First Report on Genome assembly of Talaromyces trachyspermus strain 4014 isolated from Withania somnifera

Kishore K Shende
Barkatullah University
Mahesh Dharne
National Chemical Laboratory
Karthika Nair
National Chemical Laboratory
Vinay Rajput
National Chemical Laboratory
Rachel Samson
National Chemical Laboratory
Sharda Sahu
Swami Atmanand Govt. English Medium School
Krishna Mohan Medicherla
Jaipur National University
Prashanth Suravajhala
prash.s@jaipur.manipal.edu

Manipal University Jaipur
Anil Prakash
Barkatullah University

Research Article

Keywords: Talaromyces trachyspermus, Whitania somnifera, Endophyte, Fungal Genome, Mitochondria Genome

Posted Date: May 4th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9244707/v1

Additional Declarations: No competing interests reported.

First Report on Genome assembly of *Talaromyces trachyspermus* strain 4014 isolated from *Withania somnifera*

Kishor K. Shende,[a] Mahesh Dharne,[b] Karthika Nair,[b] Vinay Rajput,[b] Rachel Samson,[b] Sharda Sahu,[c] Krishna Mohan Medicherla,[d] Prashanth Suravajhala,[e] Anil Prakash,[fg]#

[a]Department of Biotechnology and Bioinformatics Center (SubDIC), Barkatullah University, Bhopal, M.P., India. kishorekshende@gmail.com ORCID ID: 0000-0003-4239-0206
[b]NCIM Resource Center, NCL, Pune, M.S., India. ms.dharne@ncl.res.in ORCID ID: 0000-0002-3965-7320
[c]Swami Atmanand Govt. English Medium School, Bilaspur, CG, India. drsharda85@gmail.com
[d]Birla Institute of Scientific Research, Jaipur, Rajasthan, India. mkrishnamohan@gmail.com ORCID ID: 0000-0001-7099-7721
[e]Amrita School of Biotechnology, Amrita Vishwa Vidyapeetham, Clappana 690525, Kerala, India. prash@bioclues.org ORCID ID: 0000-0002-8535-278X
[f]Department of Microbiology, Barkatullah University, Bhopal, M.P., India
[g]Centre for Sustainable Agricultural and Environment, Prof. H.S. Srivastava Foundation for Science and Society, Uttarathia, Lucknow, U.P., India. dranilprakash98@gmail.com ORCID ID: 0000-0003-1527-7611

#Corresponding Author: Anil Prakash, dranilprakash98@gmail.com
*Present address: Centre for Sustainable Agricultural and Environment, Prof. H.S. Srivastava Foundation for Science and Society, Uttarathia, Lucknow, U.P., India

ABSTRACT

The genus *Talaromyces* encompasses fungal species inhabiting various environments, including soil, marine sponges, food materials, and plant parts. *Talaromyces trachyspermus* 4014, a filamentous fungus isolated from the leaves of the medicinal plant *Withania somnifera*, exhibits plant growth-promoting and pathogen-antagonistic properties. This study presents the first *de novo* whole genome sequence assembly of *T. trachyspermus* 4014, including the complete mitochondrial genome. The strain was confirmed based on ITS region homology, showing a close relation to *T. assuitensis, T. gossypii*, and *T. aurantiacus*. The assembled genome size is approximately 29 mb with 49.3% GC content, comprising 10,892 protein-coding genes. The key genes involved in IAA production, siderophore biosynthesis, and various hydrolases were identified, highlighting the strain's potential for agricultural applications. Comparative analysis revealed its mitochondrial genome's high similarity with *T. marneffei*. This comprehensive genomic data enhances our understanding of *T. trachyspermus* 4014's beneficial traits, laying the groundwork for future research and biotechnological exploitation.

KEYWORDS: *Talaromyces trachyspermus*, *Whitania somnifera*, Endophyte, Fungal Genome, Mitochondria Genome

INTRODUCTION

The genus *Talaromcyes* comprises fungal species that inhabit diverse environments such as soil and forest soil [1, 2], marine sponges [3]; food materials [4]; House dust [5], HIV/AIDS patient [6], decaying wood [7] , plants parts [8-10] exploring life cycle either as a saprophytic, symbiotic, pathogenic or combination and survive in diverse ecological niche. *Talaromyces* genus is a group of species with soft ascospore, interwoven hyphae, and yellow ascomata [11], later included the imperfect species of subgenus *Biverticillium* of genus *Penicillium* [12, 13] and also characterized based on internal transcribed spacer (ITS) sequence, small and large rRNA [14, 15]. The filamentous fungi, *Talaromyces trachyspermus* species belongs to the kingdom: Fungi, sub-kingdom: Dikarya, phylum: Ascomycota (sac fungi), sub-phylum: Pezizomycotina, class: Eurotiomycetes, order: Eurotials, family: Trichocomaceae and genus: *Talaromyces* that produces asci in chain [15]. Talaromyces species are industrially [16, 17], environmentally [18], and medically [19] important and have the potential for the production of various bioactive molecules. It is reported that around 221 bioactive molecules are produced by the species of genus *Talaromyces* composed of alkaloids, peptides, esters, polyketides, quinones, steroids, and terpenoids [20].

