



ELSEVIER

Contents lists available at ScienceDirect

Data in brief

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



Data Article

Draft genome sequence data of *Bifidobacterium longum* strain VKPM Ac-1636, a prospective probiotic isolated from human gut



V.V. Zinina^a, A.A. Korzhenkov^a, A.V. Tepliuk^a,
Yu A. Trimasova^b, M.V. Patrushev^a, I.V. Kublanov^a,
S.V. Toshchakov^{a,*}

^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

ARTICLE INFO

Article history:

Received 11 January 2019

Received in revised form 26 February 2019

Accepted 8 March 2019

Available online 19 March 2019

ABSTRACT

Data on the draft genome sequence of *Bifidobacterium longum* subsp. *longum* strain Ac-1636 is presented in this report. This strain, isolated from the digestive tract of one-year old healthy infant, was deposited in the Russian National Collection of Industrial Microorganisms as a prospective candidate for development of probiotics and probiotic foods. The 2,321,741 bp draft genome consists of 73 scaffolds with N50 of 162,253 bp. Genome annotation revealed the presence of multiple determinants of probiotic properties of this strain. The draft genome sequence data of strain Ac-1636 is available in DBJ/EMBL/GenBank under the accession nos. RZHL00000000, PRJNA511803 and SAMN10644101 for Genome, Bioproject and Biosample, respectively.

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

* Corresponding author.

E-mail address: stepan.toshchakov@gmail.com (S.V. Toshchakov).

Specifications table

Subject area	Biology
More specific subject area	Microbial genomics
Type of data	Genomic sequence, predicted genes and annotation
How data was acquired	Whole genome sequencing of fragment libraries with Illumina MiSeq platform, following <i>de novo</i> genomic assembly
Data format	Annotated draft genome assembly
Experimental factors	Extraction of genomic DNA, fragment library preparation, Illumina sequencing, <i>de novo</i> assembly and annotation procedures
Experimental features	Extraction of genomic DNA was performed with standard phenol-chloroform method; fragment library was prepared with KAPA HyperPlus™ Kit; sequencing was performed with Illumina MiSeq™ system; genome assembly and annotation was performed with SPAdes assembler and RAST annotation system, respectively,
Data source location	The culture of strain VKPM Ac-1636 is deposited in Russian National Collection of Industrial Microorganisms (VKPM) in Moscow, Russian Federation. http://eng.genetika.ru/service-offer/vkpm/
Data accessibility	The draft genomes sequence data of strain Ac-1636 is available in DBJ/EMBL/GenBank under the accession nos. RZHL00000000 (https://www.ncbi.nlm.nih.gov/nucleotide/RZHL00000000.1/), PRJNA511803 (https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA511803) and SAMN10644101 (https://www.ncbi.nlm.nih.gov/biosample/?term=SAMN10644101) for Genome, Bioproject and Biosample, respectively.
Related research article	D.A. Sela, J. Chapman, A. Adeuya, J.H. Kim, F. Chen, T.R. Whitehead, A. Lapidus, D.S. Rokhsar, C.B. Lebrilla, J.B. German, N.P. Price, P.M. Richardson, D.A. Mills, The genome sequence of <i>Bifidobacterium longum</i> subsp. <i>infantis</i> reveals adaptations for milk utilization within the infant microbiome., Proc. Natl. Acad. Sci. U. S. A. 105 (2008) 18964–9. https://doi.org/10.1073/pnas.0809584105

Value of the data

- Initial microbiological tests revealed the capability of with this strain to inhibit the opportunistic pathogens growth. Draft genome assembly gives an opportunity to search for novel proteins and pathways of secondary metabolite biosynthesis.
- Draft genome data may be used by scientific community working in the field of probiotic microorganisms to discover molecular mechanisms of probiotic production and activity.
- Draft genome data may be used to broaden the knowledge on phylogenetic diversity of *Bifidobacteria* and, specifically, species *Bifidobacterium longum*

1. Data

In this paper we present the draft genome sequence and results of genome annotation of *Bifidobacterium longum* subsp. *longum* strain VKPM Ac-1636 obtained from the Russian National Collection of Industrial Microorganisms (RNCIM) as a potential probiotic culture. This strain, isolated in early two-thousandths from the digestive tract of one-year old healthy infant [1] was initially described as a strain of *Bifidobacterium infantis* 37b, which showed some probiotic characteristics and could be used in pediatric practice as biologically active supplement [1]. The genome was sequenced in the frame of Russian program “Genomes of industrially-relevant microorganisms” in 2018 to identify genomic determinants of its probiotic properties.

