



HAL
open science

Complete genome sequences of two rhizobium strains producing azol(in)e-modified antibiotics

Dmitrii Travin, Dmitry Sutormin, Peter Mergaert, Konstantin Severinov

► To cite this version:

Dmitrii Travin, Dmitry Sutormin, Peter Mergaert, Konstantin Severinov. Complete genome sequences of two rhizobium strains producing azol(in)e-modified antibiotics. *Microbiology Resource Announcements*, 2022, 2022 (11), pp.e00722-22. 10.1128/mra.00722-22 . hal-03833088

HAL Id: hal-03833088

<https://hal.science/hal-03833088v1>

Submitted on 28 Oct 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

1 Complete genome sequences of two *Rhizobium* strains producing azol(in)e- 2 modified antibiotics

3
4 Dmitrii Y. Travin^{1,2}, Dmitry Sutormin^{1,2}, Peter Mergaert³, Konstantin Severinov^{4,*}

5 ¹ Center of Life Sciences, Skolkovo Institute of Science and Technology, Moscow, Russia

6 ² Institute of Gene Biology, Russian Academy of Science, Moscow, Russia

7 ³ Université Paris-Saclay, CEA, CNRS, Institute for Integrative Biology of the Cell (I2BC), Gif-sur-
8 Yvette, France.

9 ⁴ Waksman Institute for Microbiology, Rutgers, Piscataway, New Jersey, USA

10 *Corresponding author

11 12 13 ABSTRACT

14 Rhizobia are known for their ability to establish symbiotic relationships with plants. Specialized
15 metabolism of these bacteria remains understudied. Here, we report whole-genome sequences of
16 two rhizobia producing narrow-spectrum antirhizobial azol(in)e-modified peptides: that of
17 *Rhizobium* sp. Pop5, a phazolicin producer, and another of *Rhizobium anhuiense* T24, a
18 trifolitoxin producer.

19 20 ANNOUNCEMENT

21 Symbiotic nitrogen-fixing rhizobia compete for the host plants nodulation using a variety
22 of mechanisms including the production of antimicrobial compounds (1). Phazolicin (2) and
23 trifolitoxin (3) are antirhizobial azol(in)e-containing antibiotics belonging to ribosomally
24 synthesized and post-translationally modified peptides (RiPPs). The genome of the trifolitoxin
25 producer, isolated from the nodules of *Trifolium dubium* and originally classified as *Rhizobium*
26 *trifolii* T24 (4) was not sequenced to date except for two small, 7.1 and 7.3 kbp, regions (5, 6).
27 The genome of phazolicin-producing *Rhizobium* sp. Pop5, isolated from the nodules of
28 *Phaseolus vulgaris*, was available in contigs only (GCA_000295895.1).

29 *Rhizobium* sp. Pop5 was provided by Prof. Esperanza Martinez-Romero, and *R.*
30 *anhuiense* T24 was obtained from USDA-ARS Culture Collection (USDA2124). Both strains
31 were maintained on YM medium. Genomic DNA was extracted from overnight cultures grown
32 at 28°C using GeneJET Genomic DNA purification kit (ThermoFisher). Pop5 DNA was
33 sequenced using Oxford Nanopore Technologies (ONT) sequencing, while a combination of
34 ONT and Illumina sequencing was performed for T24 DNA. For Nanopore sequencing, libraries
35 were prepared with non-sheared DNA using the Ligation Sequencing Kit 1D (SQK-LSK109,
36 ONT) and EXP-NBD104 native barcoding expansion according to standard protocols. Genomic
37 DNA was subjected to End Repair and A tailing by NEBNext FFPE DNA Repair mix and
38 NEBNext Ultra II End repair/dA-tailing module (NEB). Sequencing adaptors were ligated using

39 NEBNext Ultra II DNA Ligation module (NEB) after AMPure XP (Beckman Coulter)
40 purification. The final product was cleaned using LFB buffer. MinION sequencing was
41 performed using R9.4 flow cell (FLO-MIN106D, ONT). Base calling was performed using
42 Guppy 6.0.1 (7) with the "sup" mode. For Illumina sequencing, DNA was sheared using Covaris
43 sonicator and 200-400 bp fragments were size-selected with AMPure XP beads. Library was
44 prepared using NEBNext Ultra II DNA Library Prep kit (NEB). Sequencing was performed on
45 MiSeq with the 250+250 bp paired-end protocol.

