Genomics Data 11 (2017) 95–97

Contents lists available at ScienceDirect

Genomics Data

journal homepage: www.elsevier.com/locate/gdata

Data in Brief

Genome sequencing and annotation of *Geobacillus* sp. 1017, a hydrocarbon-oxidizing thermophilic bacterium isolated from a heavy oil reservoir (China)

Vitaly V. Kadnikov [a], Andrey V. Mardanov [a], Andrey B. Poltaraus [b], Diyana S. Sokolova [c], Ekaterina M. Semenova [c], Nikolay V. Ravin [a], Tatiyana P. Tourova [c], Tamara N. Nazina [c],*

[a] *Institute of Bioengineering, Research Center of Biotechnology, Russian Academy of Sciences, Prospect 60-letiya Oktyabrya, 7/1, Moscow 117312, Russian Federation*
[b] *Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, ul. Vavilova, 32, Moscow 119991, Russian Federation*
[c] *Winogradsky Institute of Microbiology, Research Center of Biotechnology, Russian Academy of Sciences, Prospect 60-letiya Oktyabrya, 7/2, Moscow 117312, Russian Federation*

ARTICLE INFO

Article history:
Received 21 December 2016
Accepted 24 December 2016
Available online 29 December 2016

Keywords:
Geobacillus sp.
Whole genome sequencing
Petroleum reservoir
Hydrocarbons
Biodegradation

ABSTRACT

The draft genome sequence of *Geobacillus* sp. strain 1017, a thermophilic aerobic oil-oxidizing bacterium isolated from formation water of the Dagang high-temperature oilfield, China, is presented here. The genome comprised 3.6 Mbp, with the G + C content of 51.74%. The strain had a number of genes responsible for numerous metabolic and transport systems, exopolysaccharide biosynthesis, and decomposition of sugars and aromatic compounds, as well as the genes related to resistance to metals and metalloids. The genome sequence is available at DDBJ/EMBL/GenBank under the accession no MQMG00000000. This genome is annotated for elucidation of the genomic and phenotypic diversity of new thermophilic alkane-oxidizing bacteria of the genus *Geobacillus*.

Specifications	
Organism/cell line/tissue	*Geobacillus* sp.
Sequencer or array type	Sequencer; Roche 454
Data format	Processed
Experimental factors	Microbial strains
Experimental features	Draft genome sequence of *Geobacillus* sp. strain 1017 assembly and annotation
Consent	N/A
Sample source location	A production water sample from the Dagang high-temperature heavy oil reservoir (Hebei Province, China) (38°40′4.7″ N, 117°22′38.0″ E).

1. Direct link to deposited data

https://www.ncbi.nlm.nih.gov/nuccore/MQMG01000000

2. Experimental design, materials and methods

The genus *Geobacillus* was originally proposed in 2001 [1]. Members of this genus growing at temperatures between 35 and 75 °C are widespread throughout the world and can be recovered from various natural environments with temperatures ranging from 80 °C to that of glacier ice (such as hot springs, petroleum reservoirs, compost, marine hydrothermal vents, soils, lake and ocean sediments, and arctic waters) [1]. A recent revision of this genus [2] based on analysis of the 16S rRNA gene sequences resulted in a considerable decrease in the number of *Geobacillus* species. Apart from the 16S rRNA gene, sequence similarity analyses of the *recN*, *gyrB*, and *parE* genes was proposed for differentiation of species within the genus *Geobacillus* [3,4].

The strain 1017 (=VKM B-3132) was isolated from a production water sample of the Dagang high-temperature heavy oil reservoir (Hebei Province, China) (38°40′4.7″ N, 117°22′38.0″ E) [4,5]. Strain 1017 is an aerobic endospore-forming bacterium, motile during the early exponential growth phase, and able to grow within the temperature range from 38 to 68 °C with an optimum at 60 °C. The strain utilizes fatty acids, sugars, amino acids, aromatic compounds, and *n*-alkanes of crude oil. The 16S rRNA gene sequence of strain 1017 has 99.5% similarity with the respective gene of *Geobacillus kaustophilus* R-35646 (FN428694) [5]. Analysis of the *gyrB* and *parE* genes of the new strain

---

* Corresponding author.
E-mail address: nazina@inmi.ru (T.N. Nazina).

