data

MDPI

Data Descriptor

Draft Genome Sequence of the Commercial Strain *Rhizobium ruizarguesonis* bv. *viciae* RCAM1022

Olga A. Kulaeva, Evgeny A. Zorin, Anton S. Sulima, Gulnar A. Akhtemova and Vladimir A. Zhukov *

All-Russia Research Institute for Agricultural Microbiology, 196608 St. Petersburg, Russia; okulaeva@arriam.ru (O.A.K.); ezorin@arriam.ru (E.A.Z.); asulima@arriam.ru (A.S.S.); gakhtemova@arriam.ru (G.A.A.)
* Correspondence: vzhukov@arriam.ru

Abstract: Legume plants enter a symbiosis with soil nitrogen-fixing bacteria (rhizobia), thereby gaining access to assimilable atmospheric nitrogen. Since this symbiosis is important for agriculture, biofertilizers with effective strains of rhizobia are created for crop legumes to increase their yield and minimize the amounts of mineral fertilizers required. In this work, we sequenced and characterized the genome of *Rhizobium ruizarguesonis* bv. *viciae* strain RCAM1022, a component of the 'Rhizotorfin' biofertilizer produced in Russia and used for pea (*Pisum sativum* L.).

Dataset: The raw genome sequencing data of Illumina NovaSeq 6000 were submitted to the NCBI SRA database under the accession number PRJNA1038712. The assembled genome is available in the NCBI database under the BioProject accession number PRJNA1038702.

Keywords: *Rhizobium ruizarguesonis*; genome sequencing; nitrogen fixation

Citation: Kulaeva, O.A.; Zorin, E.A.; Sulima, A.S.; Akhtemova, G.A.; Zhukov, V.A. Draft Genome Sequence of the Commercial Strain *Rhizobium ruizarguesonis* bv. *viciae* RCAM1022. *Data* **2024**, *9*, 19. https://doi.org/10.3390/data9020019

Academic Editors: Leopoldo Palma, Diego Herman Sauka and Baltasar Escriche

Received: 20 November 2023
Revised: 15 January 2024
Accepted: 18 January 2024
Published: 23 January 2024

1. Summary

Soil alphaproteobacteria attributed to the group of rhizobia, or nodule bacteria, are capable of fixing atmospheric nitrogen, i.e., converting it into an assimilable form. This process is carried out in symbiosis with legumes (family Fabaceae) within the specialized symbiotic organs called root nodules [1]. Owing to this symbiosis, the plant satisfies its need for nitrogen while the bacteria gain access to the products of photosynthesis.

Legume–rhizobial symbiosis is quite specific [2] and involves a complex molecular dialogue between partners at all stages of its development; thus, certain plant species often interact only with a limited range of rhizobia species and even strains. Despite this, even bacteria that have successfully entered symbiosis differ in the positive effect they have on the plant, i.e., their symbiotic effectiveness [3]. Since there are many crops among legumes and, moreover, they themselves are often used in crop rotations to enrich soils with assimilable nitrogen, studying the mechanisms underlying symbiotic effectiveness may be of great practical importance.

Host specificity and symbiotic effectiveness are multicomponent and complex traits [4]. Some bacteria may poorly fix nitrogen but successfully colonize the roots of the host plant, and vice versa. To increase the fitness and yields of crops via symbiosis, it is necessary to select or design such 'plant-bacterium' pairs that exhibit both specificity and effectiveness.

For all major legume crops, biofertilizers containing specific (i.e., compatible with a particular plant species) rhizobial strains are developed. For pea (*Pisum sativum* L.), the biofertilizer named 'Rhizotorfin' is produced in Russia, and several field experiments have shown the positive effect of this preparation on pea yield. The strain used in 'Rhizotorfin', *Rhizobium ruizarguesonis* bv. *viciae* RCAM1022, has proved its effectiveness on various cultivars of pea [5–7]. However, the molecular basis of such effectiveness remains poorly

studied. Here, we sequenced, assembled, and performed the basic annotation of this strain's genome. Through comparative analysis with a database of known *Rh. ruizarguesonis* genomes, the closest strain was identified, being *Rh. ruizarguesonis* bv. *viciae* RCAM1026 [8]. Our study may be of use for future research into the mechanisms of effective interaction between rhizobia and legumes.

