

МИКРОБИОЛОГИЯ / MICROBIOLOGY

DOI: <https://doi.org/10.18454/jbg.2024.23.1>

COMPLETE GENOME SEQUENCE OF THE TYPE STRAIN BACTERIUM *SPHAEROCHAETA ASSOCIATA*
GLS2^T (VKM B-2742)^T

Data paper

Troshina O.Y.¹, Tepeeveva A.N.², Arzamastseva V.O.³, Whitman W.B.⁴, Varghese N.J.⁵, Shapiro N.⁶, Woyke T.⁷, Kyrpidides N.C.⁸, Vasilenko O.V.^{9,*}

¹ ORCID : 0000-0003-3319-043X;

² ORCID : 0000-0002-2895-8339;

³ ORCID : 0000-0001-6228-8730;

⁵ ORCID : 0000-0003-0580-0675;

⁶ ORCID : 0000-0002-5405-7761;

⁷ ORCID : 0000-0002-9485-5637;

⁸ ORCID : 0000-0002-6131-0462;

⁹ ORCID : 0000-0002-3333-5558;

^{1,2,3,9} Pushchino Scientific Center for Biological Research of the Russian Academy of Sciences, Pushchino, Russian Federation

⁴ University of Georgia, Athens, USA

^{5,6,7,8} DOE Joint Genome Institute, Berkeley, USA

* Corresponding author (ovvasilenko[at]gmail.com)

Abstract

This study reports the complete genome sequence of *Sphaerochaeta associata* GLS2^T (=VKM B-2742^T =DSM 26261^T), which was isolated from a consortium with methanogenic archaeon *Methanosarcina mazei* JL01. The consortium was collected from permafrost of the Kolyma lowland in Russia. The hybrid approach, combining paired-end Illumina reads with Oxford Nanopore Technologies MinION reads, was used to assemble the genome. The final assembly resulted in a circular chromosome that is 3,554,903 bp long. This high-quality genome assembly serves as a basis for algorithmic pathway reconstruction and postgenomic analysis. To further this research, the genome was imported into research portals for the algorithmic reconstruction of metabolic pathways, in both common sense (KEGG) and with special attention to carbohydrate metabolism (CAZy). These portals offer high-quality workplaces for in-depth studies.

Keywords: genome sequencing, Sphaerochaeta, hybrid genome assembly.

ЗАВЕРШЕННАЯ ГЕНОМНАЯ ПОСЛЕДОВАТЕЛЬНОСТЬ ТИПОВОГО ШТАММА БАКТЕРИИ
SPHAEROCHAETA ASSOCIATA GLS2^T (VKM B-2742)^T

Статья с данными

Трошина О.Ю.¹, Тепеева А.Н.², Арзамасцева В.О.³, Уитмен У.Б.⁴, Варгезе Н.Д.⁵, Шапиро Н.⁶, Войк Т.⁷, Кирпидес Н.С.⁸, Василенко О.В.^{9,*}

¹ ORCID : 0000-0003-3319-043X;

² ORCID : 0000-0002-2895-8339;

³ ORCID : 0000-0001-6228-8730;

⁵ ORCID : 0000-0003-0580-0675;

⁶ ORCID : 0000-0002-5405-7761;

⁷ ORCID : 0000-0002-9485-5637;

⁸ ORCID : 0000-0002-6131-0462;

⁹ ORCID : 0000-0002-3333-5558;

^{1,2,3,9} Пушчинский научный центр биологических исследований РАН, Пушкино, Российская Федерация

⁴ Университет Джорджии, Атенс, Соединенные Штаты Америки

^{5,6,7,8} Объединенный институт генома Министерства энергетики США, Беркли, Соединенные Штаты Америки

* Корреспондирующий автор (ovvasilenko[at]gmail.com)

