genomeA
Journals.ASM.org

Genome Sequence of *Geobacillus* sp. Strain ZGt-1, an Antibacterial Peptide-Producing Bacterium from Hot Springs in Jordan

Rawana N. Alkhalili,[a] Rajni Hatti-Kaul,[a] Björn Canbäck[b]

Department of Biotechnology, Center for Chemistry and Chemical Engineering, Lund University, Lund, Sweden[a]; Department of Biology, Microbial Ecology Group, Lund University, Lund, Sweden[b]

This paper reports the draft genome sequence of the firmicute *Geobacillus* sp. strain ZGt-1, an antibacterial peptide producer isolated from the Zara hot spring in Jordan. This study is the first report on genomic data from a thermophilic bacterial strain isolated in Jordan.

Received 22 June 2015 **Accepted** 22 June 2015 **Published** 23 July 2015

Citation Alkhalili RN, Hatti-Kaul R, Canbäck B. 2015. Genome sequence of *Geobacillus* sp. strain ZGt-1, an antibacterial peptide-producing bacterium from hot springs in Jordan. Genome Announc 3(4):e00799-15. doi:10.1128/genomeA.00799-15.

Address correspondence to Rawana N. Alkhalili, rawana.alkhalili@biotek.lu.se.

Lately, the species of the genus *Geobacillus* have been gaining interest as antimicrobial peptide producers (1, 2). *Geobacillus* sp. strain ZGt-1, isolated from the Zara hot spring in Jordan, has been shown to produce an as-yet-uncharacterized antimicrobial peptide (3). In order to screen for the antibacterial protein-encoding genes and to identify potential novel genes associated with antibacterial peptide biosynthesis, we performed a whole-genome sequencing of the bacterium that was already identified by sequencing the PCR-amplified 16S rRNA gene (GenBank accession no. KT026965).

Here, we report the genome sequence of *Geobacillus* sp. strain ZGt-1. Total genomic DNA was extracted from pure cultures of the isolate using ZR Fungal/Bacterial DNA MiniPrep (Zymo Research). A DNA library was constructed using the Nextera protocol with modifications as described earlier (4). Input to the assembly consisted of 680,000 single-end Illumina reads with a length of 151 nucleotides. Quality control was performed by the FastQC version 0.11.2 software (http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc). Reads were assembled using Velveth and Velvetg, both with version 1.2.10 (5). This resulted in an assembly containing 9,625 contigs with a total length of 3.7 million bp. Taking into account that genome sequences from closely related strains were available, it was decided to produce ZGt-1 scaffolds based on the genome sequence of *Geobacillus kaustophilus* HTA426 (GenBank accession number NC_006510.1). This was conducted online with the Scaffold_builder tool (6) using default settings. This resulted in a new assembly with 241 scaffolds and a total length 3,483,107 bp. On this final assembly, gene prediction was carried out with Prodigal version 2_60 using default settings (7). The predicted number of protein-encoding genes was 3,546, which is close to the reported number of genes from *G. kaustophilus* HTA426 (3,397 protein-encoding genes). The GC content was calculated to 52.2% and gene density to 88%.

Genome analysis using antiSMASH version 3.0 software (8) revealed that strain ZGt-1 harbors a lantipeptide biosynthetic gene cluster, where one of the genes encodes for a lantipeptide similar to geobacillin I. The presence of this cluster was also confirmed using BAGEL version 3.0 software (9). The antiSMASH also revealed that the strain harbors another cluster containing a gene encoding for a bacteriocin similar to Linocin M18. A number of putative genes found in the lantipeptide and bacteriocin clusters showed low percentage identity with already described genes. This indicates that the ZGt-1 strain possibly possesses novel genes related to antibacterial peptide production.

Combining the *in silico* analysis of the draft genome of strain ZGt-1 with *in vitro* experimentation is likely to lead to the discovery of novel bioactive compounds.

Nucleotide sequence accession numbers. This whole-genome shotgun project has been deposited in DDBJ/EMBL/GenBank under the accession number LDPD00000000. The version described in this paper is the first version, LDPD01000000.

ACKNOWLEDGMENTS

This work was supported by Erasmus Mundus Partnership (JOSYLEEN).

We thank Gerton Lunter and the High-Throughput Genomics Group at the Wellcome Trust Centre for Human Genetics (funded by Wellcome Trust grant reference 090532/Z/09/Z), Oxford, United Kingdom, for the generation of sequencing data.

REFERENCES

1. **Garg N, Tang W, Goto Y, Nair SK, van der Donk WA.** 2012. Lantibiotics from *Geobacillus thermodenitrificans*. Proc Natl Acad Sci USA **109:** 5241–5246. http://dx.doi.org/10.1073/pnas.1116815109.
2. **Pokusaeva K, Kuisiene N, Jasinskyte D, Rutiene K, Saleikiene J, Chitavichius D.** 2009. Novel bacteriocins produced by *Geobacillus stearothermophilus*. Open Life Sciences **4:**196–203. http://dx.doi.org/10.2478/s11535-009-0009-1.
3. **Alkhalili R, Dishisha T, Mamo G, Hatti-Kaul R.** Abstr 667, Abstr 3rd Int Conf Antimicrob Res, 1–3 October 2014, Madrid, Spain.
4. **Lamble S, Batty E, Attar M, Buck D, Bowden R, Lunter G, Crook D, El-Fahmawi B, Piazza P.** 2013. Improved workflows for high throughput library preparation using the transposome-based nextera system. BMC Biotechnol **13:**104. http://dx.doi.org/10.1186/1472-6750-13-104.
5. **Zerbino DR, Birney E.** 2008. Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. Genome Res **18:**821–829. http://dx.doi.org/10.1101/gr.074492.107.
6. **Silva GG, Dutilh BE, Matthews TD, Elkins K, Schmieder R, Dinsdale EA,**

Edwards RA. 2013. Combining *de novo* and reference-guided assembly with scaffold_builder. Source Code Biol Med **8:**23. http://dx.doi.org/10.1186/1751-0473-8-23.
7. **Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ**. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics **11:**119. http://dx.doi.org/10.1186/1471-2105-11-119.
8. **Weber T, Blin K, Duddela S, Krug D, Kim HU, Bruccoleri R, Lee SY, Fischbach MA, Müller R, Wohlleben W, Breitling R, Takano E, Medema MH**. 6 May 2015. antiSMASH 3.0—a comprehensive resource for the genome mining of biosynthetic gene clusters. Nucleic Acids Res. http://dx.doi.org/10.1093/nar/gkv437.
9. **van Heel AJ, de Jong A, Montalbán-López M, Kok J, Kuipers OP**. 2013. BAGEL3: automated identification of genes encoding bacteriocins and (non-)bactericidal posttranslationally modified peptides. Nucleic Acids Res **41:**W448–W453. http://dx.doi.org/10.1093/nar/gkt391.