DATA NOTE

Draft genome sequence of *Bacillus* sp. EE-W1 isolated from a biogas reactor

Waleed S. Mohammed[1], Elvira E. Ziganshina[2], Ramilya F. Akhmetova[2] and Ayrat M. Ziganshin[2*]

Abstract

Objectives: The genus *Bacillus* comprises spore-forming rod-shaped Gram-positive bacteria, which usually grow aerobically or anaerobically. Members of this genus are common environmental microorganisms. Also, they can be monitored in the food production chain. Genome sequence of *Bacillus* sp. strain EE-W1 will provide helpful information to understand its ecology and genetics. Draft genome data may be useful in the field of using *Bacillus* species in industrial biotechnology. Also, these data can be a useful resource for the study of comparative genomics.

Data description: Here, we present the draft genome sequence of *Bacillus* sp. strain EE-W1 isolated from a biogas reactor, Kazan, Russia. The assembled genome size was 5,769,164 bp, with a GC content 35.1%. This draft genome data can be accessed at DDBJ/ENA/GenBank under the accession WIPE00000000.

Keywords: Draft genome, *Firmicutes*, *Bacillus*, Biogas reactor, Illumina sequencing

Objective

The genus *Bacillus* includes spore-forming rod-shaped Gram-positive aerobic or facultative anaerobic bacteria. A distinctive feature of the genus *Bacillus* is its ability to form endospores in response to various environmental and nutritional stresses. Members of this genus are common environmental microorganisms. Also, they can be monitored in the food production chain [1]. Moreover, some species of *Bacillus* with high proteolytic activity have demonstrated a sustained survival in some protein-fed anaerobic digesters, what finally led to an improved biogas productivity [2]. However, the *Bacillus* group also includes pathogenic species, therefore, it is necessary to develop proper hygiene and sanitation procedures to decrease the potential risk of transmission of the infectious disease [3]. Genome sequence of *Bacillus* sp. strain EE-W1 will provide helpful information to understand its ecology and genetics. Draft genome data can be also

useful in the field of using *Bacillus* species in industrial biotechnology. Also, these data can be a useful resource for the study of comparative genomics.

Data description

The *Bacillus* sp. EE-W1 was isolated from a laboratory-scale mesophilic biogas reactor. Briefly, 1 g (wet weight) of digested wastes was consistently diluted with sterile tap water. 100 µL of each obtained dilution was spread onto LB agar plates and then cultured at +37 °C. Finally, single colonies were transferred to new plates, and the cultivation process was repeated until pure cultures were received. For DNA extraction, the bacterial strain *Bacillus* sp. EE-W1 was cultured on LB agar at +37 °C for 2 days. Genomic DNA from the *Bacillus* sp. EE-W1 was then extracted using a commercially available DNA extraction kit, FastDNA spin kit (#116540600; MP Biomedicals, USA), and assessed by spectrophotometric analysis as reported previously [4, 5]. The traditional identification of the strain EE-W1 was carried out using morphological characteristics and biochemical tests followed by the PCR-amplified 16S rRNA gene sequence analysis (Table 1) at the EzBioCloud Database [6]. The phylogenetic tree was then

*Correspondence: ayrat.ziganshin@kpfu.ru
[2] Department of Microbiology, Kazan (Volga Region) Federal University, Kremlyovskaya str. 18, Kazan 420008, Russia
Full list of author information is available at the end of the article

Mohammed *et al. BMC Res Notes*   (2020) 13:152

Page 2 of 3

Table 1 Overview of data files/data sets

Label	Name of data file/data set	File types (file extension)	Data repository and identifier (DOI or accession number)
Bacillus sp. EE-W1, whole genome shotgun sequencing project	Whole genome shotgun project	FASTA	Genbank (https://identifiers.org/ncbi/insdc:WIPE00000000.1)
Bacillus sp. (in: Bacteria) strain EE-W1 16S ribosomal RNA gene, partial sequence	16S rRNA gene sequence	FASTA	Genbank (https://identifiers.org/ncbi/insdc:MN611309.1)

constructed with the MEGA 7 software [7]. The genome sequence of *Bacillus* sp. EE-W1 was received by Illumina sequencing as described previously [5]. Sequence read quality was checked using FastQC v0.11.5 [8] and trimmed with Trimmomatic version 0.36 [9]. The filtered reads were assembled into contigs with minimum contig length 500 bp using de novo assembler Velvet version 1.2.10 [10]. Contigs were ordered using progressive Mauve version 2.4.0 [11]. The quality of genome assembly was estimated with Quast version 5.0.0 [12]. Gene prediction and annotation was performed using the Rapid Annotation System Technology (RAST) server [13]. The rRNA genes were detected by using Barrnap version 0.9 [14], while tRNA genes were identified with Aragorn version 1.2 [15]. The phylogeny based on the whole genome data (Table 1) was determined by calculating the average nucleotide identity with closely related *Bacillus* species with the Orthologous Average Nucleotide Identity Tool (OrthoANI) [16].

