Data Article

# Draft genome sequence data of *Bacillus subtilis* strain *9407*, isolated from healthy apples in China

Qingchao Zeng [a], Jianbo Xie [b], Yan Li [a], Xiaofei Gu [a], Qi Wang [a, *]

[a] College of Plant Protection, Key Laboratory of Plant Pathology, Ministry of Agriculture, China Agricultural University, Beijing, China
[b] College of Biological Sciences and Technology, Key Laboratory of Genetics and Breeding in Forest Trees and Ornamental Plants, Ministry of Education, Beijing Forestry University, Beijing, China

### ARTICLE INFO

### ABSTRACT

*Bacillus subtilis* strain 9407 is an endophyte which was isolated from healthy apples from an infested orchard that exhibits strong inhibitory activity against *Botryosphaeria dothidea*. Whole-genome sequencing of *B. subtilis* 9407 was performed using the Illumina Hiseq platforms. Here, we report the draft genome sequence of *B. subtilis* strain 9407 containing 16 scaffolds (4,062,615 bp), 4033 coding sequences, and an average 43.66% G + C content. The genome contains genes responsible for the production of several bioactive secondary metabolites, including the lipopeptides fengycin and surfactin. The genome information will provide fundamental knowledge of the organism. This whole-genome shotgun data has been deposited at DDBJ/EMBL/GenBank under the accession numbers PISO00000000.

---

* Corresponding author.
  *E-mail address:* wangqi@cau.edu.cn (Q. Wang).

Specifications Table

| Subject | Biology |
| --- | --- |
| Specific subject area | Microbiology and genomics |
| Type of data | Draft genome sequence data in FASTA format, figure and image |
| How data were acquired | Genome sequencing with Illumina Hiseq at Beijing Berrygenomics Bioinformatics Technology Co., Ltd. |
| Data format | Analyzed and assembled genome sequences |
| Parameters for data collection | Genomic DNA was extracted from a pure culture of B. subtilis 9407 |
| Description of data collection | Whole-genome sequencing, assembly, and annotation |
| Data source location | B. subtilis 9407 was isolated from healthy apple in China |
| Data accessibility | This whole-genome shotgun data has been deposited at DDBJ/EMBL/GenBank under the accession numbers PISO00000000 (https://www.ncbi.nlm.nih.gov/nuccore/PISO00000000). The sequence data have been registered in the NCBI SRA database under the accession number SSR8935609 (https://www.ncbi.nlm.nih.gov/sra/?term&equals;SRR8935609) and SRR8935610 (https://www.ncbi.nlm.nih.gov/sra/?term&equals;SRR8935610). |

**Value of the Data**
- The draft genome sequence of B. subtilis 9407 provides fundamental knowledge of this organism and insight for biotechnological application in agriculture.
- The draft genome sequence data are useful for comparative genomic analysis of Bacillus species.
- It will accelerate functional genomics research.

## 1. Data

Plant pathogens are one of the major challenges for sustainable food production and ecosystem stability worldwide [1]. Producers have been relying heavily on agrochemicals because they are a reliable and economical method to protect crops. However, the heavy use of chemicals has negatively affected the environment. Biocontrol formulations, as an alternative to chemical pesticides, have shown good application prospects in sustainable agriculture [2,3]. Bacillus spp. are biocontrol agents that have been extensively studied owing to their strong stress resistance and outstanding environmental adaptability. Many Bacillus strains are known to produce a diverse spectrum of secondary metabolites with antimicrobial activity [4,5]. Bacillus subtilis is a model Gram-positive bacterium that has been widely used for plant disease control for several decades [6]. B. subtilis is widespread in nature. The species neither causes disease in humans or animals nor pollutes the environment; in addition, it can produce a variety of antibiotics and inhibitory enzymes, such as iturin, bacilysin, and fengycin, with broad-spectrum antimicrobial/antifungal activity [6].

B. subtilis 9407 was isolated from healthy apples from an infested orchard in China. Our previous result showed that strain 9407 exhibits strong antagonistic activity toward apple ring rot and 18 other pathogens that cause plant diseases in apples, grapes, pears and other plants [7]. Here, we report the draft genome sequence of strain 9407 to facilitate further studies on pathways related to biocontrol and broaden the current understanding of the biological control mechanism of B. subtilis 9407. The draft genome of the strain 9407 was obtained by direct sequencing of genomic DNA using Illumina sequencing technology. After quality filtering, we obtained 6,963,036 (1.3 Gb for short insert) and 8,288,121 (1.54 Gb for long insert) paired-end reads.

