Data on draft genome sequence of *Bacillus* sp. strain VKPM B-3276 isolated from *Culex pipiens* larvae

V.V. Zinina, A.A. Korzhenkov, A.V. Tepliuk, A.A. Kanikovskaja, M.V. Patrushev, I.V. Kublanov, S.V. Toshchakov

*a* National Research Center “Kurchatov Institute”, Moscow, 123182, Russia

*b* Federal Institution “State Research Institute of Genetics and Selection of Industrial Microorganisms” of the National Research Center “Kurchatov Institute”, Moscow, 117545, Russia

**Abstract**

The draft genome sequence of *Bacillus* sp. VKPM B-3276, a mesophilic, gram-positive bacterium, isolated from dead *Culex pipiens* larvae is presented. This strain was deposited in the Russian National Collection of Industrial Microorganisms as a prospective candidate for development of new entomopathogenic agents. The genome of *Bacillus* sp. VKPM B-3276 was 6,126,346 bp in length with predicted completeness of 99.43%. Genome analysis identified 6518 protein-coding sequences and 111 RNAs genes. 13% (271) of the protein-coding genes were assigned to “Carbohydrates” subsystem category, according to RAST/SEED. Among them about 50 enzymes, responsible for chitin, peptidoglycan and related molecules decomposition, were found. The draft genome of strain VKPM B-3276 was deposited at DBJ/EMBL/GenBank under the accession nos. RZHM00000000, PRJNA511803 and SAMN10644103 for Genome, Bioproject and Biosample, respectively.

© 2019 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).
1. Data

Bacillus sp. strain VKPM B-3276 was isolated from *Culex pipiens* larvae as an entomopathogenic agent [1]. Its genome was sequenced using Illumina MiSeq platform to identify genes, responsible for its entomopathogenic properties. *De novo* assembly resulted in 176 contigs with average coverage of 44x. Total length of the assembly was equal to 6,126,346 bp with a G+C content of 35%. Automatic annotation by RAST (Rapid Annotation using Subsystems Technology) server [2] identified 6518 protein-coding and 111 RNA genes. Protein-coding sequences were organized in 358 subsystems, among which the most numerous were “Amino acids and derivatives” (395 genes), “Carbohydrates” (271), “Protein” (191) and “Cofactors, Vitamins, Prosthetic Groups, Pigments” (187). From almost 200 Carbohydrate Active enZymes (CAZymes) [3], detected using the dbCAN server [4] 50 were predicted to participate in decomposition of chitin and peptidoglycan and their derivatives (Fig. 1). The latter is well correlated with the isolation source of this strain as well as its entomocidic capabilities [1]. According to sequence comparisons, some of these enzymes are only distantly related to currently known members of CAZyme families and/or representing recently proposed families with limited number of members. E.g. VKPM B-3276 genome possesses a gene for GH129, a family, for which the only characterized member - α-N-acetylgalactosaminidase, possibly involved in mucin degradation [5]. This observation emphasizes the potential of this strain for other than insecticide-related applications. Analysis of genes, responsible for secondary metabolite biosynthesis, showed that VKPM B-3276 has a number of pathogen-related features. Gene clusters for bacillibactin and antrachelin siderophore
biosynthesis, system for petrobactin-mediated iron uptake, as well as multiple toxin/antitoxin systems were found. VKPM B-3276 also possesses a heme utilization system, characteristic to gram-positive pathogens [6]. Interestingly, large proportion of pathogen-related gene clusters show a high level of synten with extremely pathogenic B. anthracis [7], therefore not only accentuating the importance of this strain as a prospective insecticide, but also indicating requirement of extensive safety studies before implementation of this strain in agricultural industry.

According to the analysis of 16S rRNA genes, strain VKPM B-3276 belongs to “Bacillus cereus group” of species, including B. cereus, B. anthracis and well known entomocidic strain Bacillus thuringiensis [8]. For the purpose of refinement of strain B-3276 phylogenetic position, average nucleotide identity (ANI) was calculated between B-3276 and all available genomes of “Bacillus cereus group”. ANI analysis showed that B. thuringiensis serovar berliner ATCC 10792 (96.48%) and B. cereus strain NCTC2599 (95.96% ANI) were the closest relatives of strain VKPM B-3276 forming with it a distinctive cluster (Fig. 2). Digital DNA:DNA Hybridization analysis (DDH) showed that the probability of the hypothesis, that these strains are from the same subspecies, is less than 25%.