Аннотация

В статье приведен завершённый геном бактерии *Sphaerochaeta associata* GLS2^T (=VKM B-2742^T =DSM 26261^T), которая была выделена из консорциума с метаногенной археей *Methanosarcina mazei* JL01, собранного из мерзлоты Колымской низменности в России. Геном был собран при помощи гибридного подхода, который объединил парные риды Illumina и технологию MinION Nanopore Technologies. Готовая последовательность состоит из одной циклической хромосомы длиной 3,554,903 п.н. Данная качественная сборка генома является основой для алгоритмической реконструкции метаболических путей и постгеномного анализа. Данный геном был загружен на исследовательские порталы для алгоритмической реконструкции метаболических путей в общепринятом смысле (KEGG), а также с особым вниманием к обмену углеводов (CAZy). Эти порталы представляют собой высококачественные места работы для углубленных исследований.

Ключевые слова: геномное секвенирование, Sphaerochaeta, гибридная геномная сборка.

Introduction

Currently, eight species from the family *Sphaerochaetaceae* are validly described, namely *Sphaerochaeta associata*, *S. globosa*, *S. halotolerans*, *S. pleomorpha*, *Parasphaerochaeta coccoides*, *Pleomorphochaeta caudata*, *P. naphthae*, and *P. multiformis* [1], [3], [5], [6], [8]. For *S. globosa* (CP002541), *S. pleomorpha* (CP003155), *S. halotolerans* (QUWK00000000), and *P. coccoides* (CP002659), whole genome sequences are available [9], [10], [11].

The draft genome of GLS2^T was generated at the DOE Joint Genome Institute (JGI) (Berkeley, CA, USA) under the umbrella of the Genomic Encyclopedia of Type Strains, Phase III project [12]. The GLS2^T genome project (Gp0157006) has been registered in the Genomes OnLine Database [13], the draft sequence was annotated with the IMG annotation pipeline [14], [15] and deposited in GenBank (FXUH00000000.1). The complete genome sequence of GLS2^T was subsequently obtained using a combination of Illumina and Oxford Nanopore Technologies (ONT) sequencing platforms.

Materials and Methods

S. associata strain GLS2^T was grown anaerobically in 1,1 L flasks containing 0,7 L of MS medium at 30°C [1]. Cells in the exponential growth phase were collected by centrifugation (20 min at 4500 rpm), washed with saline, and used for DNA isolation. Genomic DNA for Illumina sequencing was extracted using guanidinium thiocyanate and Triton X-100, followed by purification using the Cleanup Standard BC022 kit (Evrogen, Russia). Illumina shotgun libraries were constructed with an insert size ranging from 262 to 287 bp, and sequencing was carried out on the Illumina HiSeq 2500 platform, generating 2×151-bp paired-end reads. To ensure data quality, all raw Illumina sequence data (7,415,066 reads) were filtered using BBDuk (sourceforge.net/projects/bbmap) to remove known Illumina artifacts and PhiX sequences. Reads with more than one N, average quality scores (before trimming) below 8, or reads shorter than 51 bp (after trimming) were discarded. The remaining reads were aligned to masked versions of the human, cat, and dog references using BMAP (sourceforge.net/projects/bbmap), and reads with identities above 95% were filtered out. Sequence masking was performed using BBMask (sourceforge.net/projects/bbmap).

For long-read sequencing with ONT MinION, high molecular weight genomic DNA was extracted using a combination of guanidinium thiocyanate-Triton X-100 lysis, followed by enzymatic treatments with proteinase K and RNase A. Subsequently, DNA was further processed using the Circulomics nanobind plant nuclei big DNA kit, and size selection was performed with the Circulomics SRE XL kit. The library was prepared using the SQK-LSK110 kit, following the manufacturer's specifications, with a final long DNA selection using the large fragment buffer (LFB). Sequencing was conducted in a MinION R9.4.1 flow cell. Basecalling was performed using Guppy v6.0.1+652ffd179 in GPU mode, utilizing the super-accurate model (sup) with the --config dna_r9.4.1_450bps_sup.cfg option. A total of 137,155 "passed" reads were obtained, with an N50 of 28,900 bp.

