ГЕНЕТИКА / GENETICS

COMPLETE GENOME SEQUENCE OF THE BACTERIUM *PSEUDOMONAS SHIRAZICA HY376* FROM NATURAL WATERS OF ALGIERS

Data paper

Haffaressas Y.[1], *, Seghouani N.[2], Arzamastseva V.O.[3], Tepeeva A.N.[4], Vasilenko O.V.[5]
[1] ORCID : 0009-0001-2404-2809;
[3] ORCID : 0000-0001-6228-8730;
[4] ORCID : 0000-0002-2895-8339;
[5] ORCID : 0000-0002-3333-5558;
[1] Russian Biotechnological University, Puschino, Russian Federation
[2] Pasteur Institute of Algeria, Algiers, Algeria
[3, 4, 5] Pushchinsky Scientific Center for Biological Research of the Russian Academy of Sciences, Puschino, Russian Federation

* Corresponding author (yacinechabani[at]yandex.com)

Abstract

Pseudomonas shirazica (the synonymous with *Pseudomonas asiatica)* is highly diverse bacterial species with important phenotypic traits relevant to human activities. We sequenced and assembled the genome of *P. shirazica* strain HY376 using long-read sequencing. The assembly resulted in a circular chromosome of 5,683,422 bp with 62.6% G+C content. The annotation revealed 5,181 genes, including 4,951 coding genes, 130 pseudogenes, 22 rRNAs, 74 tRNAs, and four non-coding RNAs. Additionally, it uncovered gene multiplicity and intra-genomic variations in key rRNA genes, which are important for molecular taxonomy. This study adds the eighth complete-level *P. shirazica* genome to GenBank, along with two complete-level *P. shirazica* and five *P. asiatica* genomes.

Keywords: genome sequencing, Pseudomonas shirazica, Pseudomonas asiatica, long reads genome assembly.

ЗАВЕРШЕННАЯ ГЕНОМНАЯ ПОСЛЕДОВАТЕЛЬНОСТЬ БАКТЕРИИ *PSEUDOMONAS SHIRAZICA* HY376 ПРИРОДНЫХ ВОД Г. АЛЖИР

Статья с данными

Хаффарессас Я.[1], *, Сегвани Н.[2], Арзамасцева В.О.[3], Тепеева А.Н.[4], Василенко О.В.[5]
[1] ORCID : 0009-0001-2404-2809;
[3] ORCID : 0000-0001-6228-8730;
[4] ORCID : 0000-0002-2895-8339;
[5] ORCID : 0000-0002-3333-5558;
[1] Российский биотехнологический университет, Пущино, Российская Федерация
[2] Институт Пастера Алжира, Алжир, Алжир
[3, 4, 5] Пущинский научный центр биологических исследований РАН, Пущино, Российская Федерация

* Корреспондирующий автор (yacinechabani[at]yandex.com)

Аннотация

Pseudomonas shirazica (синонимичен с *Pseudomonas asiatica)* - это высоко вариабельный бактериальный вид с важными фенотипическими свойствами, важными в том числе для человеческой деятельности. Мы произвели секвенирование и сборку генома штамма *P. shirazica* HY376 с использованием длинных прочтений. Сборка показала, что хромосома состоит из 5,683,422 пар оснований с содержанием GC в 62.6%. Аннотация генома выявила 5,181 ген, включая 4951 кодирующий ген, 130 псевдогенов, 22 гена рРНК, 74 гена тРНК и четыре не кодирующие РНК, а также многокопийность и внутригеномные вариации важных для молекулярной таксономии генов рРНК. Это исследование добавляет восьмой геном P. shirazica на уровне «завершенных» в ГенБанк, где уже размещены два генома *P. shirazica* и пять геномов *P. asiatica* на уровне завершенности.

Ключевые слова: геномное секвенирование, Pseudomonas shirazica, Pseudomonas asiatica, геномная сборка из длинных ридов.