http://dx.doi.org/10.1016/j.gdata.2016.12.011

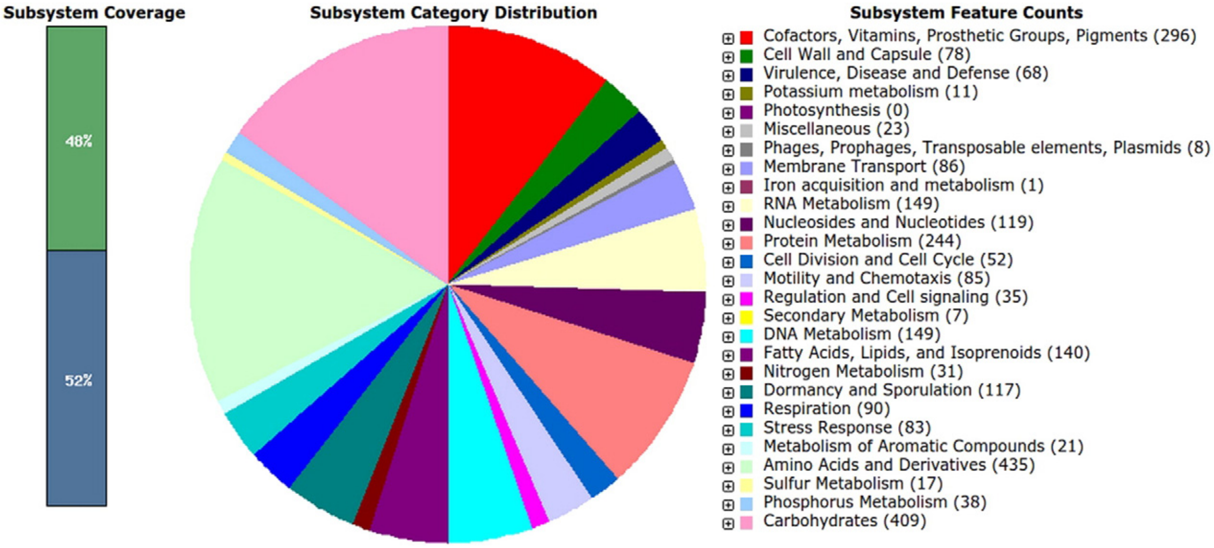

Fig. 1. Subsystems of *Geobacillus* sp. 1017 based on SEED database.

1017 revealed 96.0% and 97% similarity to the relevant genes of *G. kaustophilus* NBRC 102445^{T}, (NZ_BBJV01000018, BBJV01000017), which made it possible to suggest this strain as a new *Geobacillus* species. Analysis of the genome of strain 1017 was carried out in order to specify its taxonomic position and to elucidate the genotypic diversity within the genus *Geobacillus*.

Genomic DNA was isolated from the biomass using the DNeasy Blood & Tissue Kit (Qiagen, Germany) according to the manufacturer's instructions. *Geobacillus* sp. 1017 genomic DNA was sequenced with a Roche Genome Sequencer (GS FLX), using the Titanium XL+ protocol for a shotgun genome library. About 356 Mb of raw sequences with an average read length of 5394 nucleotides were produced providing approximately 100-fold genome coverage. The reads were *de novo* assembled into contigs using Newbler Assembler version 2.9 (454 Life Sciences, Branford, CT). The draft genome of *Geobacillus* sp. 1017 consists of 146 contigs longer than 500 bp, with a total length of 3,562,450 bp. The N50 contig size of the genome is 52,882 bp and the G + C content of 51.74%.

Gene search and annotation were performed for all contigs longer than 500 bp using the RAST server [6]. SEED viewer was used for assignment of the predicted genes to functional categories [7]. The genome was predicted to contain 3896 protein-coding sequences, 86 tRNAs and 24 coded rRNAs (5S, 16S, and 23S). Most of the annotated genes determined amino acids and derivatives synthesis (435), carbohydrate metabolism (409), cofactors, vitamins, prosthetic groups and pigment formations (296), protein metabolism (244), and RNA metabolism (149) (Fig. 1).

The 16S rRNA, *gyrB* and *parE* gene sequences revealed in the genome had 100%, 100% and 99.9% similarity, respectively, with those determined earlier for the strain 1017 [5]. The results of phenotypic studies characterizing the strain as an aerobic spore-forming bacterium utilizing a wide range of organic substrates were confirmed by detection of respective genes. Despite utilization of $C_{10,11}$ and C_{13}–$C_{19,22}$ *n*-alkanes by strain 1017, neither *alk*B nor *ladA* genes were found in the draft genome. A range of protein-coding sequences involved in the pathways for the oxidation of aromatic compounds was detected in the genome of the strain. Further studies of the genes involved in biodegradation of alkanes and aromatic compounds from crude oil by aerobic thermophilic bacteria are necessary. Overall, *in silico* results confirmed that the strain 1017 was a member of the genus *Geobacillus*, though further characterization work is required to determine its species.