The *de novo* hybrid assembly was performed using the Unicycler pipeline v.0.5.0 [16]. Attempts to assemble a single circular chromosome using all reads from both the Illumina and ONT platforms proved unsuccessful. The optimal assembly was achieved by moderately covering both types of reads. The best assembly incorporated only the first 10⁶ Illumina reads from the SRR5832273 file and 8,872 random ONT reads (submitted as SRR18671923). As a result, the genome was assembled into a single circular contig with an overall sequence coverage of approximately 90X, with a roughly equal contribution of long and short reads.

The assembled genome was annotated by the NCBI Prokaryotic Genome Annotation Pipeline (PGAP) v6.1, employing the best-placed reference protein set, GeneMarkS-2+ [17].

Results and Discussion

The genome consists of a circular 3,554,903 bp long chromosome with a G+C content of 51 %. Of the 3,302 total genes, 3,241 CDSs (total), 3 pseudogenes, 9 rRNAs (3 of 5S, 3 of 16S and 3 of 23S), 49 tRNAs and 3 noncoding RNAs were predicted. All three copies of 16S rRNA as well as of 5S rRNA gene sequences are identical, but 23S rRNA gene sequences reveal intragenomic variation (T – CC substitution in one copy).

The genome also was deposited in the JGI GOLD with the GOLD ID's (Study ID: Gs0156768; Project ID: Gp0622633; Analysis ID: Ga0536652) as *Sphaerochaeta associata* VKM B-2742. The annotation was conducted by IMG Annotation Pipeline v.5.1.5. The comprehensive analysis revealed the gene products that connected to the metabolic pathways and the gene families. The results are presented in the form of a multi-level text database with convenient hyperlinks.

The complete genome assembly of *S. associata* GLS2^T provides the basis for algorithmic pathway reconstruction and other studies, including comparative genomics and postgenomic analysis, on specialized public Internet portals. However, it is important to note that only high-quality completed genomes are useful for this purpose. The finalized assembly of the *S. associata* GLS2^T genome was easily imported into research portals, such as KEGG [18] and CAZy [19], for algorithmic reconstruction of metabolic pathways with a focus on carbohydrate metabolism. These portals provide high-quality environments for in-depth studies.

Data availability

The genome sequence and raw sequencing reads for VKM B-2742^T were deposited under GenBank accession number CP094929 (version CP094929.1), BioProject accession number PRJNA822125, BioSample accession number SAMN27176868, and SRA accession numbers SRR18671923 (MinION reads) and SRR5832273 (Illumina reads). The JGI GOLD's ID are: Gs0156768 (Study ID); Gp0622633 (Project ID); Ga0536652 (Analysis ID). The KEGG database identifier is T-number: T08141. The reference for *S. associata* GLS2^T in the CAZy database is www.cazy.org/b24437.html [19].

Финансирование

Работа (предложение: 10.46936/10.25585/60001401), проведенная Объединенным институтом генома Министерства энергетики США, подразделением Министерства энергетики США по научным исследованиям, поддержана Управлением науки Министерства энергетики США по контракту № DE-AC02-05CH11231. Работа, проведенная в коллекции ВКМ ФИЦ “Пушкинский научный центр биологических исследований РАН” поддержана грантом Министерства науки и высшего образования Российской Федерации (грантовое соглашение № 075-15-2021-1051).

Конфликт интересов

Не указан.

Рецензия

Все статьи проходят рецензирование. Но рецензент или автор статьи предпочли не публиковать рецензию к этой статье в открытом доступе. Рецензия может быть предоставлена компетентным органам по запросу.

Funding

This research work was supported by grants from U.S. Department of Energy (Contract No. DE-AC02-05CH11231) and from the Ministry of Science and Higher Education of the Russian Federation (Grant agreement № 075-15-2021-1051). The work (proposal: 10.46936/10.25585/60001401) conducted by the U.S. Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, is supported by the Office of Science of the U.S. Department of Energy operated under Contract No. DE-AC02-05CH11231.