2. Experimental design, materials and methods

2.1. Strain isolation and deposition into collection

VKPM strain B-3276 was isolated from Culex pipiens larvae cadaver [2] and deposited in Russian National Collection of Industrial Microorganisms (VKPM). In 2018 it was sequenced in the frame of Russian program “Genomes of industrially-relevant microorganisms”.

2.2. DNA extraction, library preparation and sequencing

Genomic DNA was extracted and purified with standard phenol-chloroform method. DNA integrity was assessed by electrophoresis in agarose gel. Fragmentation of DNA was performed with Bioruptor™ sonicator (Diagenode, Belgium) to achieve an average fragment length of 500 bp. Additional step of size-selection with electrophoresis was performed before library preparation to get fragments in range from 400 to 600 bp. Further steps of library preparation were performed with KAPA™ HyperPlus fragment library kit (Roche) according to the manufacturer’s instructions. Sequencing was done with
Illumina MiSeq™ platform (Illumina, USA) using 500 cycles paired-end sequencing cartridge. 579,166 read pairs were obtained from the sequencing run.

2.3. De novo assembly

Removal of low-quality reads, bases and sequencing adapters was made with fastq-mcf [9] using the following parameters: Phred score ≥ 25, window size = 5. Genome were assembled with SPAdes v 3.10 [10] in “careful” mode. To check the quality of the assembly, reads were mapped back to contigs with bowtie2 [11], mapping file was processed with samtools [12].

2.4. Genome annotation and analysis

Genome was annotated with RAST [2] using RASTtk scheme [13]. Functional analysis was performed using the tools embedded in SEED portal [14]. CAZymes [3] prediction was done using the dbCAN meta server [4]. Analysis of genes involved in the biosynthesis of secondary metabolites was made with ANTISMASH [15] server. Average nucleotide identity was calculated using ani.rb script (https://github.com/lmrodriguez/enveomics). ANI heatmap was plotted using ggplot2 library for R. Probability of being a new species or subspecies was assessed with GGDC 2.1 [16].
Acknowledgements

This work was supported by the NRC “Kurchatov institute” (internal grant #1972 from 09.08.2018 “Genomes of industrially-relevant microorganisms”). Computationally-intensive bioinformatic part of this work has been carried out using computing resources of the federal collective usage center Complex for Simulation and Data Processing for Mega-science Facilities at NRC “Kurchatov Institute”, http://ckp.nrcki.ru/. Authors also thank Drs. Konstantin Voyushin and Konstantin Sidoruk from the Russian National Collection of Industrial Microorganisms for their valuable input.

Transparency document

Transparency document associated with this article can be found in the online version at https://doi.org/10.1016/j.dib.2019.103757.

References

[1] V.V. Shevtsov, V.I. Gajtan, O.A. Krajnova, M.P. Khovrychev, S.P. Rasnitsyn, A.A. Vojtisik, E.A. Ganina, A.I. Pakhtuev, T.F. Lomovskaja, Strain of Bacterium Bacillus Sphaericus - a Producer of Entomopathogenic Preparation against Blood-Sucking Mosquito Larvae, 1996. Russian Patent No 1305916, http://www1.fips.ru/wps/portal/IPS_Ru/#1546427624754. (Accessed 2 January 2019).

[2] R.K. Aziz, D. Bartels, A. Best, M. Dejongh, T. Disz, R.A. Edwards, K. Formsmo, S. Gerdes, E.M. Glass, M. Kubal, F. Meyer, G.J. Olsen, R. Olson, A.L. Osterman, R.A. Overbeek, L.K. McNeil, D. Paarmann, T. Paczian, B. Parrello, G.D. Pusch, C. Reich, R. Stevens, O. Vassieva, V. Vonstein, A. Wilke, O. Zagnitko, The RAST Server: rapid annotations using subsystems technology, BMC Genomics 9 (2008) 75, https://doi.org/10.1186/1471-2164-9-75.