An endophytic fungi, *Talaromyces trachyspermus* 4014 was isolated from the leaves of the medicinal plant *Withania somnifera* obtained from Vindhya Herbal, Barkhera Pathani, Bhopal, Madhya Pradesh, India [9]. The endophytic fungi in symbiotic association with plants stimulate the synthesis of phytohormones and bioactive molecules [21, 22]. In our study on *T. trachyspermus* 4014, a fungal strain was detected to produce IAA, siderophore, protease, cellulase, chitinase, amylase, lipase, and phosphatase, which indicates the plant growth-promoting characteristics [9]. The antagonistic activity was shown by this strain against plant pathogens [9] as also reported by Dethoup et al. [23] in case of *T. trachyspermus* strain KUFA 0021.

We present the first report on the de novo whole genome sequence assembly of *T. trachyspermus* 4014, including the complete mitochondrial genome sequence. The identification of the *T. trachyspermus* strain was confirmed based on ITS region homology, showing a close relation to *T. assuitensis*, *T. gossypii*, and *T. aurantiacus*. Mitochondrial genome comparison revealed its closeness to *T. Marneffeia* and *T. pinophilus*. This study focuses on genome assembly, gene prediction, annotation, metabolic pathway prediction, and analysis of predicted proteins involved in the processes of plant growth promotion and antagonistic properties. The related data are submitted to the NCBI genome database. The data related to BioSample: SAMN15692527, BioProject: PRJNA650128 and genome: JACGXQ010000001- JACGXQ010000015 and mitochondrial genome: CM035349 are submitted to NCBI genome database (https://ncbi.nlm.nih.gov/genome, submission date: 05-10-2021)

MATERIALS AND METHODS

1. Fungal strain Isolation and Culturing

The wild plant of *Whitania somnifera* (L) was obtained from the nursery of Vindhay Herbal, a Madhya Pradesh State Forest Produce (T & D) Co. Fed., Bhopal in the year 2015 and planted at Department of Microbiology, Barkatullah University, Bhopal. The plant species was identified by Vindhya Herbal and by Prof. Anil Prakash at Department of Microbiology, Barkatullah University. The leaf samples were surface sterilized using a series of washes in 70% ethanol for 1 minute, 2% sodium hypochlorite for 5 minutes, and rinsed three times in sterile distilled water. The surface-sterilized leaves were cut into small segments (5 mm x 5 mm) and placed on potato dextrose agar (PDA) plates supplemented with chloramphenicol (100 µg/mL) to inhibit bacterial growth. The plates were incubated at 28°C for 7 days. The emerging fungal colonies were subcultured on fresh PDA plates to obtain pure cultures. The potential endophytic fungal strain was identified at National Fungal Culture Collection of India (NFCCI), Agarkar Research Institute, Pune, India; as *Talaromyces trachyspermus.* The fungal strain was submitted to NFCCI, Pune with an accession number NFCCI4014.

2. Genomic DNA Extraction and Sequencing

Genomic DNA was extracted from the mycelium of *T. trachyspermus* 4014 grown in liquid PDA medium for 5 days at 28°C. The DNA was extracted from the around 250mg fungal biomass using DNeasy PowerSoil Pro Kit (Qiagen) according to the protocol. The quality and concentration of the DNA was assessed using Nanodrop Lite Spectrophotometer (Thermo Fisher Scientific), and Qubit fluorometer (Thermo Fisher Scientific). The extracted DNA was subjected to nanopore sequencing. The library preparation was carried out following the manufacturer's protocol from Oxford Nanopore Technology (ONT). Briefly, 400 ng of the extracted DNA was used in triplicates for end-repair and barcoding. Each replicate was barcoded independently using Native Barcoding Kit 24 V14 (SQK-NBD114.24, ONT). Following this, the barcoded samples were adapter ligated and then after elution, the library was loaded onto the flow cell (FLO-MIN-106, R9 Version). The sequencing platform used for this experiment was MinION™ Mk1B (ONT), with Parallel Basecalling, FAST basecalling mode of GUPPY5.0.1.