Conflict of Interest

None declared.

Review

All articles are peer-reviewed. But the reviewer or the author of the article chose not to publish a review of this article in the public domain. The review can be provided to the competent authorities upon request.

Список литературы на английском языке / References in English

1. Troshina O. *Sphaerochaeta Associata* sp nov., a Spherical Spirochaete Isolated from Cultures of *Methanosarcina Mazei* JL01 / O. Troshina, V. Oshurkova, N. Suzina et al. // *Int. J. Syst. Evol. Microbiol.* — 2015. — 65(12). — p. 4315-4322. — DOI: 10.1099/ijsem.0.000575.
2. Rivkina E. Biogeochemistry of Methane and Methanogenic Archaea in Permafrost / E. Rivkina, V. Shcherbakova, K. Laurinavichius et al. // *FEMS Microbiology Ecology.* — 2007. — 61(1). — p. 1-15. — DOI: 10.1111/j.1574-6941.2007.00315.x.
3. Ritalahti K.M. *Sphaerochaeta Globosa* gen. nov., sp nov and *Sphaerochaeta Pleomorpha* sp nov., Free-living, Spherical Spirochaetes / K.M. Ritalahti, S.D. Justicia-Leon, K.D. Cusick et al. // *Int. J. Syst. Evol. Microbiol.* — 2012. — 62(1). — p. 210-216. — DOI: 10.1099/ijms.0.023986-0.
4. Bidzhieva S.K. *Sphaerochaeta Halotolerans* sp. nov., a Novel Spherical Halotolerant Spirochete from a Russian Heavy Oil Reservoir, Emended Description of the Genus *Sphaerochaeta*, Reclassification of *Sphaerochaeta Coccoides* to a New Genus *Parasphaerochaeta* gen. nov. as *Parasphaerochaeta Coccoides* comb. nov. and Proposal of *Sphaerochaetaceae* fam. nov. / S.K. Bidzhieva, D.S. Sokolova, D.S. Grouzdev et al. // *Int. J. Syst. Evol. Microbiol.* — 2020. — 70(8). — p. 4748-4759. — DOI: 10.1099/ijsem.0.004340.
5. Dröge S. *Sphaerochaeta Coccoides* sp. nov., a Novel Coccoid Spirochete from the Hindgut of the Termite *Neotermes Castaneus* / S. Dröge, J. Fröhlich, R. Radek et al. // *Appl. Environ. Microbiol.* — 2006. — 72(1). — p. 392-397. — DOI: 10.1128/aem.72.1.392-397.2006.
6. Arroua B. *Pleomorphochaeta Caudata* gen. nov., sp nov., an Anaerobic Bacterium Isolated from an Offshore Oil Well, Reclassification of *Sphaerochaeta Multiformis* MO-SPC2T as *Pleomorphochaeta Multiformis* MO-SPC2T comb. nov as the Type Strain of This Novel Genus and Emended Description of the Genus *Sphaerochaeta* / B. Arroua, A. Ranchou-Peyruse, M. Ranchou-Peyruse et al. // *Int. J. Syst. Evol. Microbiol.* — 2017. — 67(2). — p. 417-424. — DOI: 10.1099/ijsem.0.001641.
7. Arroua B. *Pleomorphochaeta Naphthae* sp. nov., a New Anaerobic Fermentative Bacterium Isolated from an Oil Field / B. Arroua, R. Grimaud, A. Hirschler-Réa et al. // *Int. J. Syst. Evol. Microbiol.* — 2018. — 68(12). — p. 3747-3753. — DOI: 10.1099/ijsem.0.003048.
8. Miyazaki M. *Sphaerochaeta Multiformis* sp nov.; an Anaerobic, Psychrophilic Bacterium Isolated from Subseafloor Sediment, and Emended Description of the Genus *Sphaerochaeta* / M. Miyazaki, S. Sakai, K.M. Ritalahti et al. // *Int. J. Syst. Evol. Microbiol.* — 2014. — 64(12). — p. 4147-4154. — DOI: 10.1099/ijms.0.068148-0.
9. Caro-Quintero A. The Chimeric Genome of *Sphaerochaeta*: Nonspiral Spirochetes That Break with the Prevalent Dogma in Spirochete Biology / A. Caro-Quintero, K.M. Ritalahti, K.D. Cusick et al. // *Mbio.* — 2012. — 3(3). — DOI: 10.1128/mbio.00025-12.
10. Abt B. Complete Genome Sequence of the Termite Hindgut Bacterium *Sphaerochaeta Coccoides* Type Strain (SPN1T), Reclassification in the Genus *Sphaerochaeta* as *Sphaerochaeta Coccoides* comb. nov and Emendations of the Family *Sphaerochaetaceae* and the Genus *Sphaerochaeta* / B. Abt, C. Han, C. Scheuner et al. // *Stand. Genomic Sci.* — 2012. — 6(2). — p. 194-209. — DOI: 10.4056/signs.2796069.
11. Grouzdev D. S. Draft Genome Sequence of a Fermenting Bacterium, "*Sphaerochaeta halotolerans*" 4-11T, from a Low-Temperature Petroleum Reservoir in Russia / D. S. Grouzdev, S. K. Bidzhieva, D. S. Sokolova et al. // *Microbiol. Resour. Announc.* — 2018. — 7(21). — p. e01345-18. — DOI: 10.1128/mra.01345-18.
12. Whitman W. B. Genomic Encyclopedia of Bacterial and Archaeal Type Strains, Phase III: the genomes of soil and plant-associated and newly described type strains / W. B. Whitman, T. Woyke, H. P. Klenk et al. // *Stand. Genomic Sci.* — 2015. — 10(26). — DOI: 10.1186/s40793-015-0017-x.
13. Mukherjee S. Genomes OnLine Database (GOLD) v.8: overview and updates / S. Mukherjee, D. Stamatis, J. Bertsch et al. // *Nucleic Acids Res.* — 2021. — 49(D1). — p. D723-D733. — DOI: 10.1093/nar/gkaa983.