The draft genome of B. subtilis 9407 contains 16 scaffolds with a total length of 4,062,615 bp with an $N_{50}$ of 2,111,374 bp. The genome contains 4033 protein-coding genes (CDSs), 8 rRNA genes, and 62 tRNAs genes, with a G + C content of 43.65%. The protein-coding genes have an average length of 884 bp and account for 89.21% of the genome sequence. Some of these predicted genes and functions, such as those related to chemotaxis-related proteins, antimicrobial biosynthesis, biofilm formation, are known to be associated with growth promotion and biocontrol. The average nucleotide identity value between strain 9407 and 168[T] was obtained by the JSpecies [8] software to exceed 98%.

Subsequent analyses of the genome content of B. subtilis 9407 and its comparison with phylogenetically related strains will help to determine key aspects of its interaction with the plants.

## 2. Experimental design, materials, and methods

*B. subtilis* 9407 was cultivated as previously described [7]. A single colony of *B. subtilis* 9407 was cultured in Lurina-Bertani medium at 37 °C overnight. Then, the bacterial cells were centrifuged and pelleted for DNA extraction. The genomic DNA of *B. subtilis* 9407 was extracted using the phenol-chloroform method [9] and sequenced at Beijing Berrygenomics Bioinformatics Technology Co., Ltd. Genomic DNA was used for construction of 350 bp and 2 kb libraries, using the TruSeq Nano DNA Sample Preparation Kit and Mate Pair Library Prep Kit v2, respectively. The manufacturer's instructions were followed in both cases. The libraries were sequenced on Illumina Hiseq 2500 in 2 × 100 paired-end format which achieved approximately 741x coverage. Adapters and low-quality sequences were removed using the software of Cutadapt and Sickle, respectively [10,11]. Assembly of filtered Illumina reads was done using the software SOAP*denovo* version 2.04 which the kmer value was 83 [12]. Gaps inside scaffolds were closed by using GapCloser version 1.12 [12]. Annotation was performed using PROKKA software [13].

## Acknowledgments

## Conflict of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.dib.2020.105143.

## References

[1] L. Wu, et al., Novel routes for improving biocontrol activity of Bacillus based bioinoculants, Front. Microbiol. 6 (1395) (2015).
[2] G. Liu, et al., Whole-genome sequencing of Bacillus velezensis LS69, a strain with a broad inhibitory spectrum against pathogenic bacteria, J. Biotechnol. 249 (2017) 20–24.
[3] S.Y. Kim, et al., Complete genome sequence of Bacillus velezensis M75, a biocontrol agent against fungal plant pathogens, isolated from cotton waste, J. Biotechnol. 241 (2017) 112–115.
[4] S.P. Chowdhury, et al., Biocontrol mechanism by root-associated Bacillus amyloliquefaciens FZB42-a review, Front. Microbiol. 6 (780) (2015).
[5] H. Pan, Q. Li, J. Hu, The complete genome sequence of Bacillus velezensis 9912D reveals its biocontrol mechanism as a novel commercial biological fungicide agent, J. Biotechnol. 247 (2017) 25–28.
[6] H. Liu, et al., Complete genome sequence of Bacillus subtilis BSD-2, a microbial germicide isolated from cultivated cotton, J. Biotechnol. 230 (2016) 26–27.
[7] H. Fan, et al., Fengycin produced by Bacillus subtilis 9407 plays a major role in the biocontrol of apple ring rot disease, Microbiol. Res. 199 (2017) 89–97.
[8] M. Richter, R. Rossello-Mora, Shifting the genomic gold standard for the prokaryotic species definition, Proc. Natl. Acad. Sci. U.S.A. 106 (45) (2009) 19126–19131.
[9] G. Zhang, et al., A mimicking-of-DNA-methylation-patterns pipeline for overcoming the restriction barrier of bacteria, PLoS Genet. 8 (e10029879) (2012).
[10] M. Martin, Cutadapt removes adapter sequences from highthroughput sequencing reads, EMBnet Journal 17 (2011) 10–12.
[11] N.A. Joshi, J.N. Fass, Sickle—a Sliding-Window, Adaptive, Qualitybased Trimming Tool for FASTQ Files, 2011.
[12] R. Luo, et al., SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler, GigaScience 1 (18) (2012).
[13] T. Seemann, Prokka: rapid prokaryotic genome annotation, Bioinformatics 30 (14) (2014) 2068–2069.