

Telomere-to-telomere genome assembly of *Talaromyces purpureogenus* strain ISA502 isolated from maize rhizosphere

Jun Zhou,¹ Maria Elena Antinori,² Gabriele Bellotti,² Michel Chalot,¹ Edoardo Puglisi^{2,3}

AUTHOR AFFILIATIONS See affiliation list on p. 3.

ABSTRACT We present a telomere-to-telomere genome assembly of *Talaromyces purpureogenus* strain ISA502, isolated from the maize rhizosphere. The 30.33 Mb genome, assembled using hybrid sequencing, comprises 8 chromosomes with 99% BUSCO completeness, offering improved resolution over existing fragmented genomes and advancing studies on its ecological and biotechnological significance.

KEYWORDS *Talaromyces purpureogenus*, genome, maize

The rhizosphere hosts diverse fungi essential to nutrient cycling and plant health. Among these, *Talaromyces purpureogenus* stands out for its ability to produce bioactive metabolites such as antimicrobial compounds, pigments, and enzymes, making it highly relevant for agricultural and biotechnological applications (1, 2). Although six genome sequences of *T. purpureogenus* are publicly available, they remain fragmented (3), limiting structural resolution and downstream analyses. Here, we report the chromosome-scale genome assembly of *T. purpureogenus* strain ISA502, representing a significant improvement over existing draft genomes.

Strain ISA502 was isolated from the maize rhizosphere in Pavia, Italy (GPS: N 45.137293, E 8.776014) following the protocol of Barillot et al. (4). Briefly, rhizospheric soil was suspended in 0.9% NaCl solution, and dilutions of the suspension were plated on malt extract agar (MEA). Single mycelial colonies were subcultured on MEA, and strain purity was confirmed via single-spore isolation. Mycelial cultures were grown on MEA at 28°C (Fig. 1), and genomic DNA was extracted using the HiPure Universal DNA Kit D301 (Genepioneer Biotechnologies). A hybrid sequencing strategy was employed, combining Illumina short reads for base-level accuracy and Oxford Nanopore Technologies (ONT) long reads for resolving complex genomic structures. DNA libraries for Illumina sequencing were prepared using the VAHTS Universal DNA Library Prep Kit for Illumina V3 ND607 (Vazyme). Sequencing was performed using the Illumina NovaSeq 6000 platform (PE150), yielding 19,028,384 reads and 5.71 Gb of data with an average depth of coverage of ~177×. Quality control, performed with fastp (v0.20.0) (5), removed adapter sequences, low-quality reads (Q5), and reads with over five ambiguous bases. For ONT sequencing, genomic DNA was purified using 0.4× magnetic beads to remove short fragments (<2 kb). Libraries were prepared with the SQK-LSK110 ligation kit and sequenced on a PromethION 48 system with R9.4.3 flow cells. The run yielded 419,591 reads totaling 4.95 Gb with a read N50 of 21,477 bp (~152× depth). Reads were basecalled with Guppy (hac) and filtered using Filtlong (v0.2.1) to remove low-quality or <1,000 bp reads. Assembly began with self-correction of ONT reads using Canu (v2.1.1) (6), followed by hybrid error correction using LoRDEC (v0.9) (7) with Illumina reads. Corrected long reads were assembled using NextDenovo (v2.5) (8). Misassemblies were identified and resolved manually by aligning ONT reads with minimap2 (v2.1) (9) and inspecting dot plots generated by MUMmer4 to assess synteny and structural inconsistencies. The genome was polished in three rounds using NextPolish (v1.4.0) (10)

Editor Jason E. Stajich, University of California
Riverside, Riverside, California, USA

Address correspondence to Jun Zhou,
jun.zhou@univ-fcomte.fr.

The authors declare no conflict of interest.

Received 27 February 2025

Accepted 27 May 2025

Published 18 June 2025

Copyright © 2025 Zhou et al. This is an open-access
article distributed under the terms of the [Creative
Commons Attribution 4.0 International license](#).



FIG 1 *Talaromyces purpureogenus* ISA502 grown 10 days on MEA (upper side of the plate). From the bottom, the strain appears pale yellow.

with ONT reads, and then in three rounds using Pilon (v1.23) (11) with Illumina reads. Default parameters were used for all software unless otherwise specified.

The final assembly consists of 8 chromosomes totaling 30.33 Mb, with a GC content of 44.44% and an N50 of 4.12 Mb. The largest chromosome spans 5.80 Mb. Assembly completeness, evaluated using BUSCO (v5.2.1) (12), showed 99% completeness (97.9% single copy, 1.1% duplicated, 0.1% fragmented, and 0.9% missing) (Table 1). Read coverage analysis using minimap2 showed uniform coverage across the genome. Telomeric repeats (CCCTAA)_n were detected at both ends of all contigs, indicating complete chromosomal assemblies.

ACKNOWLEDGMENTS

Financial support has been provided to SIRAM by PRIMA, a program supported by H2020, the European Programme for Research and Innovation, by the French ANR under ANR-22-PRIM-0001-05 and by the Italian MUR (SIRAM PRIMA CALL 2021 SECTION 2 CUP J33C22000510008).