14. Huntemann M. The Standard Operating Procedure of the DOE-JGI Microbial Genome Annotation Pipeline (MGAP v.4) / M. Huntemann, N. N. Ivanova, K. Mavromatis et al. // *Stand. Genomic Sci.* — 2015. — 10. — p. 86. — DOI: 10.1093/nar/gkaa983.
15. Chen I. A. The IMG/M Data Management and Analysis System v.6.0: New Tools and Advanced Capabilities / I. A. Chen, K. Chu, K. Palaniappan et al. // *Nucleic Acids Res.* — 2021. — 49(D1). — p. D751-D763. — DOI: 10.1093/nar/gkaa939.
16. Wick R. R. Unicycler: Resolving Bacterial Genome Assemblies from Short and Long Sequencing Reads / R. R. Wick, L. M. Judd, C. L. Gorrie et al. // *PLoS Comput. Biol.* — 2017. — 13(6). — p. e1005595. — DOI: 10.1371/journal.pcbi.1005595.
17. Li W. RefSeq: Expanding the Prokaryotic Genome Annotation Pipeline Reach with Protein Family Model Curation O'Neill / W. Li, K. R. O'Neill, D. H. Haft // *Nucleic Acids Res.* — 2021. — 49(D1). — p. D1020-D1028. — DOI: 10.1093/nar/gkaa1105.
18. KEGG GENOME: *Sphaerochaeta associata*. — 2023 — URL: https://www.kegg.jp/kegg-bin/show_organism?org=sass (accessed: 16.05.2023)
19. CAZy: Bacteria. — 2023 — URL: <http://www.cazy.org/b24437.html> (accessed: 16.05.2023)