46 Illumina reads were filtered and trimmed with Trimmomatic v 0.39 (8). For T24, a draft
47 genome was assembled with Unicycler (v 0.5.0) (9) using ONT reads. The assembly was
48 subsequently polished with medaka (v 1.5.0), Polypolish (v 0.5.0) (10), and POLCA (v 4.0.8)
49 (11) using ONT and Illumina reads. For Pop5, a draft genome was assembled with Unicycler (v
50 0.5.0), and the assembly was polished with medaka (v 1.5.0). Default parameters were used for
51 all software unless otherwise specified. Replicons were circled and rotated automatically
52 using Unicycler and annotated using PGAP 6.1.

53 Both genomes contain multiple replicons including the chromosome and several plasmids
54 (**Table 1**). An antiSMASH (12) search revealed 12 biosynthetic gene clusters per genome
55 (**Fig. 1A**). The trifolditoxin *tfxABCDEFG* cluster is located on the chromosome, while the
56 phazolicin cluster *phzEACBDR* is encoded on the smallest plasmid (**Fig. 1A**). According to
57 average nucleotide identity (ANI) analysis by JSpeciesWS (13), the genome of
58 *R. leguminosarum* T24 is 99.08% identical (89.78% aligned nucleotides) to that of *R. anhuiense*
59 CCBAU 23252^T (GCA_003985145.1) (14) (**Fig. 1B**). Accordingly, the strain is reclassified as
60 *R. anhuiense* bv. *trifolii* T24.

61

62 **Data availability.**

63 The genome sequences reported were deposited in DDBJ/ENA/GenBank databases; the raw
64 reads were deposited in the NCBI Sequence Read Archive. Accession numbers are provided in
65 Table 1.

66

67 **ACKNOWLEDGEMENTS**

68 We thank Prof. Esperanza Martinez-Romero (Universidad Nacional Autónoma de México) and
69 USDA-ARS Culture Collection for providing the strains *Rhizobium* sp. Pop5 and *R. anhuiense*
70 bv. *trifolii* T24 (USDA2124) respectively. NGS library preparation and sequencing were
71 performed at Skoltech Sequencing Core Facilities. The work was supported by a grant from the
72 Ministry of Science and Higher Education of the Russian Federation (agreement No. 075-10-
73 2021-114 from 11 October 2021).

