Data Article

Data on draft genome sequence of *Bacillus* sp. strain AN2 isolated from agricultural soil in Brazil

ABSTRACT

Aims: The aim of this research is to report the genome sequence of the *Bacillus* sp. strain AN2, isolated from agricultural soil from Rio de Janeiro, Brazil.
Study design: This study is designed to assess novel microbial genes with biotechnology potential.
Place and Duration of Study: Laboratory of Environmental Biotechnology, Western Rio Janeiro State University in Brazil, between January 2019 and December 2019.
Methodology: High-quality genomic DNA was extracted using a GenElute Bacterial Genomic DNA kit. The Nextera XT DNA Library Prep Kit was used for genomic library construction. Paired-end sequence reads were generated by an Illumina MiSeq instrument with the 600 cycles MiSeq Reagent Kit v3. Sequence data were assembled with A5-MiSeq pipeline software and the contigs were annotated by Rapid Annotation using Subsystems Technology (RAST).
Results: The obtained genome sequence of *Bacillus* AN2 included 21 contigs with a calculated size of 3,681,081 bp in length. The G + C content for the draft genome is 41.4%. A total of 3824 coding sequences (CDS) were predicted and encoded at least 88 tRNAs. The strain AN2 possesses numerous genes involved in deconstructing plant cell walls to improve the efficiency of processing biomass from agriculture residues by enzymatic hydrolysis such as cellulases, xylanases, and amylases. Many genes responsible for the bacterial strain resistance to several antibiotics and various toxic compounds were also identified.
Conclusion: From this study it can be deduced that the novel strain has a high capacity for biocontrol against soilborne pathogens and bioconversion of agro-industrial residues. The data represent the first characterization of genome sequence data of a newly isolated *Bacillus* strain and it may holds a great promise to improve crop yield and biofuel industry.

Keywords: Biofertilizers; Biotechnology; Inoculants; Soil microbiota.

DATA DESCRIPTION

With the aim of revealing novel genes from *Bacillus* species, some bacterial strains were isolated by mixing 1 g of soil with 9 ml of NaCl 0.8% and incubating at 30°C for 1 h. The solution was decanted, and 1 ml of the supernatant was spread on MEMB agar plates [1] and incubated at 30°C for 4 days under aerobic conditions. One of the colonies recovered was named AN2, and its genome was sequenced. The genomic DNA was purified using a GenElute Bacterial Genomic DNA kit (Sigma-Aldrich, USA), and the library was constructed using a Nextera XT kit (Illumina). Paired-end sequence reads were generated by an Illumina MiSeq instrument with the 600 cycles MiSeq Reagent Kit v3. Reads were filtered for a Phred quality score of at least 20 and were assembled with A5-MiSeq pipeline software [2]. The

contigs were annotated using the NCBI Prokaryotic Genomes Automatic Annotation Pipeline (PGAAP) and Rapid Annotation using Subsystems Technology (RAST) [3]. *Bacillus* species identification was achieved by using the JSpeciesWS server online with average nucleotide identity (ANIb) and MUMmer average nucleotide identity (ANIm) analysis [4]. The AN2 strain was identified as a member of *Bacillus* group. The draft genome of *Bacillus* sp. AN2 consists of 3,681,081 bp distributed in 21 contigs, with an average GC content of 41.4%. The draft genome comprises 88 tRNAs. The bioinformatic analyses indicates that this strain carries many antimicrobial resistance genes and also most of the key genes associated with deconstructing plant cell walls to improve the efficiency of processing biomass from agriculture residues by enzymatic hydrolysis such as cellulases, xylanases, and amylases have been identified.

Comment [IH1]: Please add figure of Heatmap (for exemple) indicating the OrthoANI values of your Bacillus sp. Strain and closely related Bacillus species.

Comment [IH2]: You may add a table describing a genome statistics of your Bacillus sp. strain

Comment [IH3]: Could you please add a diagram showng number of genes encoding for protein involved in degradation of cell walls.

Limitations

Current data is based on the draft level genome such that the exact length of the genome, and the number of rRNA genes and repetitive elements, cannot be absolutely determined.

Availability of data

The data described in this Data Article can be freely and openly accessed on DDBJ/ENA/GenBank under the accession WUQP00000000. The version described in this paper is version WUQP01000000.

REFERENCES

1. Chong NM. 2005. Development of a tool for measuring the degradation capacity of microorganisms for a xenobiotic. Enzyme and Microbial Technology. 37:467-471.

2. Coil D, Jospin G, Darling AE. 2015. A5-miseq: an updated pipeline to assemble microbial genomes from Illumina MiSeq data. Bioinformatics 31:587-589.

3. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O. 2008. The RAST server: rapid annotations using subsystems technology. BMC Genomics 9:75.

4. Richter M, Rosselló-Móra R, Glöckner FO, Peplies J. 2016. *JSpeciesWS: a Web server for prokaryotic species circumscription based on pairwise genome comparison.* Bioinformatics *32:929–931. doi:10.1093/bioinformatics/btv681*