3. Genome Assembly and Annotation

The Quality of the Oxford Nanopore raw reads was analyzed by FastQC 0.11.8 [24] software tool. The read correction, trimming, and assembly were carried out by Canu v2.0 assembler [25] with default parameters and a minimum read length of 1000 bps Genome completeness was checked by BUSCOv5 [26] and CEGMA [27]. The assembly quality was assessed using QUAST v5.2 [28].

The genome assembly scaffolds were annotated by the Genome Sequence Annotation Server (GenSAS) [29]. Genome completeness was checked by BUSCO v3.1.0 software [30]. Genomic repeat regions were analyzed by RepeatMasker v4.1.1 [31] and de novo repeat masker, RepeatModeler v2.0.1. [32]. The tools, Augustus 3.3.1 [33], GeneMarkES 4.48 [34], and GlimmerM 2.5.1 [35] were used for gene prediction. The consensus of predicted genes was obtained using the Evidence Modeler 1.1.1 [36] and refined by PASA 2.4.1 [37]. Further, the predicted genes were aligned with the set of reference fungal proteins from the UniProtKB database by BLASTX [38] and DIAMOND 2.0.6 [39] and NCBI fungal transcript sequences by BLASTN 2.11.0 [38], PASA 2.4.1 [37] and pBLAT [40]. Functional analysis of the official gene set was done by BLASTP 2.11.0 [38], DIAMOND 2.0.6 [39], InterProScan 4.44-79.0 [41], Pfam 1.6 [42] analysis and SignalP 5.0b [43]. Ribosomal RNA and tRNA were predicted by the tool, RNAmmer [44] and tRNAScan-SE [45], respectively.

The mitochondrial genome was separated from the whole genome assembly as contig6 by the virtue of circularization detected by CANU software and its low GC content (25.3 %). The initial assembly, contig6 of a length of 58664 bps was analyzed using the VISTA, PipMaker [46], and Mauve [47] for duplication of genomic regions that were removed to get the final assembly of length 38945 bps.

4. Bioinformatics Analysis

Functional categorization of the predicted proteins was conducted using the Gene Ontology (GO) database. Metabolic pathways were reconstructed using the Kyoto Encyclopedia of Genes and Genomes (KEGG), KO (KEGG Orthology) assignment by BlastKOALA and GhostKOALA [48]. The gene clusters of secondary metabolite synthesis were predicted by antiSMASH 8.0 [49]. Phylogenetic analysis was performed using the ITS region sequences and mitochondrial genome sequence of *T. trachyspermus* 4014 and closely related species. Multiple sequence alignment was conducted using ClustalW, and a phylogenetic tree was constructed using MEGA X with the Maximum Likelihood method with 1,000 bootstrap replications.

5. Plant Growth-Promoting Traits and Antagonistic Activity

The production of indole-3-acetic acid (IAA) by *T. trachyspermus* 4014 was quantified using the Salkowski reagent method. Siderophore production was assessed using Chrome Azurol S (CAS) agar plates. The production of hydrolytic enzymes (protease, cellulase, chitinase, amylase, lipase, and phosphatase) was determined using specific substrate-based assays. Antagonistic activity against plant pathogenic fungi was

evaluated using a dual culture assay. *T. trachyspermus* 4014 and the test pathogen were inoculated at opposite sides of a PDA plate and incubated at 28°C for 7 days. The inhibition zone was measured to assess the antagonistic effect.

6. Data Submission

GENSAS generated output files, '.gff3', 'interproscan', and protein BLAST output file were processed along with reference protein sequences of SWISS-PROT databases by the software tool 'ANNIE' [50] to extract the domain and motif annotation, gene functions, and protein products. The annotation output file of Annie program, along with '.gff3' and genome assembly fasta file was processed by the GAG [51] tool to generate the table format (.tbl) file that fed to program table2asn [52] program along with genome assembly file to generate the NCBI genome database submission format '.asn' The whole genome sequence of *T. trachyspermus* 4014 has been submitted to the NCBI genome database under accession number JACGXQ010000001-JACGXQ010000015; Mitochondrial genome: CM035349. The raw sequencing reads are available in the NCBI Sequence Read Archive (SRA) under accession number SRR25836654.

RESULTS

Quality Analysis of Nanopore Reads

The Quality of a total of 706979 reads were assessed by the FastQC tool. The reads were of length in the range of 7- 48161 bps and composed of 46% GC content. The quality score per sequence read is low in the range of 5-22 (average = 14) for the sequences with length <28000bp. The read sequences of length >28000 are found to be of good quality.