2. Data Description

2.1. Strain Description

According to the catalog of the Russian Collection of Agricultural Microorganisms (RCAM) at the All-Russia Research Institute for Agricultural Microbiology (ARRIAM), Saint Petersburg, Russia, *Rh. ruizarguesonis* strain RCAM1022 was isolated from the roots of grain pea (cultivar 'Moskovsky 572') originating from the Moscow region in the Scientific Research Institute of Agriculture of the Non-Chernozem Zone in 1954 (the original name of the strain was 245a). The strain was deposited in the RCAM at the ARRIAM in 1968. Currently, the strain (accession number RCAM1022) is stored at −80 °C in the automated Tube Store (LiCONiC Instruments, Mauren, Lichtenstein) in the RCAM at the ARRIAM (Saint Petersburg, Russia).

Long-term field experiments have shown that this strain is characterized by effective nitrogen fixation [9]. RCAM1022 serves as a basis for the biofertilizer 'Rhizotorfin' (ULTRASTIM®, Ekos Biopreparaty, Saint Petersburg, Russia) for pea (*Pisum sativum* L.). Several types of 'Rhizotorfin' with rhizobia specific to a particular plant species increase the yield of peas, lentils, lathyrus, vetch, and broad beans due to the formation of nodules with effective nitrogen fixation [5–7,9].

For present genome sequencing and analysis, the strain RCAM1022 was retrieved from the Russian Collection of Agricultural Microorganisms (RCAM) (All-Russia Research Institute for Agricultural Microbiology).

2.2. Genome Assembly and Annotation

To sequence the genome of the RCAM1022 strain, a short-read sequencing library was prepared using the Illumina DNA Prep kit (Illumina, Inc., San Diego, CA, USA) and sequenced using the Illumina NovaSeq 6000 (Illumina, Inc., San Diego, CA, USA). In total, 60,035,897 2 × 150-bp reads were obtained. The adapters and low-quality sequences (Phred33 score ≥ 20) were removed using the BBDuk tool from BBMap suite ver. 39.03 (https://sourceforge.net/projects/bbmap/ (accessed on 15 January 2024)) [10]. After filtering, 112,324,092 high-quality paired reads were obtained. The genome was assembled using the Unicycler ver. 0.4.8 program [11] with default options. The assembly consists of 124 contigs, with a total length of 7,245,707 bp, an N50 of 242,536 bp, the longest contig being 600,404 bp, and a GC content of 60.85%. Of the 124 contigs, 97 are longer than 200 bp (Table 1).

Table 1. The main characteristics of the draft genome assembly of the *Rh. ruizarguesonis* strain RCAM1022.

Total amount of contigs	124
Total length of contigs	7,245,707
N50	242,536
N90	70,305
Longest segment	600,404
Contigs longer than 200 bp	97
Coverage	2447x
GC%	60.85%
Reads mapped onto assembly	100%

The Prokka pipeline ver. 1.14.5 [12] was used to annotate the assembly. The final assembly contains 6901 protein-coding genes, 2 rRNA operons, and 51 tRNA genes (Table 2).

Table 2. General description of annotated genome regions of *Rh. ruizarguesonis* strain RCAM1022.

Contigs	97
Bases	7,240,119
CDS	6954
Protein-coding genes	6901
rRNA	2
tRNA	50
tmRNA	1

MicrobeAnnotator ver. 2.0.5 [13] was used to describe the metabolic modules in which annotated genes are involved (Supplementary Table S1). Serine and threonine metabolism, pyrimidine metabolism, polyamine biosynthesis, lipid metabolism, histidine metabolism, fatty acid metabolism, and cofactor and vitamin metabolism are the most complete categories among annotated genes (Supplementary Table S1). Also, six *nod*, three *fix*, and five *nif* genes were found in the annotated genome (Supplementary Table S2).

The obtained assembly was compared with all *Rh. ruizarguesonis* and *Rh. leguminosarum* strains available in NCBI databases using the fastANI ver.1.33 algorithm [14] incorporated in the pyani python 3.8 module [15]. According to the analysis, the closest related strain to the studied one is *Rh. ruizarguesonis* bv. *viciae* strain RCAM1026 [8], with an average nucleotide identity of 99.993%. In addition, we showed a high percentage of identity with a number of other strains (Table 3).

