DATA NOTE

The genome sequence of a basidiomycete yeast, *Tausonia pullulans* (Lindner) X.Z. Liu, F.Y. Bai, M. Groenew. & Boekhout, 2016 (Mrakiaceae)

[version 1; peer review: 1 approved, 2 approved with reservations]

Richard Wright[1], Brian Douglas[2],
Royal Botanic Gardens Kew Genome Acquisition Lab,
Darwin Tree of Life Barcoding collective,
Wellcome Sanger Institute Tree of Life Management, Samples and Laboratory team,
Wellcome Sanger Institute Scientific Operations: Sequencing Operations,
Wellcome Sanger Institute Tree of Life Core Informatics team,
Tree of Life Core Informatics collective, Darwin Tree of Life Consortium

[1]Cardiff University, Cardiff, Wales, UK
[2]Royal Botanic Gardens Kew, Richmond, England, UK

v1 **First published:** 20 Nov 2024, **9**:685
https://doi.org/10.12688/wellcomeopenres.23350.1
Latest published: 20 Nov 2024, **9**:685
https://doi.org/10.12688/wellcomeopenres.23350.1

Abstract

We present a genome assembly from an individual *Tausonia pullulans* (a basidiomycete yeast; Basidiomycota; Tremellomycetes; Cystofilobasidiales; Mrakiaceae). The genome sequence is 23.9 megabases in span. Most of the assembly is scaffolded into 20 chromosomal pseudomolecules. The mitochondrial genome has also been assembled and is 18.82 kilobases in length.

Keywords

Tausonia pullulans, basidiomycete yeast, genome sequence, chromosomal, Cystofilobasidiales

This article is included in the Tree of Life gateway.

Open Peer Review

Approval Status ? ✓ ?

	1	2	3
version 1 20 Nov 2024	? view	✓ view	? view

1. **David Pizarro**, Universidad Complutense de Madrid, Madrid, Spain
2. **Reyim Mamut**, Xinjiang University, Urumqi, China
3. **David Diaz Escandon**, University of Alberta, Edmonton, Canada

Any reports and responses or comments on the article can be found at the end of the article.

Corresponding author: Darwin Tree of Life Consortium (mark.blaxter@sanger.ac.uk)

Author roles: Wright R: Investigation, Resources, Writing – Original Draft Preparation, Writing – Review & Editing; **Douglas B**: Investigation, Resources;

Competing interests: No competing interests were disclosed.

Grant information: This work was supported by Wellcome through core funding to the Wellcome Sanger Institute [206194, https://doi.org/10.35802/206194] and the Darwin Tree of Life Discretionary Award [218328, https://doi.org/10.35802/218328].
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

How to cite this article: Wright R, Douglas B, Royal Botanic Gardens Kew Genome Acquisition Lab *et al.* **The genome sequence of a basidiomycete yeast, *Tausonia pullulans* (Lindner) X.Z. Liu, F.Y. Bai, M. Groenew. & Boekhout, 2016 (Mrakiaceae) [version 1; peer review: 1 approved, 2 approved with reservations]** Wellcome Open Research 2024, **9**:685
https://doi.org/10.12688/wellcomeopenres.23350.1

First published: 20 Nov 2024, **9**:685 https://doi.org/10.12688/wellcomeopenres.23350.1

Species taxonomy

Eukaryota; Opisthokonta; Fungi; Dikarya; Basidiomycota; Tremellomycetes; Tremellomycetidae; Cystofilobasidiales; Mrakiaceae; *Tausonia*; *Tausonia pullulans* ((Lindner) Xin Zhan Liu, F.Y. Bai, M. Groenew. & Boekhout, 2015) (NCBI:txid82525)

Background

Tausonia pullulans is a psychrotolerant basidiomycete yeast in Tremellomycetes (Trochine *et al.*, 2022a). It produces budding cells that are ovoid to cylindrical and may be enteroblastic. It can produce extensive true hyphae that disarticulate to form arthroconidia, its key method of asexual reproduction (Fell & Guého-Kellermann, 2011; Fell & Scorzetti, 2004). It is not known to sexually reproduce and lacks the clamp connections and chlamydospore production of the closely related *T. pamirica* (Sampaio, 2011).

This yeast is known from all continents (GBIF, 2023), from a broad range of substrates and habitats, including records from soil, herbaceous plants, trees, pigeon faeces, sawdust, tree sap of birch, maple, and chestnut, butter, frozen beef, packaged kimchi (Kim *et al.*, 2020) and a nasal infection in a cat (Fell & Guého-Kellermann, 2011). *T. pullulans* is most frequently reported from soils (Yurkov, 2018) and woody substrates (Cadete *et al.*, 2017).

It is known to produce higher levels of lignocellulose decay enzymes, as well as glucoamylases and pectinolytic enzymes, when compared to other Tremellomycetes (Trochine *et al.*, 2022a), and has shown cellulolytic activity (Dennis, 1972).

This species has gone through a number of taxonomic rearrangements, first described as *Oidium pullulans* by Lindner in 1901, sampled from storage barrels in an experimental brewery (Lindner, 1901). Most recently it was transferred from *Trichosporon* to a new monotypic genus, *Guehomyces* (Fell & Scorzetti, 2004), based on DNA; and subsequently moved to *Tausonia* as part of a phylogenetic reclassification of the Tremellomycetes based on analyses of seven gene regions (Liu *et al.*, 2015).

T. pullulans' enzymatic activities have been applied to the degradation of industrial pollutants (Romero *et al.*, 2002), wood pulping processes (Sláviková *et al.*, 2002), hydrolysis of lactose in milk (Nakagawa *et al.*, 2006), pectin extraction from limes (Bezus *et al.*, 2022), prevention of post-harvest decay in pears (Yao *et al.*, 2004), and starch hydrolysis (Trochine *et al.*, 2022b), amongst other applications.

Although reported as human pathogen by Moylett *et al.* (2003), it is considered unlikely due to it being a low temperature organism which has a maximum growth temperature of 25 °C (Fell & Guého-Kellermann, 2011).

T. pullulans represents an ecologically interesting species that may play an important role in the wood decay process and has many potential biotechnology applications. Further discoveries, alongside improved taxonomic understanding, will be aided by the production of this genome.

Genome sequence report

The genome was sequenced from a specimen of *Tausonia pullulans* from Siston, Gloucestershire, UK. A total of 458-fold coverage in Pacific Biosciences single-molecule HiFi long reads was generated. Primary assembly contigs were scaffolded with chromosome conformation Hi-C data. Manual assembly curation corrected 12 missing joins or mis-joins, reducing the scaffold number by 12.00%, and decreasing the scaffold N50 by 5.78%.