3. Nucleotide sequence accession numbers

This *Geobacillus* sp. 1017 whole genome shotgun project has been deposited at DDBJ/ENA/GenBank under the accession MQMG00000000. The version described in this paper is version MQMG01000000.

Conflict of interest

The authors declare that there is no conflict of interests on the work published in this paper.

Acknowledgements

This work was performed using the scientific equipment of the Core Research Facility 'Bioengineering' (Research Center of Biotechnology RAS) and supported by the Russian Science Foundation (grant number 16-14-00028).

References

[1] T.N. Nazina, T.P. Tourova, A.B. Poltaraus, E.V. Novikova, A.A. Grigoryan, A.E. Ivanova, A.M. Lysenko, V.V. Petrunyaka, G.A. Osipov, S.S. Belyaev, M.V. Ivanov, Taxonomic study of aerobic thermophilic bacilli: descriptions of *Geobacillus subterraneus* gen. nov., sp. nov. and *Geobacillus uzenensis* sp. nov. from petroleum reservoirs and transfer of *Bacillus stearothermophilus*, *Bacillus thermocatenulatus*, *Bacillus thermoleovorans*, *Bacillus kaustophilus*, *Bacillus thermoglucosidasius* and *Bacillus thermodenitrificans* to *Geobacillus* as the new combinations *G. stearothermophilus*, *G. thermocatenulatus*, *G. thermoleovorans*, *G. kaustophilus*, *G. thermoglucosidasius* and *G. thermodenitrificans*. Int. J. Syst. Evol. Microbiol. 51 (2001) 433–446, http://dx.doi.org/10.1099/00207713-51-2-433.

[2] A. Coorevits, A.E. Dinsdale, G. Halket, L. Lebbe, P. De Vos, A. Van Landschoot, N.A. Logan, Taxonomic revision of the genus *Geobacillus*: emendation of *Geobacillus*, *G. stearothermophilus*, *G. jurassicus*, *G. toebii*, *G. thermodenitrificans* and *G. thermoglucosidans* (nom. corrig., formerly '*thermoglucosidasius*'); transfer of *Bacillus thermantarcticus* to the genus as *G. thermantarcticus* comb. nov.; proposal of *Caldibacillus debilis* gen. nov., comb. nov.; transfer of *G. tepidamans* to *Anoxybacillus* as *A. tepidamans* comb. nov.; and proposal of *Anoxybacillus caldiproteolyticus* sp. nov. Int. J. Syst. Evol. Microbiol. 62 (2012) 1470–1485, http://dx.doi.org/10.1099/ijs.0.030346-0.

[3] D.R. Zeigler, Application of a *recN* sequence similarity analysis to the identification of species within the bacterial genus *Geobacillus*. Int. J. Syst. Evol. Microbiol. 55 (2005) 1171–1179, http://dx.doi.org/10.1099/ijs.0.63452-0.

[4] T.P. Tourova, A.V. Korshunova, E.M. Mikhailova, D.S. Sokolova, A.B. Poltaraus, T.N. Nazina, Application of *gyrB* and *parE* sequence similarity analyses for differentiation of species within the genus *Geobacillus*. Microbiology (Moscow) 79 (2010) 356–369, http://dx.doi.org/10.1134/S0026261710030124.

[5] T.P. Tourova, D.S. Sokolova, E.M. Semenova, E.S. Shumkova, A.V. Korshunova, T.L. Babich, A.B. Poltaraus, T.N. Nazina, Detection of *n*-alkane biodegradation genes *alkB* and *ladA* in thermophilic hydrocarbon-oxidizing bacteria of the genera *Aeribacillus* and *Geobacillus*. Microbiology (Moscow) 85 (2016) 710–724, http://dx.doi.org/10.1134/S0026261716060199.

[6] T. Brettin, J.J. Davis, T. Disz, R.A. Edwards, S. Gerdes, G.J. Olsen, R. Olson, R. Overbeek, B. Parrello, G.D. Pusch, M. Shukla, J.A. Thomason III, R. Stevens, V. Vonstein, A.R. Wattam, F. Xia, RASTtk: a modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. Sci. Rep. 5 (2015) 8365, http://dx.doi.org/10.1038/srep08365.

[7] R. Overbeek, R. Olson, G.D. Pusch, G.J. Olsen, J.J. Davis, T. Disz, R.A. Edwards, S. Gerdes, B. Parrello, M. Shukla, V. Vonstein, A.R. Wattam, F. Xia, R. Stevens, The SEED and the rapid annotation of microbial genomes using subsystems technology (RAST). Nucleic Acids Res. 42 (D1) (2014) D206–D214, http://dx.doi.org/10.1093/nar/gkt1226.