[3] V. Lombard, H. Golconda Ramulu, E. Drula, P.M. Coutinho, B. Henriassit, The carbohydrate-active enzymes database (CAZY) in 2013, Nucleic Acids Res. 42 (2014) D490–D495, https://doi.org/10.1093/nar/gkt1178.

[4] Y. Yin, X. Mao, J. Yang, X. Chen, F. Mao, Y. Xu, DbCAN: a web resource for automated carbohydrate-active enzyme annotation, Nucleic Acids Res. 40 (2012) W445–W451, https://doi.org/10.1093/nar/gks479.

[5] M. Kiyohara, T. Nakatomi, S. Kurihara, S. Fushinobu, H. Suzuki, T. Tanaka, S.I. Shoda, M. Kitaoaka, T. Katayama, K. Yamamoto, H. Ashida, α-N-acetylgalactosaminidase from infant-associated bifidobacteria belonging to novel glycoside hydrodrolase family 129 is implicated in alternative mucin degradation pathway, J. Biol. Chem. 287 (2012) 693–700, https://doi.org/10.1074/jbc.M111.277384.

[6] H. Contreras, N. Chim, A. Credali, C.W. Goulding, Heme uptake in bacterial pathogens, Curr. Opin. Chem. Biol. 19 (2014) 34–41, https://doi.org/10.1016/j.cbpa.2013.12.014.

[7] J. Ravel, L. Jiang, S.T. Stanley, M.R. Wilson, R.S. Decker, T.D. Read, P. Worsham, P.S. Keim, S.L. Salzberg, C.M. Fraser-Liggett, D.A. Rasko, The complete genome sequence of Bacillus anthracis Ames “Ancestor”, J. Bacteriol. 91 (2009) 445–446, https://doi.org/10.1128/JB.01347-08.

[8] E. Helgason, O.A. Økstad, D.A. Caugant, H.A. Johansen, A. Fouet, M. Mock, I. Hegna, A.B. Kolstø, Bacillus anthracis, Bacillus cereus, and bacillus thuringiensis - one species on the basis of genetic evidence, Appl. Environ. Microbiol. 66 (2000) 2627–2630, https://doi.org/10.1128/AEM.66.6.2627-2630.2000.

[9] E. Aronesty, Comparison of sequencing utility programs, Open Bioinf. J. 7 (2013) 1–8, https://doi.org/10.2174/187536201360710001.

[10] A. Banklevich, S. Nurk, D. Antipov, A.A. Gurevich, M. Dvorkin, A.S. Kulikov, V.M. Lesin, S.I. Nikolenko, S. Pham, A.D. Prjibelski, A.V. Pyszkin, A.V. Sirotkin, N. Vyahhi, G. Tesler, M.A. Alekseyev, P.A. Pevzner, SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing, J. Comput. Biol. 19 (2012) 455–477, https://doi.org/10.1089/cmb.2012.0021.

[11] B. Langmead, S.L. Salzberg, Fast gapped-read alignment with Bowtie 2, Nat. Methods 9 (2012) 357–359, https://doi.org/10.1038/nmeth.1923.

[12] H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, The sequence alignment/ map format and SAMtools, Bioinformat. 25 (2009) 2078–2079, https://doi.org/10.1093/bioinformatics/btp552.

[13] 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, 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, https://doi.org/10.1038/srep08365.

[14] 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 (2014) D206–D214, https://doi.org/10.1093/nar/gkt2226.

[15] K. Blin, T. Wolf, M.G. Chevrete, X. Xu, C.J. Schwalen, S.A. Kautsar, H.C. Suarez Duran, E.L.C. de Los Santos, H.J. Kim, M. Nave, J.S. Dickshat, D.A. Mitchell, E. Shelest, R. Breitling, E. Takano, S.Y. Lee, T. Weber, M.H. Medema, antiSMASH 4.0: improvements in chemistry prediction and gene cluster boundary identification, Nucleic Acids Res. 45 (2017) W36–W41, https://doi.org/10.1093/nar/gkx319.

[16] J.P. Meier-Kolthof, A.F. Auch, H.P. Klenk, M. Göker, Genome sequence-based species delimitation with confidence intervals and improved distance functions, BMC Bioinf. 14 (2013) 60, https://doi.org/10.1186/1471-2105-14-60.