Table 3. Strains with the highest average nucleotide identity to *Rh. ruizarguesonis* strain RCAM1022.

Strain	Average Nucleotide Identity, %	NCBI Reference Sequence
Rhizobium ruizarguesonis bv. *viciae* RCAM1026	99.993	NZ_CP084696.2
Rhizobium leguminosarum bv. *viciae* CZP3H6 RLVCZP3H601	98.859	SJNB01000001.1
Rhizobium leguminosarum bv. *viciae* CZP2H1 RLVCZP2H101	98.848	SJNE01000001.1
Rhizobium leguminosarum bv. *viciae* P2.5 RLVP2501	98.846	WIDV00000000.1
Rhizobium leguminosarum bv. *viciae* GB29 RLVGB2901	98.834	WIEK00000000.1

In total, 29 genes with variations in the nucleotide sequence between *Rh. ruizarguesonis* bv. *viciae* strain RCAM1022 and *Rh. ruizarguesonis* bv. *viciae* strain RCAM1026 were found. Among them, substitutions lead to a change in the amino acid sequence of 15 genes. These genes, in particular, encode proteins such as nodulation protein O, glutathione amide reductase, protein-export protein SecB, periplasmic oligopeptide-binding protein, etc. (Supplementary Table S3).

3. Methods

3.1. DNA Extraction

DNA was extracted from a bacterial liquid culture grown for 2 days in TY medium [16] at 28 °C. Five milliliters (mL) of bacterial culture was concentrated in a 1.5 mL tube via

successive, quick (1–2 min) centrifugations at 2700 relative centrifugal force units (RCF) (Eppendorf centrifuge 5430R, Eppendorf, Hamburg, Germany) and supernatant removal. DNA preparation was based on a method described in Wilson, 2001 [17], with essential additions necessary to reduce the concentration of polysaccharides. A total of 500 microliters (µL) of 1M NaCl were added to the sample and vortexed until the precipitate was fully resuspended. The tube was centrifuged for 10 min at 2700 RCF at room temperature (RT). The supernatant was carefully removed, and the procedure was repeated two more times. Then, the precipitate was resuspended in 500 µL of Milli-Q Water (MQ water), and the sample was centrifuged for 10 min at 3000 RCF at RT. The supernatant was removed, and the precipitate was resuspended in 567 µL of TE buffer. Then, further DNA extraction procedures were carried out in accordance with Wilson, 2001. At the final stage of isolation, the DNA precipitate was dried for 5 min at RT and then dissolved in 50 µL of TE buffer via vortexing and overnight incubation at +4 °C. Then, DNA was stored at −20 °C until library preparation and sequencing.

3.2. DNA Quality and Quantity Controls

DNA concentration was measured using the dsDNA Quantitation Broad Range Kit (Invitrogen, Waltham, MA, USA) and NanoDrop OneC (Thermo Fisher Scientific, Waltham, MA, USA). DNA quality control (the presence of contaminating components) was performed by analyzing the sample at different absorbing ratios with the NanoDrop OneC (Thermo Fisher Scientific, Waltham, MA, USA). Additionally, DNA was visualized on a 1% agarose gel using electrophoresis in a TAE buffer.

3.3. Whole Genome Sequencing, De Novo Genome Assembly, and Annotation

A whole-genome sequencing library was prepared using the Illumina DNA Prep kit (Illumina, Inc., San Diego, CA, USA) following the manufacturer's instructions.

The libraries were sequenced using the Illumina NovaSeq 6000 (Illumina, Inc., San Diego, CA, USA) in Evrogen JSC (Moscow, Russia), and 60,035,897 2 × 150 bp paired-end reads were generated. The quality of raw reads was assessed in the FASTQC program ((https://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (accessed on 15 January 2024)). The trimming of adapter sequences from short read data was conducted using the BBDuk tool from the BBTools (https://sourceforge.net/projects/bbmap/ (accessed on 15 January 2024)) [10] package for Linux. Low-quality sequences (with a Phred 33 score < Q20) were also removed with BBDuk. After processing, reads were contained at 50 to 150 bp. In total, 56,162,046 × 2 paired reads were obtained.