The final assembly has a total length of 23.9 Mb in 21 sequence scaffolds with a scaffold N50 of 1.4 Mb (Table 1). The snail plot in Figure 1 provides a summary of the assembly statistics, while the distribution of assembly scaffolds on GC proportion and coverage is shown in Figure 2. The cumulative assembly plot in Figure 3 shows curves for subsets of scaffolds assigned to different phyla. Most (99.84%) of the assembly sequence was assigned to 20 chromosomal-level scaffolds, representing 20 autosomes. Chromosome-scale scaffolds confirmed by the Hi-C data are named in order of size (Figure 4; Table 2). While not fully phased, the assembly deposited is of one haplotype. Contigs corresponding to the second haplotype have also been deposited. The mitochondrial genome was also assembled and can be found as a contig within the multifasta file of the genome submission.

The estimated Quality Value (QV) of the final assembly is 71.1 with *k*-mer completeness of 100.0%, and the assembly has a BUSCO v5.3.2 completeness of 29.9% (single = 29.7%, duplicated = 0.2%), using the tremellomycetes_odb10 reference set (n = 4,284).

Metadata for specimens, barcode results, spectra estimates, sequencing runs, contaminants and pre-curation assembly statistics are given at https://links.tol.sanger.ac.uk/species/82525.

Methods

Sample acquisition and nucleic acid extraction

The genome was sequenced from a pure culture (SB1_H10a) obtained by Richard Wright from the heartwood of an oak core sample, collected from a standing oak tree at Siston Brook, Bristol, UK (latitude 51.47, longitude –2.45).

A specimen of *Tausonia pullulans* (specimen ID KDTOL00172, ToLID gfTauPull1) was collected from Siston Brook, Siston, Gloucestershire, UK (latitude 51.47, longitude –2.45) on 2021-06-18. The specimen was handpicked and cultured before being transferred to tubes. The specimen was collected by Richard Wright (Royal Botanic Gardens Kew) and identified by Richard Wright and Brian Douglas (Royal Botanic Gardens Kew) and preserved by snap-freezing.

Protocols developed by the Wellcome Sanger Institute (WSI) Tree of Life core laboratory have been deposited on protocols.io (Denton *et al.*, 2023). The workflow for high molecular weight (HMW) DNA extraction at the WSI includes a sequence of core procedures: sample preparation; sample homogenisation, DNA extraction, fragmentation, and clean-up. In sample preparation, the gfTauPull1 sample was weighed and dissected

Table 1. Genome data for *Tausonia pullulans*, gfTauPull1.1.

Project accession data		
Assembly identifier	gfTauPull1.1	
Species	*Tausonia pullulans*	
Specimen	gfTauPull1	
NCBI taxonomy ID	82525	
BioProject	PRJEB61351	
BioSample ID	SAMEA13759877	
Isolate information	gfTauPull1 (DNA, HiC and RNA)	
Assembly metrics*		***Benchmark***
Consensus quality (QV)	71.1	*≥ 50*
k-mer completeness	100.0%	*≥ 95%*
BUSCO**	C:29.9%[S:29.7%,D:0.2%],F:3.7%,M:66.4%,n:4,284	*C ≥ 95%*
Percentage of assembly mapped to chromosomes	99.84%	*≥ 95%*
Sex chromosomes	None	*localised homologous pairs*
Organelles	Mitochondrial genome: 18.82 kb	*complete single alleles*
Raw data accessions		
PacificBiosciences SEQUEL II	ERR11242134, ERR11242135	
Hi-C Illumina	ERR11242558	
PolyA RNA-Seq Illumina	ERR11242559	
Genome assembly		
Assembly accession	GCA_951802195.1	
Accession of alternate haplotype	GCA_951802215.1	
Span (Mb)	23.9	
Number of contigs	58	
Contig N50 length (Mb)	0.7	
Number of scaffolds	21	
Scaffold N50 length (Mb)	1.4	
Longest scaffold (Mb)	2.18	

* Assembly metric benchmarks are adapted from column VGP-2020 of "Table 1: Proposed standards and metrics for defining genome assembly quality" from (Rhie *et al.*, 2021).

** BUSCO scores based on the tremellomycetes_odb10 BUSCO set using version 5.3.2. C = complete [S = single copy, D = duplicated], F = fragmented, M = missing, n = number of orthologues in comparison. A full set of BUSCO scores is available at https://blobtoolkit.genomehubs.org/view/gfTauPull1_1/dataset/gfTauPull1_1/busco.

on dry ice (Jay *et al.*, 2023). For sample homogenisation, tissue was cryogenically disrupted using the Covaris cryoPREP® Automated Dry Pulverizer (Narváez-Gómez *et al.*, 2023).

HMW DNA was extracted using the Automated Plant MagAttract v2 protocol (Todorovic *et al.*, 2023a). HMW DNA was sheared into an average fragment size of 12–20 kb in a Megaruptor 3 system (Todorovic *et al.*, 2023b). Sheared DNA was purified by solid-phase reversible immobilisation, using AMPure PB beads to eliminate shorter fragments and concentrate the DNA (Strickland *et al.*, 2023). The concentration of the sheared and purified DNA was assessed using a Nanodrop