Raw read correction, trimming, and genome assembly

The reads were corrected 5 rounds and trimmed by the Canu v2.0 assembler [25]. The final round of read correction was operated with the parameter as minimum reads length 1000bp, kmer: 16 mers, and the rest of the other parameters as default. The total 462154 reads (1993769355 bases) with 56.96 times coverage of the genome size 35 MB (approximate), consists of 427318 reads with overlap and 34836 reads without overlap. The read correction step was followed by the trimming step with the parameters as, fraction error of 0.12, deletion of trimmed reads <1000 bps, and overlap region length is kept default. The result was 166383 trimmed reads and 44103 reads without any change. A total of 251668 reads of poor quality were deleted in the previous read correction step and 210486 reads (1373146570 bases) with 38.55 times coverage were further processed for de novo genome assembly by Canu v2.0 assembler. The unitigging was done with 22 mers parameter and keeping other parameters as default. The final consensus assembly obtained consists of 15 scaffolds of total length, 32096759 bp including one repeat region of length 10842 bp indicating that 8 chromosomes are carved out from 15 scaffolds. A total of 20677 sequences remained unassembled. The assembled genome attributes are presented in Table 1.

Table 1 Assembly Matrix of *Talaromyces trachyspemus* 4014 Genome

Property	*Talaromyces trachyspermus* strain 4014
Number of scaffolds	15
Total size of scaffolds	32096759
Total scaffold length as a percentage of assumed genome size (35 mb)	91.71%

useful amount of scaffold sequences (>=25K nt)	32038095
Longest scaffold	10530633
Shortest scaffold	9286
Number of scaffolds > 1K nt	15
Number of scaffolds > 10K nt	14
Number of scaffolds > 100K nt	9
Number of scaffolds > 1M nt	8
Number of scaffolds > 10M nt	1
N50	3814881
L50	3
NG50	3814881
LG50	3
A%	26.55
C%	23.49
G%	23.50
T%	26.46
N%	0

Genome Completeness

The completeness of genome assembly was evaluated by BUSCO v5 software [30, 53]. The BUSCO v5 analysis for queried 758 core genes against the nuclear genome assembly, it is observed that 576 (75.99%) are complete genes, 96 are (12.7%) partial genes and 86 (11.35%) genes are missing. 0.87% of genes have more than 1 copy of orthologs. The completeness analysis by CEGMA based on 248 core genes detected the presence of 228 (91.94%) complete, 15 (6.04%) partial genes, and 5 (2.02%) missing genes. The number of ortholog copies is detected for 8.77% of genes.

Genomic attributes

A total of 13 contigs are rich in 45.47% GC content, but two contigs, contig7 JACGXQ010000006) and contig15 (JACGXQ0100000015) are composed of low GC content, 37.2%, and 19.9%, respectively. Contig-6 (JACGXQ010000005) was predicted to be a Mitochondrial genome composed of 25.3% GC content and a circular nature.

Genome Annotation

The nuclear genome scaffolds were annotated by the 'Genome Sequence Annotation Server' (GenSAS) (Humann et al. 2019) online annotation pipeline in a stepwise manner. The gene prediction programs selected are Augustus v4.1.1 [33], GeneMarkES v4.48 [34], and GlimmerM v2.5.1 [35]. The consensus of predicted genes was carried out by an Evidence Modeler and gene refinement by the PASA tool. The assembly genomics features of *T. Trachyspermus* 4014 are shown in Table 2. The pineline predicted a total of 11291 genes in 14 scaffolds and no gene was predicted in scaffold-13. The largest scaffold-15 (10.53 mb) contains 3781 (36.60%) genes, and the medium-size scaffolds (1, 2, 8, and 14) of size in the range 3.00 mb to 4.36 mb contain genes 1010 to 1536. The smaller scaffolds (3, 4, 11, and 12) of size ranging from 0.81 mb to 2.4 mb, contain the genes

from 287 to 849. The smallest scaffolds (6, 7, 9, 10, and 13) of size 9.2 kb to 18.44 kb contain the genes < 4 in numbers. BUSCOv5 software predicted the 88.69% completeness (75.99% complete and 12.7% partial) of the genome therefore the total gene content of *T. trachyspermus* genome may be estimated to be around 11800 - 12000 genes.

A single copy of 18S rRNA and 28S rRNA is predicted in the scaffold-2 and 12 copies of 5S rRNA are predicted in the scaffolds, 3, 4, 8, 11, 12, 14, and 15. A total of 97 tRNA was predicted in the scaffolds, 2, 3, 4, 8, 11, 12, 14, and 15. The maximum number of tRNAs i.e. 28 are predicted in scaffold, 15.