Genome assembly was performed using the Unicycler– but we used Illumina-only read sets, which were assembled using the SPAdes ver. 3.12.0 assembler [18] built into this pipeline. After assembling, the quality of the genome was checked with QUAST ver. 5.2.0 [19] with mapping reads to the de novo assembled genome. The structural and functional annotation genome was determined with Prokka [14], excluding contigs shorter than 200 bp. Before uploading the genome assembly to the NCBI database, the short contigs were removed from the assembly. Metabolic modules were obtained using MicrobeAnnotator ver. 2.0.5 [13] in 'light' mode.

To compare our assembly to all *Rh. ruizarguesonis* and *Rh. leguminosarum* strains, assemblies were obtained from NCBI via taxid (*Rh. ruizarguesonis* taxid = NCBI:txid2081791, *Rh. leguminosarum* taxid = NCBI:txid384) using the script genbank_get_genomes_by_taxon.py from pyani ver. 2.0. python ver. 3.8 modules. The average nucleotide identity (ANI) was calculated with an alignment-free comparison of the genomes with our assembly using the FastANI tool ver.1.33 [14].

Supplementary Materials: The following supporting information can be downloaded at https://www.mdpi.com/article/10.3390/data9020019/s1, Table S1: The completeness of metabolic modules in the assembled genome; Table S2: The list of *nod*, *fix* and *nif* genes found in the annotated genome of *Rh. ruizarguesonis* RCAM1022; Table S3: The list of genes with variations between *Rh. ruizarguesonis* RCAM1022 and *Rh. ruizarguesonis* RCAM1026.

Author Contributions: Conceptualization, V.A.Z.; methodology, O.A.K., G.A.A. and A.S.S.; validation, V.A.Z. and A.S.S.; formal analysis, E.A.Z., O.A.K., V.A.Z. and A.S.S.; investigation, E.A.Z., O.A.K., V.A.Z. and A.S.S.; resources, V.A.Z.; data curation, E.A.Z.; writing—original draft preparation, O.A.K. and E.A.Z.; writing—review and editing, V.A.Z., E.A.Z., A.S.S. and O.A.K.; visualization, E.A.Z.; supervision, V.A.Z.; project administration, V.A.Z.; funding acquisition, V.A.Z. All authors have read and agreed to the published version of the manuscript.

Funding: The work of E.A.Z., A.S.S., O.A.K., G.A.A., and V.A.Z. was supported by the Russian Science Foundation (RSF) grant number 22-16-00109.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The assembly and sequence data have been uploaded to the NCBI database. The assembly is deposited under the BioProject accession number PRJNA1038702. The raw Illumina data are deposited under the accession number PRJNA1038712.

Acknowledgments: The authors thank Vera Safronova, the head of the Russian Collection of Agricultural Microorganisms (RCAM) (ARRIAM, Saint Petersburg, Russia), for providing information about the strain; Ekaterina Vasileva, Alla Saksaganskaia, and Anastasiia Kimeklis (ARRIAM, Saint Petersburg, Russia) for the discussion of methodological aspects; and Lyudmila Dvoryaninova (ARRIAM, Saint Petersburg, Russia) for her excellent technical assistance. The DNA isolation and quantity control were performed using equipment of the Core Center 'Genomic Technologies, Proteomics and Cell Biology' in ARRIAM (Saint Petersburg, Russia).