TABLE 1 Genome sequencing statistics

Parameter	Value
Illumina sequencing	
Average depth of coverage, Illumina short reads	~177×
Reads number	19,028,384
Total bases (bp)	5,708,515,200
Nanopore sequencing	
Average depth of coverage, ONT long reads	~152×
Reads number	419,591
Total bases (bp)	4,950,531,796
N50 read length (bp)	21,477
Genome assembly	
No. of chromosomes	8
Total length (bp)	30,326,944
Largest chromosome	5,800,888
GC (%)	44.44
N50 (bp)	4,121,854
L50	4
Genome assembly assessment	
Complete BUSCOs	99%
Complete and single-copy BUSCOs	97.90%
Complete and duplicated BUSCOs	1.10%
Fragmented BUSCOs	0.10%
Missing BUSCOs	0.90%

AUTHOR AFFILIATIONS

¹Université Marie et Louis Pasteur, CNRS, Chrono-environnement (UMR 6249), Montbéliard, France

²Department for Sustainable Food Process, Università Cattolica del Sacro Cuore, Piacenza, Italy

³ISA Innovations for Sustainable Agriculture R&D, Piacenza, Italy

AUTHOR ORCIDS

Jun Zhou  <http://orcid.org/0000-0003-3618-1037>

Maria Elena Antinori  <http://orcid.org/0000-0002-2258-6204>

Gabriele Bellotti  <http://orcid.org/0000-0001-8521-3256>

Michel Chalot  <http://orcid.org/0000-0001-8614-0917>

Edoardo Puglisi  <http://orcid.org/0000-0001-5051-0971>

AUTHOR CONTRIBUTIONS

Jun Zhou, Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review and editing | Maria Elena Antinori, Conceptualization, Investigation, Methodology, Visualization, Writing – review and editing | Gabriele Bellotti, Conceptualization, Investigation, Methodology, Writing – review and editing | Michel Chalot, Conceptualization, Funding acquisition, Project administration, Resources, Supervision, Writing – review and editing | Edoardo Puglisi, Conceptualization, Funding acquisition, Project administration, Resources, Supervision, Writing – review and editing

DATA AVAILABILITY

The genome and sequencing data have been deposited at NCBI under BioProject accession [PRJNA1215053](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1215053). The BioSample accession is [SAMN46395145](https://www.ncbi.nlm.nih.gov/biosample/SAMN46395145). The Illumina sequencing reads are available in the NCBI Sequence Read Archive under accession [SRR32327772](https://www.ncbi.nlm.nih.gov/sra/SRR32327772), while the nanopore sequencing reads are accessible under accession [SRR32327919](https://www.ncbi.nlm.nih.gov/sra/SRR32327919). This Whole Genome Shotgun project has been deposited in DDBJ/ENA/GenBank under the accession no. [JBLFET000000000](https://www.ncbi.nlm.nih.gov/nuclseq/JBLFET000000000). The version described in this paper is the first version, [JBLFET010000000](https://www.ncbi.nlm.nih.gov/nuclseq/JBLFET010000000).

REFERENCES

- Sahu MK, Kaushik K, Das A, Jha H. 2020. *In vitro* and *in silico* antioxidant and antiproliferative activity of rhizospheric fungus *Talaromyces purpureogenus* isolate-ABRF2. *Bioresour Bioprocess* 7:14. <https://doi.org/10.1186/s40643-020-00303-z>
- Kumari M, Taritla S, Sharma A, Jayabaskaran C. 2018. Antiproliferative and antioxidative bioactive compounds in extracts of marine-derived endophytic fungus *Talaromyces purpureogenus* *Front Microbiol* 9:1777. <https://doi.org/10.3389/fmicb.2018.01777>
- Wang Y, Yang Y, Yuan X, Ma S, Wang D, Qiao L. 2024. Draft genome sequence of endophytic fungus *Talaromyces purpureogenus* strain YAFEF302, isolated from *Juglans sigillata*. *Microbiol Resour Announc* 13:e0082923. <https://doi.org/10.1128/mra.00829-23>
- Barillot CDC, Sarde C-O, Bert V, Tarnaud E, Cochet N. 2013. A standardized method for the sampling of rhizosphere and rhizoplane soil bacteria associated to a herbaceous root system. *Ann Microbiol* 63:471–476. <https://doi.org/10.1007/s13213-012-0491-y>
- Chen S. 2023. Ultrafast one-pass FASTQ data preprocessing, quality control, and deduplication using fastp. *iMeta* 2:e107. <https://doi.org/10.1002/imt2.107>
- Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. 2017. Canu: scalable and accurate long-read assembly via adaptive *k*-mer weighting and repeat separation. *Genome Res* 27:722–736. <https://doi.org/10.1101/gr.215087.116>
- Salmela L, Rivals E. 2014. LoRDEC: accurate and efficient long read error correction. *Bioinformatics* 30:3506–3514. <https://doi.org/10.1093/bioinformatics/btu538>
- Hu J, Wang Z, Sun Z, Hu B, Ayoola AO, Liang F, Li J, Sandoval JR, Cooper DN, Ye K, Ruan J, Xiao C-L, Wang D, Wu D-D, Wang S. 2024. NextDenovo: an efficient error correction and accurate assembly tool for noisy long reads. *Genome Biol* 25:107. <https://doi.org/10.1186/s13059-024-03252-4>
- Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34:3094–3100. <https://doi.org/10.1093/bioinformatics/bty191>
- Hu J, Fan J, Sun Z, Liu S. 2020. NextPolish: a fast and efficient genome polishing tool for long-read assembly. *Bioinformatics* 36:2253–2255. <https://doi.org/10.1093/bioinformatics/btz891>
- Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, Earl AM. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 9:e112963. <https://doi.org/10.1371/journal.pone.0112963>
- Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31:3210–3212. <https://doi.org/10.1093/bioinformatics/btv351>