Announcement

The genus Pseudomonas, described formerly by Migula in 1894, is one of the most diverse and ubiquitous bacterial genera whose species were isolated worldwide from all types of environments [1], [2], encompassing 313 validly described species as of June 2023 [3]. The species *Pseudomonas shirazica* was delineated from *P.putida* clade in 2019 by Keshavarz-Tohid et al. The strain VM14 was proposed as the type strain [4]. *Pseudomonas shirazica* Keshavarz-Tohid et al. 2020 was included in validation list No. 194 in 2020 [5]. Independently, *P.asiatica* was delineated from the same clade in 2019 by Tohya et al. [6]. *Pseudomonas shirazica* Keshavarz-Tohid et al. 2020 was assigned as a later heterotypic synonym of *Pseudomonas asiatica* Tohya et al. 2019 [7]. *P.shirazica*/*P.asiatica* encompasses strains that were isolated from various environments and that exhibit a variety of phenotypes. *P.shirazica* is characterized by Gram-negative cells forming white colonies on Kings'B medium, approximately 3 mm in diameter, after 24 h of growth. The cells are rods, motile, 1.6–5 µm in length, with a growth temperature optimum of 28°C and an optimum pH of 7.0 [4].

P. shirazica strain HY376 was grown in 50 ml flasks containing 10 ml of PYE medium at 37°C. Cells in the exponential growth phase were collected by centrifugation (20 min 4500 rpm), washed with saline and used for DNA isolation. High molecular weight genomic DNA for long-read sequencing with ONT MinION was extracted using a combination of guanidinium thiocyanate-Triton X-100 lysis, followed by proteinase K and RNAse A enzymatic treatments. Subsequently, a Circulomics nanobind plant nuclei big DNA kit was used, and DNA was further size selected with the Circulomics SRE XL kit. The library was prepared using kit ONT SQK-LSK110 following the manufacturer's instructions, with a final long DNA selection using the large fragment buffer (LFB). Sequencing was performed in a MinION R9.4.1 flow cell. Basecalling was performed with Guppy v6.0.1+652ffd179 (GPU mode) using the super-accurate model (sup) and the --config dna_r9.4.1_450bps_sup.cfg option [10]. A total of 73318 "passed" reads were received (790400113 bases totally; min_len.: 131 bp; max_len.: 254026 bp; avg_len.: 10780.44 bp). Filtering with NanoFilt [8] with parameters -q 14 -l 30000 --headcrop 30 --tailcrop 30 was conducted to select best reads (Q≥14, length ≥ 30K) with appropriate expected coverage 40. This yielded 4926 reads, totaling 247722674 bases (minimum length: 29950 bp, maximum length: 253966 bp, average length: 52284.23 bp). De novo genome assembly preparation was performed using the Flye v.2.9-b1774 assembler [9]. The final assembly consisted of a single circular contig with an average coverage of 42.

The assembled genome was annotated using the NCBI Prokaryotic Genome Annotation Pipeline (PGAP) version 6.1. The annotation method employed was Best-placed reference protein set, with GeneMarkS-2+ as the gene prediction tool [11].

The genome assembly consists of a circular chromosome measuring 5683422 bp in length, with a G+C content of 62.6%. Among the total of 5181 annotated genes, 4951 are coding genes, 130 are pseudogenes, and there are 100 rRNA genes (8 copies of 5S, 7 copies of 16S, and 7 copies of 23S), 74 tRNA genes, and 4 non-coding RNA genes. The WGS sequencing revealed intragenomic variations in the copies of ribosomal RNA genes. Of particular interest is the 16S ribosomal RNA locus, where the first three copies are identical. We designated the first copy (locus 343032..344568) as the reference. The fourth and sixth copies share the same SNP g.343488C>T, while the fifth copy has the SNP g.344061A>G. The last, seventh copy, bears two SNPs, g.343488C>T and g.343163G>A. Similarly, the 23S ribosomal RNAs exhibit multiple intragenomic variations. Among the eight copies of 5S ribosomal RNAs, seven are identical, while the eighth copy has two SNPs

As mentioned in the introduction, the species P. asiatica/P. shirazica exhibits a broad range of phenotypic diversity, including traits that are highly relevant to practical human activities, such as pathogenicity and hydrocarbon degradation. To comprehensively understand this diversity of phenotypes and harness them for biotechnology purposes, a pangenomic approach is employed, involving the sequencing of genomes from multiple strains of a single species and generating high-quality assemblies. Currently, the GenBank repository includes two complete-level genomes of P. shirazica and five genomes of P. asiatica [12]. The genome presented in this article adds to this collection and represents the eighth genome of the species.