74

75

76 REFERENCES

- 77 1. Mendoza-Suárez M, Andersen SU, Poole PS, Sánchez-Cañizares C. 2021. Competition,
78 Nodule Occupancy, and Persistence of Inoculant Strains: Key Factors in the Rhizobium-
79 Legume Symbioses. *Front Plant Sci* 12:690567.
- 80 2. Travin DY, Watson ZL, Metelev M, Ward FR, Osterman IA, Khven IM, Khabibullina
81 NF, Serebryakova M, Mergaert P, Polikanov YS, Cate JHD, Severinov K. 2019. Structure
82 of ribosome-bound azole-modified peptide phazolicin rationalizes its species-specific
83 mode of bacterial translation inhibition. *Nat Commun* 10:4563.
- 84 3. Triplett EW, Barta TM. 1987. Trifolitoxin Production and Nodulation Are Necessary for
85 the Expression of Superior Nodulation Competitiveness by *Rhizobium leguminosarum* bv.
86 *trifolii* Strain T24 on Clover. *Plant Physiol* 85:335–42.
- 87 4. Schwinghamer EA, Belkengren RP. 1968. Inhibition of rhizobia by a strain of *Rhizobium*
88 *trifolii*: Some properties of the antibiotic and of the strain. *Arch Mikrobiol* 64:130–145.
- 89 5. Breil BT, Ludden PW, Triplett EW. 1993. DNA sequence and mutational analysis of
90 genes involved in the production and resistance of the antibiotic peptide trifolitoxin. *J*
91 *Bacteriol* 175:3693–3702.
- 92 6. Breil B, Borneman J, Triplett EW. 1996. A newly discovered gene, *tfuA*, involved in the
93 production of the ribosomally synthesized peptide antibiotic trifolitoxin. *J Bacteriol*
94 178:4150–6.
- 95 7. Wick RR, Judd LM, Holt KE. 2019. Performance of neural network basecalling tools for
96 Oxford Nanopore sequencing. *Genome Biol* 20:129.
- 97 8. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina
98 sequence data. *Bioinformatics* 30:2114–2120.
- 99 9. Wick RR, Judd LM, Gorrie CL, Holt KE. 2017. Unicycler: Resolving bacterial genome
100 assemblies from short and long sequencing reads. *PLOS Comput Biol* 13:e1005595.
- 101 10. Wick RR, Holt KE. 2022. Polypolish: Short-read polishing of long-read bacterial genome
102 assemblies. *PLOS Comput Biol* 18:e1009802.
- 103 11. Zimin A V., Salzberg SL. 2020. The genome polishing tool POLCA makes fast and
104 accurate corrections in genome assemblies. *PLOS Comput Biol* 16:e1007981.
- 105 12. Blin K, Shaw S, Kloosterman AM, Charlop-Powers Z, van Wezel GP, Medema MH,
106 Weber T. 2021. antiSMASH 6.0: improving cluster detection and comparison capabilities.
107 *Nucleic Acids Res* 49:W29–W35.
- 108 13. Richter M, Rosselló-Móra R, Oliver Glöckner F, Peplies J. 2016. JSpeciesWS: a web
109 server for prokaryotic species circumscription based on pairwise genome comparison.
110 *Bioinformatics* 32:929–931.
- 111 14. Zhang YJ, Zheng WT, Everall I, Young JPW, Zhang XX, Tian CF, Sui XH, Wang ET,
112 Chen WX. 2015. *Rhizobium anhuiense* sp. nov., isolated from effective nodules of *Vicia*
113 *faba* and *Pisum sativum*. *Int J Syst Evol Microbiol* 65:2960–2967.
- 114 15. Darling ACE, Mau B, Blattner FR, Perna NT. 2004. Mauve: Multiple Alignment of
115 Conserved Genomic Sequence With Rearrangements. *Genome Res* 14:1394–1403.

116

117

118 TABLE

119

120 **Table 1. Accession numbers, assembly and run statistics for the two *Rhizobium* genomes**

121

	<i>Rhizobium</i> sp. Pop5	<i>Rhizobium anhuiense</i> T24
BioSample accession no.	SAMN28987821	SAMN28987535
BioProject accession no.	PRJNA848232	PRJNA848223
GenBank accession no.	CP099398 CP099399 CP099400 CP099401 CP099402	CP098801 CP098802 CP098803 CP098804
Illumina sequencing reads	-	SRR20001171
ONT sequencing reads	SRR20000763	SRR20001172
Total reads length (Illumina), bp	-	463 844 988
Total reads length (ONT), bp	338 208 140	185 875 667
Number of reads (Illumina)	-	923994
Number of reads (ONT)	41698	25553
N50 (ONT), bp	16641	13212
Genome size, bp	6 742 135	7 327 117
GC content, %	60.93	60.97
Number of replicons	5	4
Replicon sizes, bp	4 430 159 951 562 503 518 454 824 402 072	4 746 307 1 441 560 767 332 371 918

122

123

124

125 FIGURE LEGEND

126

127 **Figure 1. (A)** Genome organization of the two sequenced *Rhizobium* strains visualized using

128 Proksee (<https://proksee.ca/>). The chromosomes and plasmids are shown not to scale. Colored

129 arrowheads indicate the locations of biosynthetic gene clusters identified by antiSMASH.

130 Schematic representation of *phz* and *txf* clusters is shown below with the functions of encoded

131 proteins listed. CDS, coding sequence; HSL, homoserine lactone; T3PKS, type III polyketide

132 synthase; NRPS, non-ribosomal peptide synthetase; LAP, linear azol(in)e-containing peptide;

133 RiPP, ribosomally synthesized posttranslationally-modified peptide. **(B)** Alignment of the whole

134 genome of *R. anhuiense* T24 and the draft genome of *R. anhuiense* CCBAU 23252^T

135 (GCA_003985145.1) (14) generated with Mauve (15).

136