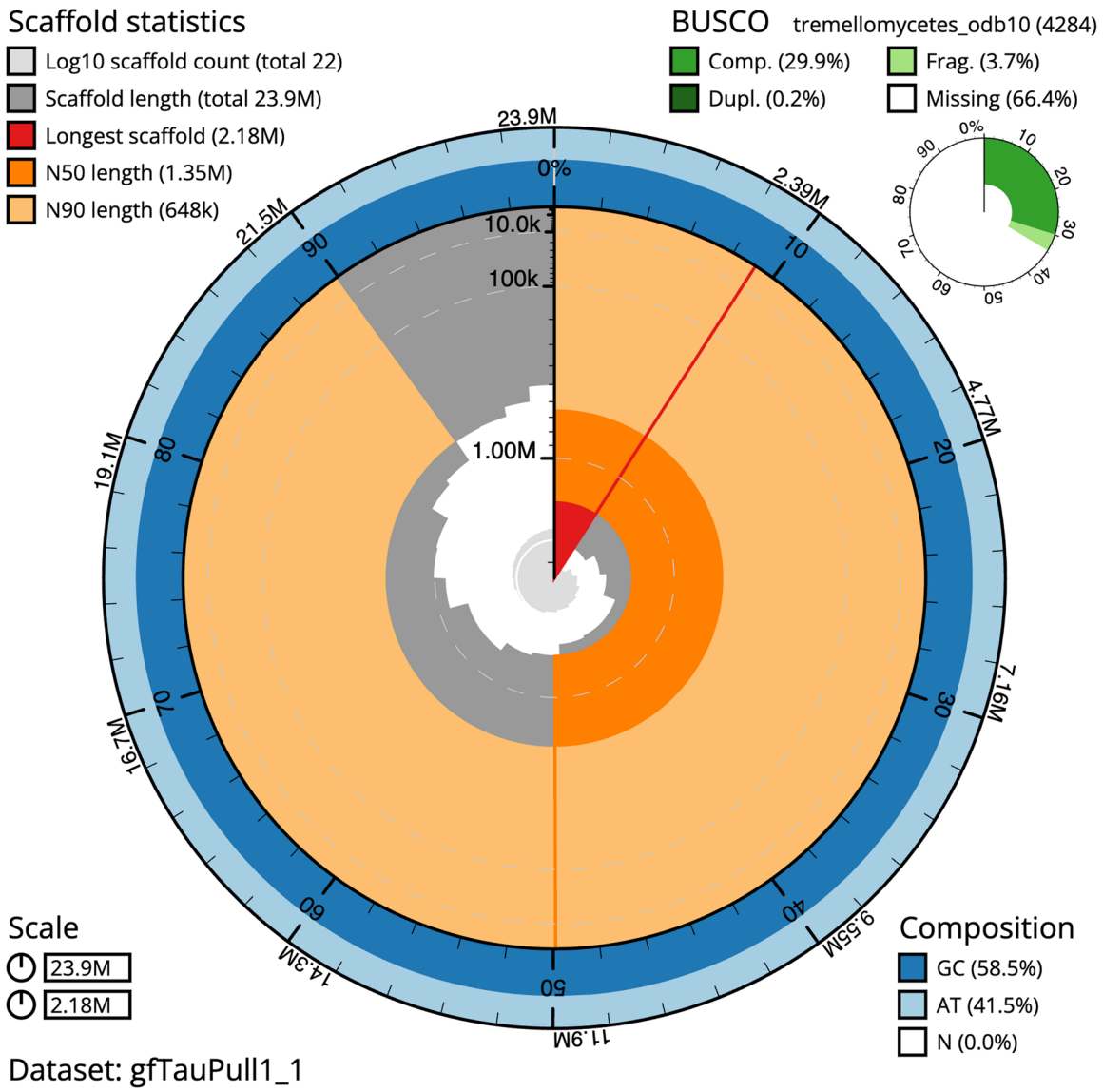

Figure 1. Genome assembly of *Tausonia pullulans*, gfTauPull1.1: metrics. The BlobToolKit Snailplot shows N50 metrics and BUSCO gene completeness. The main plot is divided into 1,000 bins around the circumference with each bin representing 0.1% of the 23,873,987 bp assembly. The distribution of scaffold lengths is shown in dark grey with the plot radius scaled to the longest scaffold present in the assembly (2,178,134 bp, shown in red). Orange and pale-orange arcs show the N50 and N90 scaffold lengths (1,354,748 and 647,864 bp), respectively. The pale grey spiral shows the cumulative scaffold count on a log scale with white scale lines showing successive orders of magnitude. The blue and pale-blue area around the outside of the plot shows the distribution of GC, AT and N percentages in the same bins as the inner plot. A summary of complete, fragmented, duplicated and missing BUSCO genes in the tremellomycetes_odb10 set is shown in the top right. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/gfTauPull1_1/dataset/gfTauPull1_1/snail.

spectrophotometer and Qubit Fluorometer and Qubit dsDNA High Sensitivity Assay kit. Fragment size distribution was evaluated by running the sample on the FemtoPulse system.

RNA was extracted from gfTauPull1 in the Tree of Life Laboratory at the WSI using the RNA Extraction: Automated MagMax™ *mir*Vana protocol (do Amaral *et al.*, 2023). The RNA concentration was assessed using a Nanodrop spectrophotometer and a Qubit Fluorometer using the Qubit RNA Broad-Range Assay kit. Analysis of the integrity of the RNA was done using the Agilent RNA 6000 Pico Kit and Eukaryotic Total RNA assay.

Sequencing

Pacific Biosciences HiFi circular consensus DNA sequencing libraries were constructed according to the manufacturers'

