Jie-ping Wang, Jian-Mei Che, Qian-Qian Chen, Yu-Jing Zhu

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

***Bacillus* sp. FJAT-27238 is a Gram-positive, spore-forming, and aerobic bacterium. Here, we report the draft genome sequence of *Bacillus* sp. FJAT-27238, with 6,134,829 bp, which will provide useful information for setting up phylogenomic analysis of the genomic taxonomy of *Bacillus*-like bacteria as well as for the functional gene mining and application of strain FJAT-27238. The genomic DNA G+C content was 47.37%.**

Received 28 July 2015 Accepted 28 July 2015 Published 24 September 2015

**Citation** Liu G-H, Liu B, Wang JP, Che JM, Chen Q-Q, Zhu Y-J. 2015. Draft genome sequence of *Bacillus* sp. FJAT-27238 for setting up phylogenomic analysis of genomic taxonomy of *Bacillus*-like bacteria. *Genome Announc* 3(5):e00985-15. doi:10.1128/genomeA.00985-15.

**Copyright** © 2015 Liu 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/4.0/).

Address correspondence to Bo Liu, fzliubo@163.com.

During the course of a study on the diversity of *Bacillus* species from Wudalianchi in Heilongjiang Province of China, strain FJAT-27238, a potentially novel species of the genus according to results of 16S rRNA phylogenetic analysis, was isolated. Recently, it has been proposed that whole-genome sequencing information be combined with the main phenotypic characteristics as a polyphasic approach strategy (taxono-genomics) to describe new bacterial taxa (1–4). In this study, a high-quality genome sequence of FJAT-27238 was sequenced that may confirm that this is a new species and also promote new research regarding the genomic taxonomy of *Bacillus*-like bacteria.

The genome of FJAT-27238 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 with a paired-end (PE) 125 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 were 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 SOAP *de novo* (5, 6) to generate scaffolds. All reads were used for further gap closure. Through the data assembly, 6,134,829 bp within 1 scaffold were obtained, and the scaffold  $N_{50}$  was 6,134,829 bp. Ninety-five percent of the clean reads were aligned back to the genome, suggesting good quality of the assembly.

Gene prediction was performed on the FJAT-27238 genome assembly by GeneMarkS (7). tRNA genes were predicted with tRNAscan-SE (8), rRNA genes were predicted with rRNAmmer (9), and sRNAs were predicted by BLAST against the Rfam (10) database. PHAST (11) was used for prophage prediction and CRISPRfinder (12) was used for CRISPR identification. A total of 5,917 genes were predicted, including 5,583 protein-coding genes, 8 small RNA (sRNA) genes, 125 tRNA genes, and 40 rRNA genes

(13 5S rRNAs, 14 16S rRNAs, and 13 23S rRNAs). Also, 6 prophage and 1 CRISPR array were found in the draft genome. The genomic DNA G+C content was 47.37%.

From the genome sequence analysis, FJAT-27238 was predicted to possess complete carbon metabolic pathways, including those for glycolysis and fatty acid biosynthesis and the pentose phosphate pathway. In particular, FJAT-27238 was predicted to be equipped with a wide variety of genes for amino acid and organic acid metabolism, in accordance with the physiological properties of FJAT-27238.

**Nucleotide sequence accession numbers.** This whole-genome shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession no. [LGJF00000000](https://www.ncbi.nlm.nih.gov/nuclseq/LGJF00000000). The version described in this paper is version LGJF00000000.1.

## 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), and the doctor fund of the Fujian Academy of Agricultural Sciences (grant no. 2014BS-3).

## REFERENCES

- Ramasamy D, Mishra AK, Lagier JC, Padhmanabhan R, Rossi M, Sentausa 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. Noncontiguous 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, Pfeiderer A, Lagier JC, Robert C, Raoult D, Fournier PE. 2013. Noncontiguous 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. Noncon

- tiguous finished genome sequence and description of *Paenibacillus senegalensis* sp. nov. *Stand Genomic Sci* 7:70–81. <http://dx.doi.org/10.4056/sigs.3056450>.
5. 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* 25:955–964. <http://dx.doi.org/10.1093/nar/25.5.0955>.
  9. 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>.
  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: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. *Nucleic Acids Res* 39: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:W52–W57. <http://dx.doi.org/10.1093/nar/gkm360>.