Table 2 *Talaromyces trachyspermus* 4014 Genome Assembly Features (GenBank Assemlby: GCA_020137715.1, Assembly name: BUMICRO_TalaroTrachy_1.1)

Contig ID	Accession number	Length	Gene freq	rRNA	tRNA	GC%	AT%	A%	C%	G%	T%
Nuclear DNA											
Contig-2-tig10	JACGXQ010000001.1	3743902	1309	2	22	47.37	52.63	26.39	23.73	23.64	26.24
Contig-3-tig15	JACGXQ010000002.1	3061800	1090	2	11	47.06	52.94	26.5	23.57	23.49	26.43
Contig-4-tig52	JACGXQ010000003.1	1328332	486	1	6	47.56	52.44	26.34	23.82	23.73	26.11
Contig-5-tig54	JACGXQ010000004.1	813339	287	1	1	46.96	53.04	26.75	23.57	23.39	26.29
Contig-7-tig308	JACGXQ010000006.1	18443	2	0	0	37.17	62.83	27.8	18.83	18.34	35.03
Contig-8-tig309	JACGXQ010000007.1	10842	2	0	0	45.93	54.07	25.06	22.51	23.43	29.01
Contig-9-tig310	JACGXQ010000008.1	3814881	1330	2	11	46.43	53.57	26.93	23.12	23.30	26.64
Contig-11-tig20989	JACGXQ010000009.1	15091	4	0	0	45.74	54.26	28.03	22.16	23.58	26.23
Contig-12-tig20990	JACGXQ010000010.1	9286	4	0	0	47.26	52.74	27.08	23.78	23.49	25.65
Contig-13-tig20991	JACGXQ010000011.1	2474372	849	2	6	46.59	53.41	26.67	23.28	23.32	26.74
Contig-14-	JACGXQ01000	18372	610	1	2	46	53.	26	23.	23	26

tig20992	0012.1	94				.32	68	.75	06	.26	.94
Contig-15-tig20993	JACGXQ010000013.1	10875	1	0	0	19.87	80.13	43.04	10.31	9.56	37.09
Contig-1-tig8	JACGXQ010000014.1	4369005	1536	0	10	47.17	52.83	26.41	23.60	23.57	26.42
Contig-10-tig20988	JACGXQ010000015.1	10530633	3781	2	28	47.24	52.76	26.40	23.62	23.62	26.36
	Total		**11291**	**13**	**97**						
Mitochondrial DNA											
Contig-6-tig72	CM035349.1	38945	2	0	48	25.31	74.69	36.6	10.89	14.41	38.10

Repeat Regions

Assembly sequence contigs were scanned for interspersed repeats and low complexity regions by the tools, RepeatMasker v4.1.1 [31] and de novo repeat masker, RepeatModeler v2.0.1. [32]. Total interspersed repeat regions of length 1754808 bp (5.48 %) were predicted in 14 scaffolds (Table 3). Different types of repeats predicted in the genomic scaffold are, (i) Retroelements (4.16%) comprised of LINEs (0.83%) and LTR elements (3.33%) and (ii) DNA transposons (0.74%), Tcl-IS630-Pogo (0.60%) and unclassified type (0.57%). The retroelements such as SINEs, Penelope, LINEs types such as CRE/SLACS, L2/CR1/Rex, R1/LOA/Jockey, R2/R4/NeSL, RTE/Bov-B, L1/CIN4, and DNA- transposons such as hobo-Activator, En-Spm, MuDR-IS905, PiggyBac, Tourist/Harbinger, Other (Mirage, P-element, Transib) and other types such as retroviral, rolling-circles, Small RNA, satellites regions, simple repeats, and low complexity regions were not predicted. Genome assembly analysis can be accessed from the Zenodo database [68].