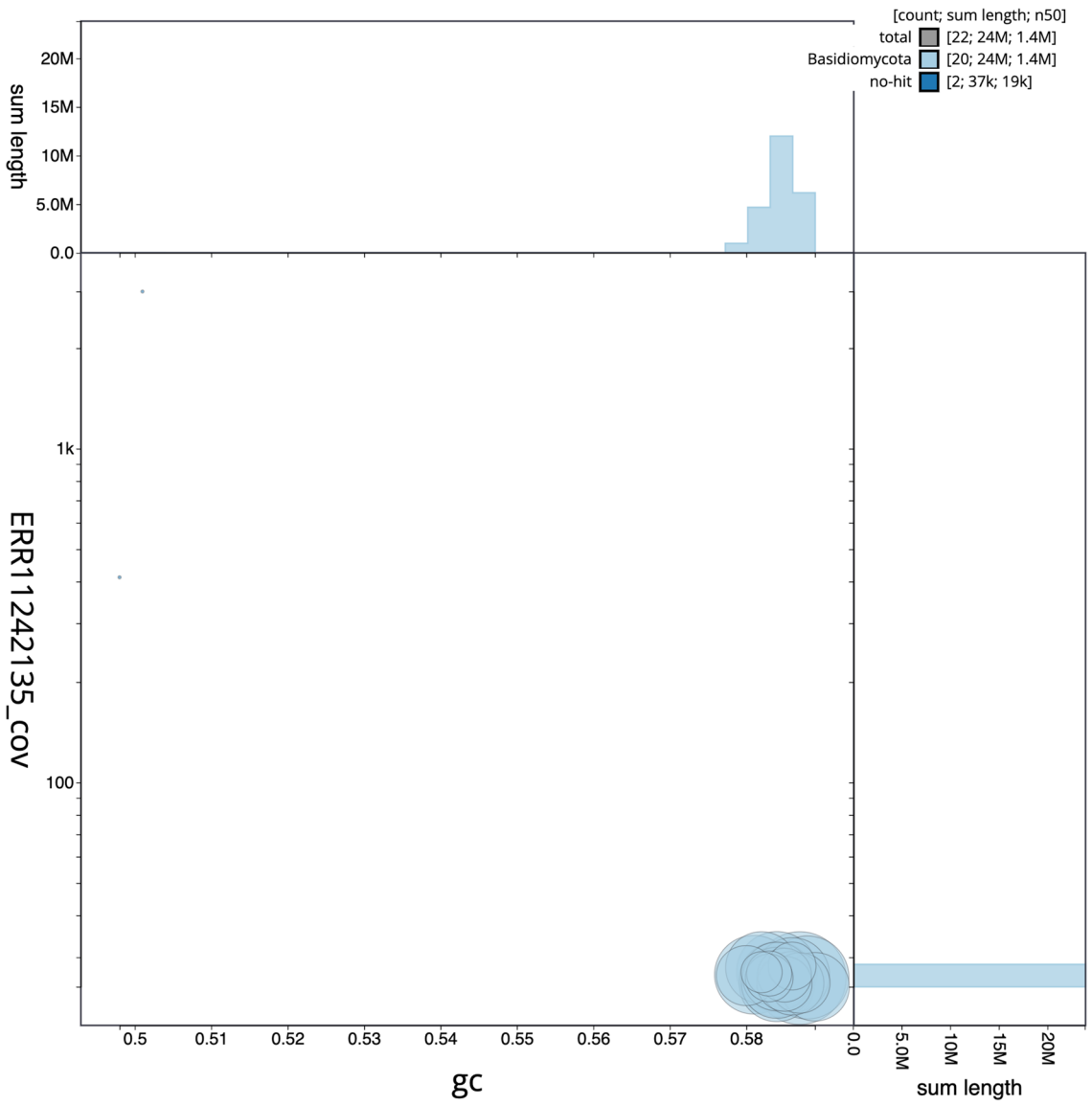

Figure 2. Genome assembly of *Tausonia pullulans*, gfTauPull1.1: BlobToolKit GC-coverage plot. Scaffolds are coloured by phylum. Circles are sized in proportion to scaffold length. Histograms show the distribution of scaffold length sum along each axis. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/gfTauPull1_1/dataset/gfTauPull1_1/blob.

instructions. Poly(A) RNA-Seq libraries were constructed using the NEB Ultra II RNA Library Prep kit. DNA and RNA sequencing was performed by the Scientific Operations core at the WSI on Pacific Biosciences SEQUEL II (HiFi) and Illumina NovaSeq 6000 (RNA-Seq) instruments. Hi-C data were also generated from cells of gfTauPull1 using the Arima2 kit and sequenced on the Illumina NovaSeq 6000 instrument.

Genome assembly, curation and evaluation

Assembly was carried out with Hifiasm (Cheng *et al.*, 2021) and haplotypic duplication was identified and removed with purge_dups (Guan *et al.*, 2020). The assembly was then scaffolded with Hi-C data (Rao *et al.*, 2014) using YaHS (Zhou *et al.*, 2023). The assembly was checked for contamination and corrected as described previously (Howe *et al.*, 2021). Manual curation was performed using HiGlass (Kerpedjiev *et al.*, 2018) and PretextView (Harry, 2022). The mitochondrial genome was assembled using MitoHiFi (Uliano-Silva *et al.*, 2023), which runs MitoFinder (Allio *et al.*, 2020) and uses these annotations to select the final mitochondrial contig and to ensure the general quality of the sequence.

A Hi-C map for the final assembly was produced using bwa-mem2 (Vasimuddin *et al.*, 2019) in the Cooler file format

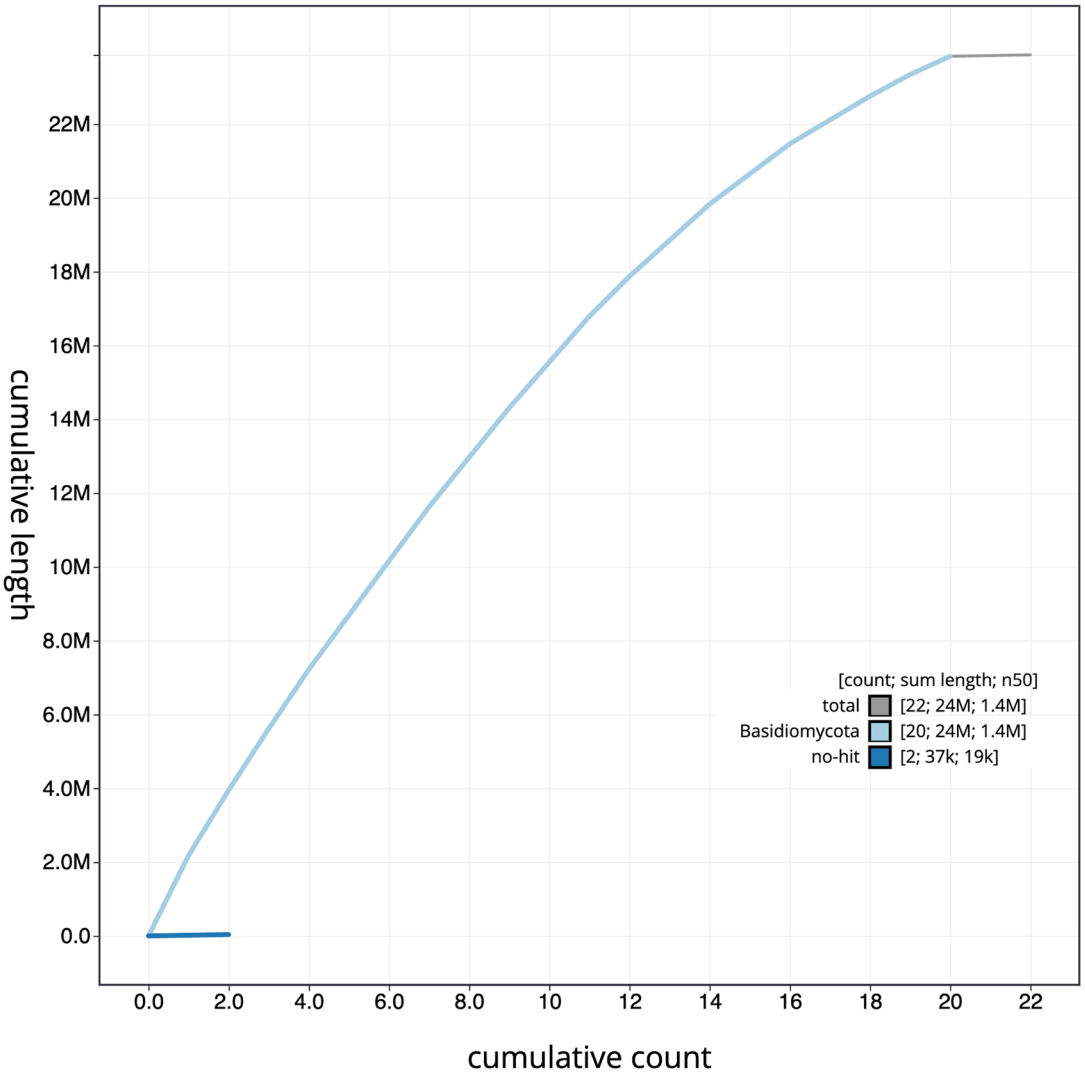

Figure 3. Genome assembly of *Tausonia pullulans*, gfTauPull1.1: BlobToolKit cumulative sequence plot. The grey line shows cumulative length for all scaffolds. Coloured lines show cumulative lengths of scaffolds assigned to each phylum using the buscogenes taxrule. An interactive version of this figure is available at https://blobtoolkit.genomehubs.org/view/gfTauPull1_1/dataset/gfTauPull1_1/cumulative.

