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

Dmitrii Y. Travin[1,2], Dmitry Sutormin[1,2], Peter Mergaert[3], Konstantin Severinov[4,*]

[1] Center of Life Sciences, Skolkovo Institute of Science and Technology, Moscow, Russia
[2] Institute of Gene Biology, Russian Academy of Science, Moscow, Russia
[3] Université Paris-Saclay, CEA, CNRS, Institute for Integrative Biology of the Cell (I2BC), Gif-sur-Yvette, France.
[4] Waksman Institute for Microbiology, Rutgers, Piscataway, New Jersey, USA
*Corresponding author

ABSTRACT

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

ANNOUNCEMENT

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

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

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

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

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

Data availability.

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

ACKNOWLEDGEMENTS

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

REFERENCES

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

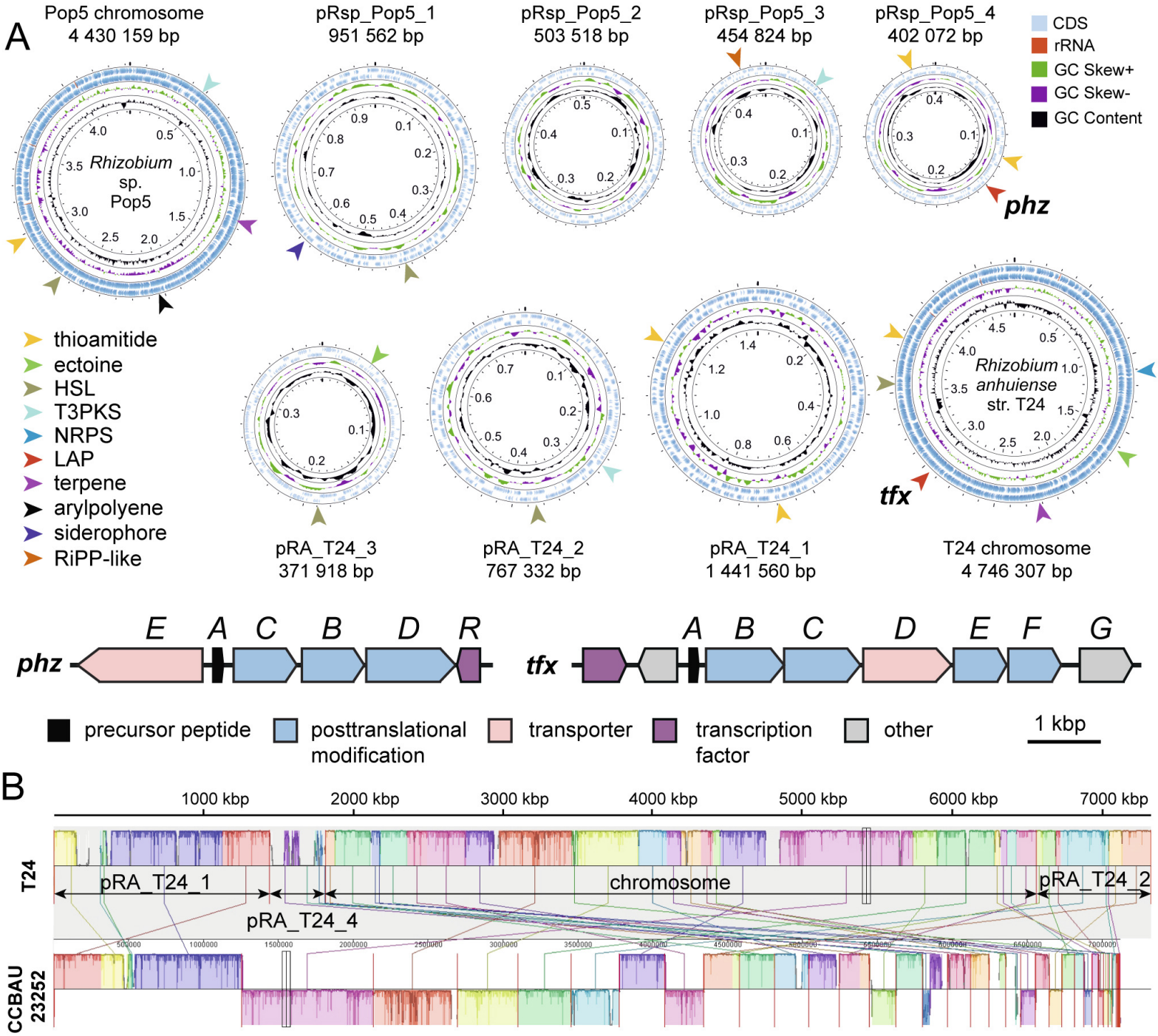
TABLE

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

	Rhizobium sp. Pop5	*Rhizobium anhuiense* 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

FIGURE LEGEND

Figure 1. (A) Genome organization of the two sequenced *Rhizobium* strains visualized using Proksee (https://proksee.ca/). The chromosomes and plasmids are shown not to scale. Colored arrowheads indicate the locations of biosynthetic gene clusters identified by antiSMASH. Schematic representation of *phz* and *tfx* clusters is shown below with the functions of encoded proteins listed. CDS, coding sequence; HSL, homoserine lactone; T3PKS, type III polyketide synthase; NRPS, non-ribosomal peptide synthetase; LAP, linear azol(in)e-containing peptide; RiPP, ribosomally synthesized posttranslationally-modified peptide. **(B)** Alignment of the whole genome of *R. anhuiense* T24 and the draft genome of *R. anhuiense* CCBAU 23252^{T} (GCA_003985145.1) (14) generated with Mauve (15).


A
Pop5 chromosome
4 430 159 bp
pRsp_Pop5_1
951 562 bp
pRsp_Pop5_2
503 518 bp
pRsp_Pop5_3
454 824 bp
pRsp_Pop5_4
402 072 bp
CDS
rRNA
GC Skew+
GC Skew-
GC Content
Rhizobium sp. Pop5
phz
thioamitide
ectoine
HSL
T3PKS
NRPS
LAP
terpene
arylpolyene
siderophore
RiPP-like
pRA_T24_3
371 918 bp
pRA_T24_2
767 332 bp
pRA_T24_1
1 441 560 bp
Rhizobium anhuiense str. T24
tfx
T24 chromosome
4 746 307 bp
phz
E A C B D R
tfx
A B C D E F G
precursor peptide
posttranslational modification
transporter
transcription factor
other
1 kbp
B
1000 kbp
2000 kbp
3000 kbp
4000 kbp
5000 kbp
6000 kbp
7000 kbp
T24
pRA_T24_1
chromosome
pRA_T24_2
pRA_T24_4
CCBAU 23252