Table 3 Description of different repeat regions predicted in the Scaffolds

Repeats Type	Number of copies	Total length	Percentage to genome
Retroelements	818	1333610 bp	4.16 %
LINEs	234	266080 bp	0.83%
LTR elements	584	1067530 bp	3.33%
Ty1/Copia	40	41933 bp	0.13%
Gypsy/DIRS1	544	1025597 bp	3.20%
DNA transposons	237	237713 bp	0.74%
Tc1-IS630-Pogo	187	198356 bp	0.62%
Unclassified	988	183485 bp	0.57%

Total interspersed repeats	1754808 bp	5.48%

Prediction of Secondary Metabolite Synthesizing Gene Cluster

The biosynthetic gene cluster (BGC) accommodated by the *Talaromcyes trachyspermus* genome, were predicted using the tool, "antibiotic and Secondary Metabolite Analysis SHell (antiSMASH 8.0) [49]. The secondary metabolite synthesis regions are not found in 5 scaffolds as JACGXQ020000006, JACGXQ020000007, JACGXQ020000009, JACGXQ020000010, and JACGXQ020000013. As described in Table 4, a total of 43 BGCs were predicted in 9 scaffolds, comprised of BGC types as, PKS or PKS like, NRPS or NRPS-like, terpene, betalactone and other (uncategorized). The BGC regions of Clavaric acid, YWA1 and Choline predicted with high similarity confidence, whereas the BGCs of Flavoglaucin, Metachelin (A,B,C, A-Ce), Dimerumic acid, Solanapyrone A, Zopfiellin, Sorbicillin, HEx-tc1 terpene, Phyllostictine (A and B), Squalestatin S1, Ustethylin A. A BGC of Phomoidride was predicted with medium level similarity confidence. Total 15 regions comprised of BGSs for polyketide, NRPS and terpene are predicted at the level of probable product synthesised. Total 28 regions of BGCs of type polyketide, NRPS, terpene and betalactone are probable BGCs with product synthesised is unclear. The antiSMASH analysis files are uploaded to Zenodo database [68].

Table 4 List of Secondary metabolite-producing gene cluster predicted in Genomics Scaffold of *Talaromyces trachyspermus* 4014

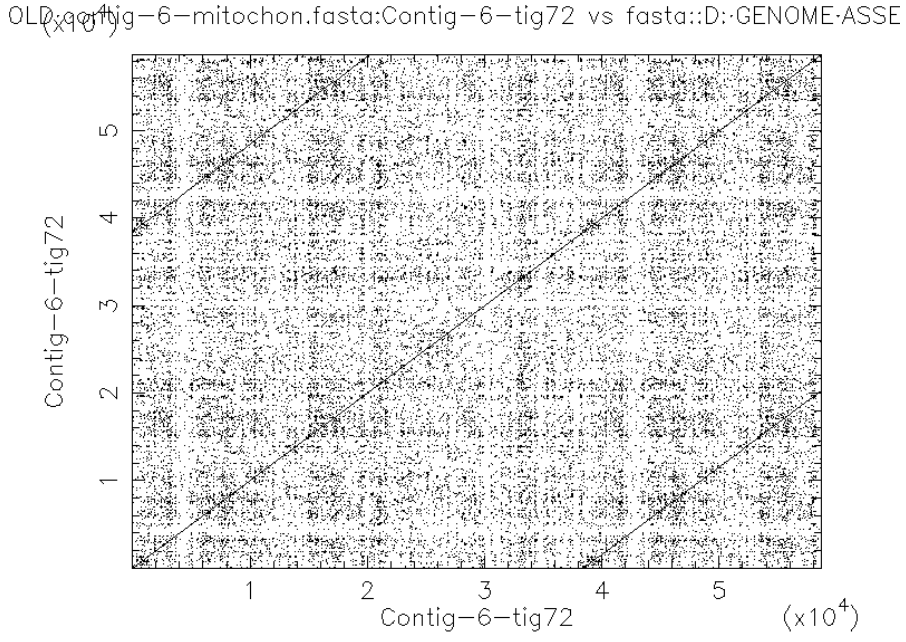
Scaffold no.	BGC type	Regions	Closest Compound from MiBIG database	BGC Type
1	T1PKS	31,094..93,431 2,747,614..2,814,865 2,824,697..2,919,449 3,177,620..3,326,279	Sorbicillin/ 2'3'-dihydrosorbicillin/ sorbicillinol/ oxosorbicillinol/ bisvertinolone/ dihydrobisvertinolone/ tetrahydrobisvertinolone/ bisorbicillinol/ 2'3'-dihydrosorbicillinol	PKS
1	NRPS and Betalactone	1,369,346..1,498,168 2,640,242..2,814,865	Choline	NRPS
1	Terpene	2,824,697..2,919,449	HEx-tc1 terpene	Terpene
2	NRPS like	515,837..582,004		NRPS
2	NRPS, NRPS like	694,189..768,585	Phyllostictine A and B	NRP + PKS
3	Terpene	9,9859..41,562 49,622..80,491	Squalestatin S1	Terpene
3	T1PKS and NRPS like	604,728..697,118 1,275,633..1,328,332	Ustethylin A	PKS
4	T1PKS and NRPS like	726,841..813,339		PKS and NRPS
8	NRPS Like	194,426..268,549		
8	T1PKS and NRPS	1,724,943..1,796,808	Phyllostictine A and B	NRPS + PKS