(Abdennur & Mirny, 2020). To assess the assembly metrics, the *k*-mer completeness and QV consensus quality values were calculated in Merqury (Rhie *et al.*, 2020). This work was done using Nextflow (Di Tommaso *et al.*, 2017) DSL2 pipelines "sanger-tol/readmapping" (Surana *et al.*, 2023a) and "sanger-tol/genomenote" (Surana *et al.*, 2023b). The genome was analysed within the BlobToolKit environment (Challis *et al.*, 2020) and BUSCO scores (Manni *et al.*, 2021) were calculated.

Table 3 contains a list of relevant software tool versions and sources.

Wellcome Sanger Institute – Legal and Governance

The materials that have contributed to this genome note have been supplied by a Darwin Tree of Life Partner. The submission of materials by a Darwin Tree of Life Partner is subject to the **'Darwin Tree of Life Project Sampling Code of Practice'**, which can be found in full on the Darwin Tree of Life website here. By agreeing with and signing up to the Sampling Code of Practice, the Darwin Tree of Life Partner agrees they will meet the legal and ethical requirements and standards set out within this document in respect of all samples acquired for, and supplied to, the Darwin Tree of Life Project.

Further, the Wellcome Sanger Institute employs a process whereby due diligence is carried out proportionate to the nature of the materials themselves, and the circumstances under which they have been/are to be collected and provided for use. The purpose of this is to address and mitigate any potential legal and/or ethical implications of receipt and use of the

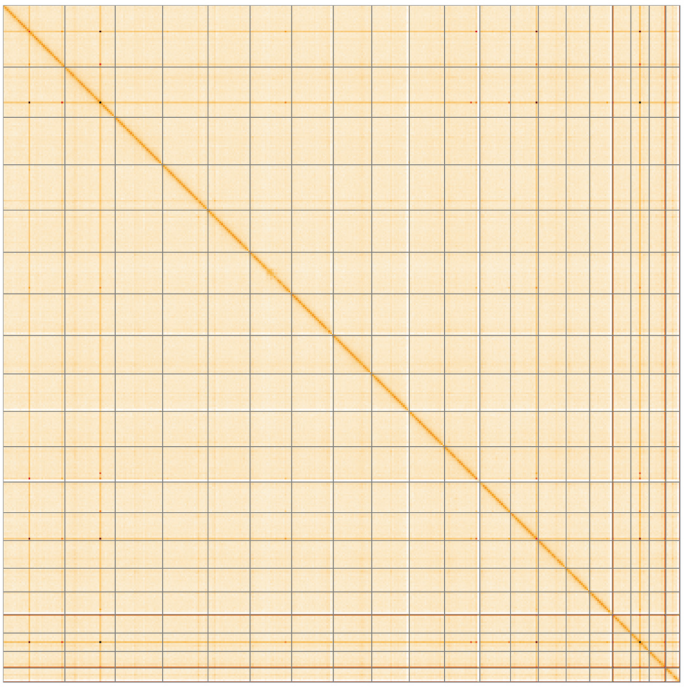

Figure 4. Genome assembly of *Tausonia pullulans*, gfTauPull1.1: Hi-C contact map of the gfTauPull1.1 assembly, visualised using HiGlass. Chromosomes are shown in order of size from left to right and top to bottom. An interactive version of this figure may be viewed at https://genome-note-higlass.tol.sanger.ac.uk/l/?d=BU4zDe9QSj6FbzspKlCP2A.

Table 2. Chromosomal pseudomolecules in the genome assembly of *Tausonia pullulans*, gfTauPull1.

INSDC accession	Chromosome	Length (Mb)	GC%
OX637620.1	1	2.18	58.5
OX637621.1	2	1.77	59.0
OX637622.1	3	1.67	58.5
OX637623.1	4	1.6	58.0
OX637624.1	5	1.48	58.5
OX637625.1	6	1.47	58.5
OX637626.1	7	1.46	58.5
OX637627.1	8	1.35	58.0
OX637628.1	9	1.33	58.5
OX637629.1	10	1.24	58.5
OX637630.1	11	1.24	59.0
OX637631.1	12	1.08	58.5
OX637632.1	13	0.98	58.5
OX637633.1	14	0.97	58.0
OX637634.1	15	0.83	58.5
OX637635.1	16	0.8	58.5
OX637636.1	17	0.65	58.5
OX637637.1	18	0.64	58.5
OX637638.1	19	0.58	58.5
OX637639.1	20	0.5	58.0
OX637640.1	MT	0.02	50.0

materials as part of the research project, and to ensure that in doing so we align with best practice wherever possible. The overarching areas of consideration are:

- Ethical review of provenance and sourcing of the material
- Legality of collection, transfer and use (national and international)

Each transfer of samples is further undertaken according to a Research Collaboration Agreement or Material Transfer Agreement entered into by the Darwin Tree of Life Partner, Genome Research Limited (operating as the Wellcome Sanger

Table 3. Software tools: versions and sources.

Software tool	Version	Source
BlobToolKit	4.1.7	https://github.com/blobtoolkit/blobtoolkit
BUSCO	5.3.2	https://gitlab.com/ezlab/busco
Hifiasm	0.16.1-r375	https://gitlab.com/chhylp123/hifiasm
HiGlass	1.11.6	https://gitlab.com/higlass/higlass
Merqury	MerquryFK	https://gitlab.com/thegenemyers/MERQURY.FK
MitoHiFi	2	https://gitlab.com/marcelauliano/MitoHiFi
PretextView	0.2	https://gitlab.com/wtsi-hpag/PretextView
purge_dups	1.2.3	https://gitlab.com/dfguan/purge_dups
sanger-tol/genomenote	v1.0	https://gitlab.com/sanger-tol/genomenote
sanger-tol/readmapping	1.1.0	https://gitlab.com/sanger-tol/readmapping/tree/1.1.0
YaHS	yahs-1.1.91eebc2	https://gitlab.com/c-zhou/yahs

Institute), and in some circumstances other Darwin Tree of Life collaborators.