De novo assembly of strain VKPM Ac-1636 resulted in 73 genomic contigs of 2,321,741 bp total length. The largest contig was 274,158 bp, N50 and N90 assembly parameters were 162,253 and 33,761 bp, respectively. 60.2%GC content well correlated with GC content of other publicly available *Bifidobacterium longum* strains. Automatic *in silico* annotation with RAST pipeline [2] revealed 2049 coding sequences, 56 tRNA and at least 3 rRNA genes. Only two-third (1326) of *in silico* predicted proteins were assigned to Clusters of Orthologous Genes (COGs) [3]. The most abundant functional COG category was “Carbohydrate transport and metabolism”, comprising of more than 8% of all identified proteins. That observation well correlates with primary ecological niche of the strain, isolated from feces of healthy breastfed infant [1], and therefore targeting human milk oligosaccharides (HMO) as a valuable carbon and energy source. Second abundant functional category, “Amino acid transport and metabolism” (7.65%), accompanied with presence of genes involved in metabolism of thiamine, folate, pyridoxine

Table 1

Code	Value	%age	Description
J	158	7.37	Translation, ribosomal structure and biogenesis
A	1	0.05	RNA processing and modification
K	95	4.43	Transcription
L	79	3.68	Replication, recombination and repair
B	0	0.0	Chromatin structure and dynamics
D	26	1.21	Cell cycle control, Cell division, chromosome partitioning
V	37	1.72	Defense mechanisms
T	57	2.66	Signal transduction mechanisms
M	83	3.82	Cell wall/membrane biogenesis
N	8	0.37	Cell motility
U	12	0.56	Intracellular trafficking and secretion
O	57	2.66	Posttranslational modification, protein turnover, chaperones
C	47	2.19	Energy production and conversion
G	172	8.02	Carbohydrate transport and metabolism
E	164	7.65	Amino acid transport and metabolism
F	78	3.64	Nucleotide transport and metabolism
H	63	2.94	Coenzyme transport and metabolism
I	42	1.96	Lipid transport and metabolism
P	51	2.38	Inorganic ion transport and metabolism
Q	18	0.84	Secondary metabolites biosynthesis, transport and catabolism
R	90	4.2	General function prediction only
S	60	2.8	Function unknown
X	25	1.17	Mobilome: prophages, transposons
–	723	33.71	Not in COGs

and riboflavin, assumes broad prototrophy of the strain regarding amino acids and vitamins. The bile salt hydrolase gene (ELS79_06255) is involved in the resistance of the strain to the bile stress in the gastrointestinal tract of the host. Thus, draft genome sequence data of strain VKPM Ac-1636 well agree with other observations of genomic features, responsible for probiotic capabilities of different strains of genus *Bifidobacteria* [4,5].

Laboratory experiments with this strain showed that it can inhibit the growth of opportunistic pathogens (*E.coli*, *K. pneumoniae*, *S. aureus*, *S. faecalis*, *C. perfringens*) [1], but no secondary metabolite-related genes was predicted by antiSMASH server [6]. On the other hand, SMIPS server algorithm [7] predicted a gene of possible polyketide synthase type I (ELS79_09510), which might be involved in synthesis of bioactive compounds [8], and a gene for UbiA prenyltransferase (ELS79_05505) which is involved in delivery of bioactive compounds to the cell membrane [9]. However, these *in silico* observations need extensive experimental work to be reliably confirmed.(see Table 1)

2. Experimental design, materials and methods

2.1. DNA extraction, library preparation and sequencing

Strain Ac-1636 was stored in Russian National Collection of Industrial Microorganisms as a lyophilized culture. For extraction of genomic DNA, it was re-cultivated using Blaurock medium, routinely used for cultivation of *Bifidobacterium*. Genomic DNA was extracted and purified with standard phenol-chloroform method. DNA quality and integrity were assessed by agarose gel electrophoresis as well as by measurement of A260/A280 and A260/A230 by Nanodrop 1000 spectrophotometer (Thermo Fisher Scientific, USA). DNA was stored at -20°C until further processing. DNA was fragmented using a Bioruptor™ sonicator (Diagenode, Belgium) to achieve an average fragment length of 500 bp. Fragmented DNA was size-selected for fragments in range from 400 to 600 bp using agarose gel electrophoresis. Further steps of library preparation were performed with KAPA™ HyperPlus fragment library kit (Roche) according to the manufacturer's instructions. Sequencing was

performed on Illumina MiSeq™ platform (Illumina, USA) using 2 × 250-cycles paired-end sequencing kit. 1,330,745 read pairs were obtained.