Data availability

The genome sequence of *P. shirazica* strain HY376 have been deposited in the GenBank database under accession number CP127845 (version CP127845.1). The corresponding BioProject accession number is PRJNA982540, while the BioSample accession number is SAMN35709901. The raw sequencing data can be accessed through the SRA (Sequence Read Archive) under accession number SRR25007542.

Конфликт интересов
Не указан.

Conflict of Interest
None declared.

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.

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. Peix A. Historical Evolution and Current Status of the Taxonomy of Genus Pseudomonas / A. Peix., M.H. Ramírez-Bahena, E. Velázquez // Infect. Genet. Evol. — 2009. — Vol. 9. — № 6. — P. 1132-1147. — DOI: 10.1016/j.meegid.2009.08.001.

2. Vásquez P.F. Phylogenetic MLSA and Phenotypic Analysis Identification of Three Probable Novel Pseudomonas Species Isolated on King George Island, South Shetland, Antarctica. / P.F. Vásquez, L.S.Higuera, M.S. Pavlov [et al.] // Braz. J. Microbiol. — 2018. — Vol. 49. — №4. — P. 695-702. — DOI: 10.1016/j.bjm.2018.02.005.

3. LPSN.dsmz.de. — URL: https://www.bacterio.net/genus/pseudomonas (accessed 16.06.23).

4. Tohid V.K. Genomic, Phylogenetic and Catabolic Re-assessment of the Pseudomonas putida clade Supports the Delineation of Pseudomonas alloputida sp. nov., Pseudomonas inefficax sp. nov., Pseudomonas persica sp. nov., and Pseudomonas shirazica sp. Nov / V.K. Tohid, J. Vacheron, A. Dubost et al. // Systematic and Applied Microbiology. — 2019. — Vol. 42. — № 4. — P. 468-480. — DOI: 10.1016/j.syapm.2019.04.004.

5. Oren A. List of New Names and New Combinations Previously Effectively, but not Validly, Published / A. Oren, G. Garrity // Int. J. Syst. Evol. Microbiol. — Vol. 70. — № 5. — P. 4043-4049. — DOI: 10.1099/ijsem.0.004244.

6. Tohya M. Pseudomonas asiatica sp. nov., Isolated from Hospitalized Patients in Japan and Myanmar / M. Tohya,S.Watanabe, K.Teramoto [et al.] // Int. J. Syst. Evol. Microbiol. — 2019. — Vol. 69. — P. 1361-1368. — DOI: 10.1099/ijsem.0.003316.

7. Tohya M. Genome Analysis-Based Reclassification of Pseudomonas fuscovaginae and Pseudomonas shirazica as Later Heterotypic Synonyms of Pseudomonas asplenii and Pseudomonas asiatica, respectively / M. Tohya, S. Watanabe,T. Tada [et al.] // Int. J. Syst. Evol. Microbiol. — 2020. — Vol. 70. — № 5. — P. 3547-3552. — DOI: 10.1099/ijsem.0.004199.

8. Coster D.W. Visualizing and Processing Long-Read Sequencing Data / D.W. Coster, S. Hert, D.T. Schultz et al. // Bioinformatics. — 2018. — Vol. 34. — № 15. — P. 2666—2669. — DOI: 10.1093/bioinformatics/bty149.

9. Kolmogorov M. Assembly of Long, Error-Prone Reads Using Repeat Graphs / M. Kolmogorov, J. Yuan, Y. Lin [et al.] // Nat. Biotechnol. — 2019. — Vol. 37. — P. 540-546. — DOI: 10.1038/s41587-019-0072-8.

10. Nanopore Guppy GPU basecalling on Windows using WSL2. — URL: https://hackmd.io/PrSp6UhqS2qxZ_rKOR18-g#Nanopore-Guppy-GPU-basecalling-on-Windows-using-WSL2 (accessed: 22.06.2023).

11. Li W. RefSeq: Expanding the Prokaryotic Genome Annotation Pipeline Reach with Protein Family Model Curation / W. Li, K.R. O'Neill, D.H. Haft [et al.] // Nucleic Acids Res. — 2021. — Vol. 49. — № D1. — P. D1020-D1028. — DOI: 10.1093/nar/gkaa1105

12. NCBI. — URL: https://www.ncbi.nlm.nih.gov/genome/browse/#!/prokaryotes (accessed: 19.06.2023).