Data availability

European Nucleotide Archive: *Tausonia pullulans*. Accession number PRJEB61351; https://identifiers.org/ena.embl/PRJEB61351. The genome sequence is released openly for reuse. The *Tausonia pullulans* genome sequencing initiative is part of the Darwin Tree of Life (DToL) project. All raw sequence data and the assembly have been deposited in INSDC databases. The genome will be annotated using available RNA-Seq data and presented through the Ensembl pipeline at the European Bioinformatics Institute. Raw data and assembly accession identifiers are reported in Table 1.

Author information

Members of the Royal Botanic Gardens Kew Genome Acquisition Lab are listed here: https://doi.org/10.5281/zenodo.4786680.

Members of the Darwin Tree of Life Barcoding collective are listed here: https://doi.org/10.5281/zenodo.4893703.

Members of the Wellcome Sanger Institute Tree of Life Management, Samples and Laboratory team are listed here: https://doi.org/10.5281/zenodo.10066175.

Members of Wellcome Sanger Institute Scientific Operations: Sequencing Operations are listed here: https://doi.org/10.5281/zenodo.10043364.

Members of the Wellcome Sanger Institute Tree of Life Core Informatics team are listed here: https://doi.org/10.5281/zenodo.10066637.

Members of the Tree of Life Core Informatics collective are listed here: https://doi.org/10.5281/zenodo.5013541.

Members of the Darwin Tree of Life Consortium are listed here: https://doi.org/10.5281/zenodo.4783558.

References

Abdennur N, Mirny LA: **Cooler: scalable storage for Hi-C data and other genomically labeled arrays.** *Bioinformatics.* 2020; **36**(1): 311–316.
PubMed Abstract | Publisher Full Text | Free Full Text

Allio R, Schomaker-Bastos A, Romiguier J, *et al.*: **MitoFinder: efficient automated large-scale extraction of mitogenomic data in target enrichment phylogenomics.** *Mol Ecol Resour.* 2020; **20**(4): 892–905.
PubMed Abstract | Publisher Full Text | Free Full Text

Bezus B, Esquivel JCC, Cavalitto S, *et al.*: **Pectin extraction from lime pomace by cold-active polygalacturonase-assisted method.** *Int J Biol Macromol.* 2022; **209**(Pt A): 290–298.
PubMed Abstract | Publisher Full Text

Cadete RM, Lopes MR, Rosa CA: **Yeasts associated with decomposing plant material and rotting wood.** *Yeasts in natural ecosystems: diversity.* 2017; 265–292.
Publisher Full Text

Challis R, Richards E, Rajan J, *et al.*: **BlobToolKit – interactive quality assessment of genome assemblies.** *G3 (Bethesda).* 2020; **10**(4): 1361–1374.
PubMed Abstract | Publisher Full Text | Free Full Text

Cheng H, Concepcion GT, Feng X, *et al.*: **Haplotype-resolved *de novo* assembly using phased assembly graphs with hifiasm.** *Nat Methods.* 2021; **18**(2): 170–175.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Dennis C: **Breakdown of cellulose by Yeast Species.** *Microbiology.* 1972; **71**(2): 409–411.
Reference Source

Denton A, Yatsenko H, Jay J, *et al.*: **Sanger Tree of Life wet laboratory protocol collection V.1.** *protocols.io.* 2023.
Publisher Full Text

Di Tommaso P, Chatzou M, Floden EW, *et al.*: **Nextflow enables reproducible computational workflows.** *Nat Biotechnol.* 2017; **35**(4): 316–319.
PubMed Abstract | **Publisher Full Text**

do Amaral RJV, Bates A, Denton A, *et al.*: **Sanger Tree of Life RNA extraction: automated MagMax™ mirVana.** *protocols.io.* 2023.
Publisher Full Text

Fell JW, Guého-Kellermann E: ***Guehomyces* Fell & Scorzetti (2004).** In: *The Yeasts.* 2011; 1773–1775.
Publisher Full Text

Fell JW, Scorzetti G: **Reassignment of the basidiomycetous yeasts *Trichosporon pullulans* to *Guehomyces pullulans* gen. nov., comb. nov. and *Hyalodendron lignicola* to *Trichosporon lignicola* comb. nov.** *Int J Syst Evol Microbiol.* 2004; **54**(Pt 3): 995–998.
PubMed Abstract | **Publisher Full Text**

GBIF: **GBIF occurrence download - Tausonia pullulans.** *GBIF Occurrence Download.* (August 31, 2023); [Accessed 31 August 2023].
Publisher Full Text

Guan D, McCarthy SA, Wood J, *et al.*: **Identifying and removing haplotypic duplication in primary genome assemblies.** *Bioinformatics.* 2020; **36**(9): 2896–2898.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Harry E: **PretextView (Paired REad TEXTure Viewer): a desktop application for viewing pretext contact maps**. 2022.
Reference Source

Howe K, Chow W, Collins J, *et al.*: **Significantly improving the quality of genome assemblies through curation.** *GigaScience.* 2021; **10**(1): giaa153.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Jay J, Yatsenko H, Narváez-Gómez JP, *et al.*: **Sanger Tree of Life sample preparation: triage and dissection.** *protocols.io.* 2023.
Publisher Full Text

Kerpedjiev P, Abdennur N, Lekschas F, *et al.*: **HiGlass: web-based visual exploration and analysis of genome interaction maps.** *Genome Biol.* 2018; **19**(1): 125.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Kim MJ, Lee HW, Kim JY, *et al.*: **Impact of fermentation conditions on the diversity of White Colony-Forming Yeast and analysis of metabolite changes by White Colony-Forming Yeast in Kimchi.** *Food Res Int.* 2020; **136**: 109315.
PubMed Abstract | **Publisher Full Text**

Lindner P: **Mikroskopische Betriebskontrolle in den Gärungsgewerben.** Open Library, 1901.
Reference Source

Liu XZ, Wang QM, Göker M, *et al.*: **Towards an integrated phylogenetic classification of the Tremellomycetes.** *Stud Mycol.* 2015; **81**: 85–147.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Manni M, Berkeley MR, Seppey M, *et al.*: **BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes.** *Mol Biol Evol.* 2021; **38**(10): 4647–4654.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Moylett EH, Chinen J, Shearer WT: **Trichosporon pullulans infection in 2 patients with chronic granulomatous disease: an emerging pathogen and review of the literature.** *J Allergy Clin Immunol.* 2003; **111**(6): 1370–1374.
PubMed Abstract | **Publisher Full Text**

Nakagawa T, Ikehata R, Uchino M, *et al.*: **Cold-active acid beta-galactosidase activity of isolated psychrophilic-basidiomycetous yeast *Guehomyces pullulans*.** *Microbiol Res.* 2006; **161**(1): 75–79.
PubMed Abstract | **Publisher Full Text**