8	Terpene like and Betalactone	2,166,370..2,197,624 3,309,429..3,351,241		Terpene
11	T1PKS	553,564..621,628 1,473,034..1,541,075		PKS
11	NRPS like	2,248,102..2,313,006		NRPS
12	T1PKS	112,527..180,837 197,167..262,437 986,830..1,051,856 1,639,455..1,705,816	Zopfiellin, Phomoidride, YWA1	PKS
12	NRPS like	338,699..401,668		NRPS
12	Terpene like	662,118..693,198		Terpene
14	T1PKS,	1..67,578	Flavoglaucin	PKS
14	NRPS like	2,756,422..2,821,224 3,146,993..3,466,612 4,312,532..4,369,005		
14	Terpene and nitropropanoic acid	1,582,000..1,649,494 3,435,786..3,466,612 3,570,228..3,589,282		
15	NRPS	1,411,614..1,477,993	Metachelin (A, B, C, A-CE), Dimerumic acid 11-mannoside, Dimerumic acid	NRPS
15	T1PKS	4,294,047..4,365,789 5,009,122..5,076,636 5,395,049..5,514,103 6,658,683..6,729,194 8,727,221..8,794,424	Solanapyrone A, YWA1	PKS
15	Terpenes and others	4,961,627..4,993,998 5,733,424..5,764,530 8,230,012..8,319,807	Clavaric acid	Terpene
The secondary metabolite regions were not predicted in the scaffolds, JACGXQ020000006, JACGXQ020000007, JACGXQ020000009, JACGXQ020000010, and JACGXQ020000013.				
*BGC: Biosynthetic Gene Cluster, PKS= Polyketide, NRPS= Non Ribosomal Protein Synthesis				

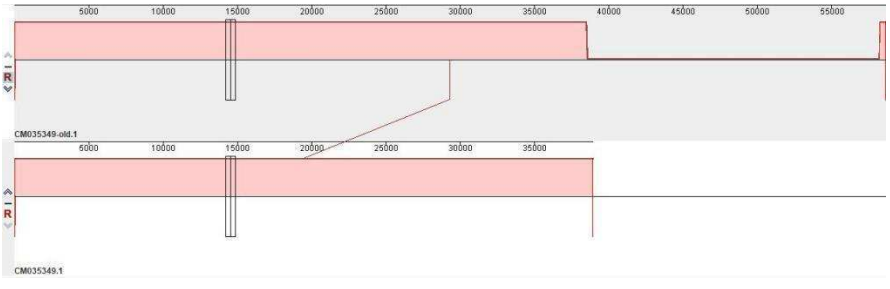
Mitochondrial Genome Assembly Features

The initial assembly of mitochondrial genomic DNA is of length 58664 bp containing duplication of length, 19719 bp genomic region due to the circular nature of the genome. The initial assembly was analyzed using the VISTA, PipMaker [46], and Mauve [47] for duplication of genomic regions **(Fig 1a and Fig 1b)**. The duplicated region was removed at overlapping regions on either end to obtain the final mitochondrial genome of length 38945 bps that is composed of low, 25.31% GC, and high, 74.69% AT content **(Table 5)**. The circular map of the mitochondrial genome of *T. trachyspermus* strain 4014, is drawn by the tool, Organellar Genome DRAW (OGDRAW) [54] (**Fig 2**). The MT genome is comprised of protein coding genes, cox1, cox2, cox3, cob, rrnL, rrnS, atp6, atp8, atp9, nad1, nad2, nad3, nad4, nad5, nad6, giy, lagli, rps3 and rps5 besides tRNA genes and

rRNA genes. The mitochondrial genome annotation and analysis files may be accessed from the Zenodo database [68].



a)



b)

Fig. 1 (a) Dotplot of the preliminary mitochondrial genome assembly (58664 bps) by Pipmaker software, showing the duplicate region indicated by the parallel lines to a diagonal line. (b) Comparison of preliminary assembly and corrected assembly (38945 bps) by Mauve software.