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Oldroyd, G.E.D.; Downie, J.A. Coordinating Nodule Morphogenesis with Rhizobial Infection in Legumes. *Annu. Rev. Plant Biol.* **2008**, *59*, 519–546. [CrossRef] [PubMed]
2. Walker, L.; Lagunas, B.; Gifford, M.L. Determinants of Host Range Specificity in Legume-Rhizobia Symbiosis. *Front. Microbiol.* **2020**, *11*, 3028. [CrossRef]
3. Lindström, K.; Mousavi, S.A. Effectiveness of Nitrogen Fixation in Rhizobia. *Microb. Biotechnol.* **2020**, *13*, 1314–1335. [CrossRef] [PubMed]
4. Thrall, P.H.; Laine, A.L.; Broadhurst, L.M.; Bagnall, D.J.; Brockwell, J. Symbiotic Effectiveness of Rhizobial Mutualists Varies in Interactions with Native Australian Legume Genera. *PLoS ONE* **2011**, *6*, e23545. [CrossRef]
5. Aleshin, M.A.; Subbotina, M.G. An Effect of Doses of Nitrogen and of "Rhizotorphin" on Pea (*Pisum sativum* L.) Productivity in the Conditions of Podzol Heavy-Loam Soil. *Agrar. Sci. J.* **2019**, *8*, 4–11. [CrossRef]
6. Yurkov, A.P.; Lactionov, Y.V.; Kojemyakov, A.P.; Stepanova, G. V Symbiotic Efficiency of Bacterial and Fungal Preparations for Forage Crops According to Seed Harvest. *Kormoproizvodstvo* **2017**, *3*, 16–21.
7. Kiselev, A.S.; Popolzukhina, N.A.; Popolzukhin, P.V.; Parshutkin, I.I.; Gaidar, A.A. Influence of Inoculation on Photosynthesis and Symbiotic Parameters of Pea and Its Crop Yield in the Conditions of the Southern Forest Steppe of Western Siberia. *Vestn. NGAU* **2018**, *3*, 23–30. [CrossRef]
8. Afonin, A.; Sulima, A.; Zhernakov, A.; Zhukov, V. Draft Genome of the Strain RCAM1026 *Rhizobium leguminosarum* bv. *viciae*. *Genom. Data* **2017**, *11*, 85–86. [CrossRef]
9. Kuzakova, O.V.; Lomovatskaya, L.A.; Goncharova, A.M.; Romanenko, A.S. Effects of *Rhizobium leguminosarum* bv. *viceae* Strains different in Their Symbiotic Effectiveness on Changes in cAMP and Hydrogen Peroxide Concentrations in Cells of Pea Seedlings. *Russ. J. Plant Physiol.* **2019**, *66*, 712–717. [CrossRef]
10. Bushnell, B. *BBMap: A Fast, Accurate, Splice-Aware Aligner*; Lawrence Berkeley National Lab. (LBNL): Berkeley, CA, USA, 2014; Volume 13.
11. Wick, R.R.; Judd, L.M.; Gorrie, C.L.; Holt, K.E. Unicycler: Resolving Bacterial Genome Assemblies from Short and Long Sequencing Reads. *PLoS Comput. Biol.* **2017**, *13*, e1005595. [CrossRef]
12. Seemann, T. Prokka: Rapid Prokaryotic Genome Annotation. *Bioinformatics* **2014**, *30*, 2068–2069. [CrossRef] [PubMed]
13. Ruiz-Perez, C.A.; Conrad, R.E.; Konstantinidis, K.T. MicrobeAnnotator: A User-Friendly, Comprehensive Functional Annotation Pipeline for Microbial Genomes. *BMC Bioinform.* **2021**, *22*, 11. [CrossRef]
14. Jain, C.; Rodriguez-R, L.M.; Phillippy, A.M.; Konstantinidis, K.T.; Aluru, S. High Throughput ANI Analysis of 90K Prokaryotic Genomes Reveals Clear Species Boundaries. *Nat. Commun.* **2018**, *9*, 5114. [CrossRef] [PubMed]
15. Pritchard, L.; Glover, R.H.; Humphris, S.; Elphinstone, J.G.; Toth, I.K. Genomics and Taxonomy in Diagnostics for Food Security: Soft-Rotting Enterobacterial Plant Pathogens. *Anal. Methods* **2016**, *8*, 12–24. [CrossRef]

16. Beringer, J.E. R Factor Transfer in Rhizobiurn Zegurninosarum. *J. Gen. Microbiol.* **1974**, *84*, 188–198. [PubMed]
17. Wilson, K. Preparation of Genomic DNA from Bacteria. *Curr. Protoc. Mol. Biol.* **2001**, *56*, 2.4.1–2.4.5. [CrossRef]
18. Bankevich, A.; Nurk, S.; Antipov, D.; Gurevich, A.A.; Dvorkin, M.; Kulikov, A.S.; Lesin, V.M.; Nikolenko, S.I.; Pham, S.; Prjibelski, A.D.; et al. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *J. Comput. Biol.* **2012**, *19*, 455–477. [CrossRef]
19. Gurevich, A.; Saveliev, V.; Vyahhi, N.; Tesler, G. QUAST: Quality Assessment Tool for Genome Assemblies. *Bioinformatics* **2013**, *29*, 1072–1075. [CrossRef]