Narváez-Gómez JP, Mbye H, Oatley G, *et al.*: **Sanger Tree of Life sample homogenisation: covaris cryoPREP® automated dry Pulverizer V.1 .** *protocols.io.* 2023.
Publisher Full Text

Rao SSP, Huntley MH, Durand NC, *et al.*: **A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping.** *Cell.* 2014; **159**(7): 1665–1680.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Rhie A, McCarthy SA, Fedrigo O, *et al.*: **Towards complete and error-free genome assemblies of all vertebrate species.** *Nature.* 2021; **592**(7856): 737–746.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Rhie A, Walenz BP, Koren S, *et al.*: **Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies.** *Genome Biol.* 2020; **21**(1): 245.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Romero MC, Hammer E, Cazau MC, *et al.*: **Isolation and characterization of biarylic structure-degrading yeasts: hydroxylation potential of dibenzofuran.** *Environ Pollut.* 2002; **118**(3): 379–382.
PubMed Abstract | **Publisher Full Text**

Sampaio JP: ***Tausonia* Bab'eva (1998).** In: *The Yeasts.* 2011; 1999–2001.
Publisher Full Text

Sláviková E, Košíková B, Mikulášova M: **Biotransformation of waste lignin products by the soil-inhabiting yeast *Trichosporon pullulans*.** *Can J Microbiol.* 2002; **48**(3): 200–203.
PubMed Abstract | **Publisher Full Text**

Strickland M, Cornwell C, Howard C: **Sanger Tree of Life fragmented DNA clean up: manual SPRI.** *protocols.io.* 2023.
Publisher Full Text

Surana P, Muffato M, Qi G: **Sanger-tol/readmapping: sanger-tol/readmapping v1.1.0 - Hebridean Black (1.1.0).** *Zenodo.* 2023a.
Publisher Full Text

Surana P, Muffato M, Sadasivan Baby C: **sanger-tol/genomenote (v1.0.dev).** *Zenodo.* 2023b.
Publisher Full Text

Trochine A, Bellora N, Nizovoy P, *et al.*: **Genomic and proteomic analysis of *Tausonia pullulans* reveals a key role for a GH15 glucoamylase in starch hydrolysis.** *Appl Microbiol Biotechnol.* 2022a; **106**(12): 4655–4667.
PubMed Abstract | **Publisher Full Text**

Trochine A, Bellora N, Nizovoy P, *et al.*: **Genomic and proteomic analysis of *Tausonia pullulans* reveals a key role for a GH15 glucoamylase in starch hydrolysis.** *Appl Microbiol Biotechnol.* 2022b; **106**(12): 4655–4667.
PubMed Abstract | **Publisher Full Text**

Todorovic M, Oatley G, Howard C: **Sanger Tree of Life HMW DNA Extraction: Automated Plant MagAttract v.2, *protocols.io*.** 2023a.
Publisher Full Text

Todorovic M, Sampaio F, Howard C: **Sanger Tree of Life HMW DNA fragmentation: diagenode Megaruptor®3 for PacBio HiFi.** *protocols.io.* 2023b.
Publisher Full Text

Uliano-Silva M, Ferreira JGRN, Krasheninnikova K, *et al.*: **MitoHiFi: a python pipeline for mitochondrial genome assembly from PacBio high fidelity reads.** *BMC Bioinformatics.* 2023; **24**(1): 288.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Vasimuddin M, Misra S, Li H, *et al.*: **Efficient architecture-aware acceleration of BWA-MEM for multicore systems.** In: *2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS).* IEEE, 2019; 314–324.
Publisher Full Text

Yao H, Tian S, Wang Y: **Sodium bicarbonate enhances biocontrol efficacy of yeasts on fungal spoilage of pears.** *Int J Food Microbiol.* 2004; **93**(3): 297–304.
PubMed Abstract | **Publisher Full Text**

Yurkov AM: **Yeasts of the soil - obscure but precious.** *Yeast.* 2018; **35**(5): 369–378.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Zhou C, McCarthy SA, Durbin R: **YaHS: yet another Hi-C scaffolding tool.** *Bioinformatics.* 2023; **39**(1): btac808.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Open Peer Review

Current Peer Review Status: ? ✓ ?

Version 1

Reviewer Report 31 December 2024

https://doi.org/10.21956/wellcomeopenres.25735.r114597

?

David Diaz Escandon
University of Alberta, Edmonton, Alberta, Canada

The paper presents a—chromosome-level—genome release for the basidiomycete yeast, *Tausonia pullulans*. The authors properly explained the rationale behind studying this organism, particularly given its recent reclassification to Cystofilobasidiales, and its nearly cosmopolitan distribution. The paper underscores the value of expanding our foundational knowledge of this yeast with this newly released genome, especially its enriched enzymatic machinery previously described in the literature and already applied to various industrial processes. The authors present a 23 Mb chromosome-level genome assembly derived from a single culture, utilizing state-of-the-art DNA extraction, sequencing, and assembly techniques, together with an 18 Kb mitochondrial genome. This represents the first and only chromosome-level assembly for this species; however, it is worth noting that four other genomes are available on NCBI.

High-quality reference genome releases are highly demanded in the current era of rapidly advancing sequencing technologies. Chromosome-level assemblies have become the gold standard, providing a necessary framework for revising and corroborating scaffold-based assemblies. This premise alone adds significant merit to the paper. However, there are a few critical issues that need to be addressed:

1. **Lack of genome/species-specific analyses in the genome release**: While this is a genome release, its current state does not differ significantly from the information currently available on NCBI or EBI. Although additional information is provided in the methods section, many of the protocols referenced are for bulk sequencing/analysis and published in 'protocols.io' and Nextflow. The lack of genome-specific discussion or analysis is a notable limitation. This could be remedied by including comparative analyses with the other available genomes or by performing additional steps, such as genome annotation. Personalized genome releases should include unique insights or findings specific to the genome in question. This might seem trivial, but at this stage it seems like a personalized genome release of a bulk generated genome.

2. **Low BUSCO Completeness**: The reported BUSCO completeness of 29.9% (benchmark >95%) using TremellomycetesODB10 requires further discussion. While the paper asserts

completeness in various ways, the current standard for genome completeness is based on BUSCO metrics, and a value of 29.9% needs to be either justified or explained. Completeness scores at the class level are often lower due to under-sampling in some classes or order selection biassed, but a value below 50% raises questions about either the genome's completeness or its taxonomic assignment unless additional context is provided. That said, It is worth noting that the TremellomycetesODB10 dataset primarily represents *Cryptococcus* and includes only 18 reference genomes, most of them from Tremellales. This narrow representation suggests that any major evolutionary shifts between orders may not be well captured. To strengthen the paper, I recommend including an assessment using BasidiomycotaODB10, which should ideally score close to 100% (92% with Busco 5.8.2 and BasidiomycotaODB10), or incorporating a secondary completeness metric such as EukCC. It would be worth discussing this BUSCO results, regards of the other metrics included, since TremellomyctesODB10 includes*Trichosporon* (Trichosporonales), *Tausonia* former classification. Additionally BUSCO completeness using version 5.8.2 and the latest updated of Tremellomycetes shows a higher completeness with 46.6%.