The calculated OrthoANI values with closely related species showed that the strain EE-W1 presented OrthoANI values of 96.33% and 94.79% with *Bacillus wiedmannii* strain FSL W8-0169 and *Bacillus mobilis* strain 0711P9-1 (type strains), respectively. The draft genome assembly of *Bacillus* sp. EE-W1 consisted of 169 contigs larger than 500 bp, with a total size of 5,769,164 bp and mean GC content 35.1%. The obtained genome sequence data was at $117\times$ coverage. The RAST server predicted 6104 coding sequences. The genome of *Bacillus* sp. EE-W1 encodes at least 3 rRNAs and 8 tRNAs. The genome of the strain EE-W1 also harbored genes related to proteins and carbohydrates degradation and fermentation (including mixed acid, lactate fermentation, acetyl-CoA fermentation to butyrate, acetoin, butanediol metabolism, and butanol biosynthesis). In addition, the genome contained various genes related to resistance to several toxic compounds and some antibiotics (such as streptothricin, tetracycline and fosfomycin).

Limitations

Current data is based on the draft level genome sequence, due to which exact length of the genome, synteny, number of rRNA and repetitive elements cannot be certainly reported.

Author details
[1] Department of Biotechnology, Faculty of Agriculture, Al-Azhar University, Cairo 11651, Egypt. [2] Department of Microbiology, Kazan (Volga Region) Federal University, Kremlyovskaya str. 18, Kazan 420008, Russia.

References
1. Checinska A, Paszczynski A, Burbank M. *Bacillus* and other spore-forming genera: variations in responses and mechanisms for survival. Annu Rev Food Sci Technol. 2015;6:351–69.
2. Kovács E, Wirth R, Maróti G, Bagi Z, Nagy K, Minárovits J, Rákhely G, Kovács KL. Augmented biogas production from protein-rich substrates and associated metagenomic changes. Bioresour Technol. 2015;178:254–61.
3. Miller RA, Beno SM, Kent DJ, Carroll LM, Martin NH, Boor KJ, Kovac J. *Bacillus wiedmannii* sp. nov., a psychrotolerant and cytotoxic *Bacillus cereus*

Mohammed *et al. BMC Res Notes*　(2020) 13:152

Page 3 of 3

group species isolated from dairy foods and dairy environments. Int J Syst Evol Microbiol. 2016;66:4744–53.

4. Ziganshina EE, Mohammed WS, Doijad SP, Shagimardanova EI, Gogoleva NE, Ziganshin AM. Draft genome sequence of *Brevibacterium epidermidis* EZ-K02 isolated from nitrocellulose-contaminated wastewater environments. Data Brief. 2018;17:119–23.

5. Ziganshina EE, Mohammed WS, Shagimardanova EI, Shigapova LH, Ziganshin AM. Draft genome sequence of *Bacillus pumilus* strain EZ-C07 isolated from digested agricultural wastes. BMC Res Notes. 2018;11:606.

6. Yoon SH, Ha SM, Kwon S, Lim J, Kim Y, Seo H, Chun J. Introducing EzBio-Cloud: a taxonomically united database of 16S rRNA and whole genome assemblies. Int J Syst Evol Microbiol. 2017;67:1613–7.

7. Kumar S, Stecher G, Tamura K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol. 2016;33:1870–4.

8. Babraham Bioinformatics. FastQC: a quality control tool for high throughput sequence data. Cambridge: Babraham Institute; 2011.

9. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30:2114–20.

10. Zerbino DR. Using the Velvet *de novo* assembler for short-read sequencing technologies. Curr Protoc Bioinform. 2010;11(11):5.

11. Rissman AI, Mau B, Biehl BS, Darling AE, Glasner JD, Perna NT. Reordering contigs of draft genomes using the Mauve aligner. Bioinformatics. 2009;25:2071–3.

12. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. Bioinformatics. 2013;29:1072–5.

13. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, et al. The RAST server: rapid annotations using subsystems technology. BMC Genomics. 2008;9:75.

14. Barrnap version 0.9. https://github.com/tseemann/barrnap.

15. Laslett D, Canback B. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. Nucleic Acids Res. 2004;32:11–6.

16. Lee I, Kim YO, Park SC, Chun J. OrthoANI: an improved algorithm and software for calculating average nucleotide identity. Int J Syst Evol Microbiol. 2016;66:1100–3.

17. Mohammed WS, Ziganshina EE, Akhmetova RF, Ziganshin AM. *Bacillus* sp. EE-W1, whole genome shotgun sequencing project. *Genbank*. 2019. https://identifiers.org/ncbi/insdc:WIPE00000000.1.

18. Mohammed WS, Ziganshina EE, Akhmetova RF, Ziganshin AM. *Bacillus* sp. (in: Bacteria) strain EE-W1 16S ribosomal RNA gene, partial sequence. *Genbank*. 2019. https://identifiers.org/ncbi/insdc:MN611309.1.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.