2.2. De novo assembly

Quality trimming, removal of sequencing adapters, and filtering of reads was performed with fastq-mcf [10] using the following parameters: Phred score ≥ 25, window size = 5. Overlapping paired reads were merged with the SeqPrep tool (<https://github.com/jstjohn/SeqPrep/>). Genome was assembled with SPAdes v 3.10 [11] in “careful” mode. To check the quality of the assembly, reads were mapped back to contigs with bowtie2 [12], mapping file was processed with samtools [13].

2.3. Genome annotation

Gene prediction and primary functional analysis was performed with RAST server [2]. Analysis of genes involved in the biosynthesis of secondary metabolites was made with ANTIMASH [6] and SMIPS [7] servers. COG annotation was performed as described previously [14].

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.2019.103847>.

References

- [1] V.M. Korshunov, B.A. Efimov, L.I. Kafarskaja, N.N. Volodin, E.A. Postnikova, Strain of *Bacterium Bifidobacterium Infantis* 37B for Preparing Biologically Active Supplements Regulating Intestine Microflora in Infants, 2003. Russian Patent No 2213140, http://www1.fips.ru/wps/portal/IP_S_Ru#1546110918128. (Accessed 29 December 2018).
- [2] R.K. Aziz, D. Bartels, A. Best, M. DeJongh, T. Disz, R.A. Edwards, K. Formsma, 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] M.Y. Galperin, K.S. Makarova, Y.I. Wolf, E.V. Koonin, Expanded microbial genome coverage and improved protein family annotation in the COG database, *Nucleic Acids Res.* 43 (2015) D261–D269, <https://doi.org/10.1093/nar/gku1223>.
- [4] D.A. Sela, J. Chapman, A. Adeuya, J.H. Kim, F. Chen, T.R. Whitehead, A. Lapidus, D.S. Rokhsar, C.B. Lebrilla, J.B. German, N.P. Price, P.M. Richardson, D.A. Mills, The genome sequence of *Bifidobacterium longum* subsp. *infantis* reveals adaptations for milk utilization within the infant microbiome, *Proc. Natl. Acad. Sci. U.S.A.* 105 (2008) 18964, <https://doi.org/10.1073/pnas.0809584105>. –9.
- [5] M. Ventura, F. Turroni, D. van Sinderen, Probiogenomics as a tool to obtain genetic insights into adaptation of probiotic bacteria to the human gut, *Bioeng. Bugs* 3 (2012) 73–79, <https://doi.org/10.4161/bbug.18540>.
- [6] K. Blin, T. Wolf, M.G. Chevrette, X. Lu, C.J. Schwalen, S.A. Kautsar, H.G. Suarez Duran, E.L.C. de Los Santos, H.U. Kim, M. Nave, J.S. Dickschat, 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>.
- [7] T. Wolf, V. Shelest, N. Nath, E. Shelest, CASSIS, SMIPS, Promoter-based prediction of secondary metabolite gene clusters in eukaryotic genomes, *Bioinformatics* 32 (2016) 1138–1143, <https://doi.org/10.1093/bioinformatics/btv713>.
- [8] G.D. Amoutzias, A. Chaliotis, D. Mossialos, Discovery Strategies of bioactive compounds synthesized by nonribosomal peptide synthetases and type-I polyketide synthases derived from marine microbiomes, *Mar. Drugs* 14 (2016) 80, <https://doi.org/10.3390/md14040080>.
- [9] W. Li, Bringing bioactive compounds into membranes: the UbiA superfamily of intramembrane aromatic prenyl-transferases, *Trends Biochem. Sci.* 41 (2016) 356–370, <https://doi.org/10.1016/j.tibs.2016.01.007>.

- [10] E. Aronesty, Comparison of sequencing utility programs, *Open Bioinf. J.* 7 (2013) 1–8, <https://doi.org/10.2174/1875036201307010001>.
- [11] A. Bankevich, 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. Pyshkin, 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>.
- [12] 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>.
- [13] 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, *Bioinformatics* 25 (2009) 2078–2079, <https://doi.org/10.1093/bioinformatics/btp352>.
- [14] S.V. Toshchakov, A.A. Korzhenkov, N.I. Samarov, I.O. Mazunin, O.I. Mozhey, I.S. Shmyr, K.S. Derbikova, E.A. Taranov, I.N. Dominova, E.A. Bonch-Osmolovskaya, M.V. Patrushev, O.A. Podosokorskaya, I.V. Kublanov, Complete genome sequence of and proposal of *Thermofilumuzonense* sp. nov. A novel hyperthermophilic crenarchaeon and emended description of the genus *Thermofilum*, *Stand. Genomic Sci.* 10 (2015) 122, <https://doi.org/10.1186/s40793-015-0105-y>.