3. **Unutilized RNA Sequencing Data**: The inclusion of RNA extraction and sequencing methods in the paper, without corresponding results or analysis, is problematic. If this is intended as a personalized genome release, the paper would benefit greatly from leveraging this data. As a rule of thumb, data mentioned in the methods should be used in the results. In this case, the RNA sequencing data is mentioned only in "Data Availability", with the suggestion that it will be used for future annotations. It would be more appropriate to either complete the annotations and include them in this genome release or exclude the RNA sequencing procedure entirely from the current manuscript.

Is the rationale for creating the dataset(s) clearly described?

Yes

Are the protocols appropriate and is the work technically sound?

Yes

Are sufficient details of methods and materials provided to allow replication by others?

Partly

Are the datasets clearly presented in a useable and accessible format?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Lichens, Genomics, Evolution

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Reviewer Report 26 December 2024

https://doi.org/10.21956/wellcomeopenres.25735.r114599

Reyim Mamut
Xinjiang University, Urumqi, Xinjiang, China

In terms of the significance of this article, it is delivers an important contribution to the genome assembly of *Tausonia pullulans*. Basidiomycete yeast have wide application value in industrial and agricultural production, biopharmaceuticals, and environmental protection. For example, they play an important systematic role in fungal evolution and have significant potential applications in developing low-temperature industrial enzymes and constructing low-temperature exogenous protein expression systems. Therefore, studying its genome is a necessary task.
The author provided a detailed Genome sequence report, outlining the standard process for obtaining genomes of *Tausonia pullulans. The methods:* Sample acquisition and nucleic acid extraction,Sequencing,Genome assembly, used in the article are reasonable and reflect a high level of genomic research.
Recently, the number of species of basidiomycetes yeast has been increasing, and some new groups with special morphologies have been discovered, which have filled many gaps in the phylogenetic tree.
Therefore, in my opinion, the section on phylogenetic analyses is needed. Phylogenetic analyses could clarify the evolutionary relationships of *T. pullulans* within the Tremellomycetes. The improvement of this part of the content will help to better understand of *T. pullulans*. On the other hand, it will reflect the value of genomic research.

Is the rationale for creating the dataset(s) clearly described?
Yes

Are the protocols appropriate and is the work technically sound?
Partly

Are sufficient details of methods and materials provided to allow replication by others?
Yes

Are the datasets clearly presented in a useable and accessible format?
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Taxonomy, Lichenology and Bioinformatic

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Reviewer Report 10 December 2024

https://doi.org/10.21956/wellcomeopenres.25735.r113616

?

David Pizarro
Universidad Complutense de Madrid, Madrid, Community of Madrid, Spain

The article delivers an important contribution by presenting the genome assembly of *Tausonia pullulans*, a yeast species with notable ecological and industrial relevance. The authors have successfully assembled a genome spanning 23.9 Mb, which includes 20 chromosomal pseudomolecules and a complete mitochondrial genome. Employing cutting-edge sequencing techniques like PacBio HiFi reads and Hi-C scaffolding, combined with manual curation, highlights the thoroughness of the work. Additionally, integrating the genome within the broader context of the yeast's enzymatic and ecological capabilities provides meaningful insights, especially for potential biotechnological applications such as lignocellulose degradation and pollutant processing. The availability of raw data and genome assembly files in public repositories also reinforces the accessibility and reproducibility of the study, which are significant strengths.
Despite these contributions, the study faces a critical challenge in the form of its low genome completeness, as reflected by a BUSCO score of 29.9%. This figure, which is well below the typical standard for reference genomes, raises questions about the representativeness of the assembly. Missing conserved single-copy orthologs may indicate technical challenges, such as an inability to resolve complex or repetitive genomic regions, or errors during assembly that led to the collapse of haplotypes. Alternatively, *T. pullulans* may exhibit biological genome reduction, as seen in certain other fungal species. If the latter is the case, presenting supporting evidence for such reduction would greatly enhance the study's impact. To address these issues, a reanalysis of the data using alternative pipelines, such as tools that are better suited for resolving heterozygosity or repetitive sequences, could improve genome completeness. Incorporating additional sequencing technologies, like ultra-long reads, may also help resolve ambiguities and gaps.
The functional annotation of the genome could be further developed to strengthen the study's claims about enzymatic activities and their biotechnological potential. While the manuscript references the yeast's role in processes like starch hydrolysis and lignocellulose degradation, these assertions are primarily based on previous research rather than novel findings derived directly from the genome. Enhanced annotation using tools such as InterProScan or KEGG would allow the authors to identify specific genes or pathways tied to these activities, offering new insights into the functional capabilities of *T. pullulans*. Furthermore, improving the annotation could help clarify whether the low BUSCO score is due to assembly gaps or a lack of functional gene content.
Another area for improvement is the evolutionary context provided in the study. While the authors offer a comprehensive account of the species' taxonomic history, the manuscript could be significantly enriched by using the assembled genome for phylogenetic analyses. Such analyses could clarify the evolutionary relationships of *T. pullulans* within the Tremellomycetes and identify traits or adaptations that distinguish it from closely related species. Insights into genome architecture, potential signs of metabolic specialization, or evidence of genome reduction could make the study more impactful.
Moreover, the data visualization and interpretation within the manuscript require greater attention. Figures such as BlobToolKit plots and Hi-C maps are useful tools for assessing genome assembly quality but are not sufficiently analyzed in the current text. Offering more detailed

explanations of these visualizations and linking them to broader conclusions about the genome's structure and organization would help readers better understand the study's findings. Expanding on the mitochondrial genome analysis could also reveal interesting aspects of the species' biology or adaptations.
In conclusion, while the genome assembly of *T. pullulans* presented in this article is a valuable resource, addressing the low BUSCO score, improving functional annotation, integrating phylogenetic analyses, and enhancing data interpretation are essential steps. These revisions would elevate the study, ensuring it becomes a robust and widely useful reference for fungal genomics and biotechnology. By tackling these challenges, the article can fulfill its potential to significantly advance research in these fields.

Is the rationale for creating the dataset(s) clearly described?
Yes

Are the protocols appropriate and is the work technically sound?
Partly

Are sufficient details of methods and materials provided to allow replication by others?
Yes

Are the datasets clearly presented in a useable and accessible format?
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Genomic, Lichenology and Bioinformatic

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.