Table 5 Assembly matrix of Mitochondrial Genome

Mitochondrial Genome	Genomic features
Genome size	38,945
GC content (%)	25.31%
Transfer RNA content	42
Ribosomal RNA content	2 (1 rnS and 1 rnL)
Gene content	cox1, cox2, cox3, cob, atp6, atp8, atp9

	nad1, nad2, nad3, nad4, nad5, nad6, giy, lagli, rps3, rps5,

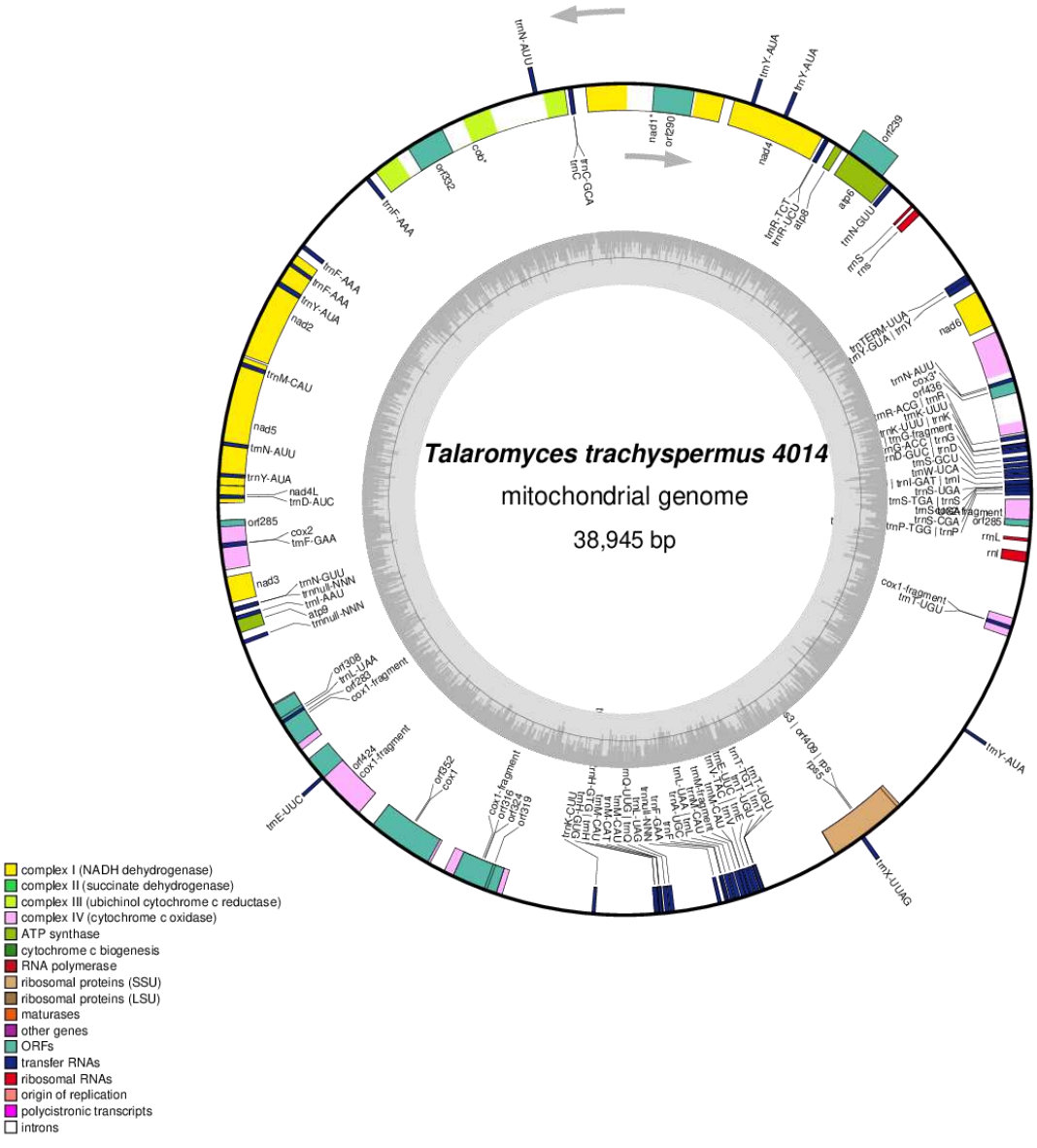


Fig. 2 Presentation of Mitochondrial Circular genome with gene locations

Mitochondrial Genome Homology Search (BLASTN Analysis)

The homology sequence analysis by BLASTN [38] tool of the Mitochondrial genome of *T. trachyseprmus* strain 4014 was performed against the mitochondrial genome dataset of fungi belonging to Eurotiomycetidae (taxid: 451871) keeping expectation value as 0.01, word size as 24 in the first search and 128 in the second search, hit to 10, and other parameters as default. The mitochondrial genome of *T. trachyspermus* 4014 shows high similarity with *T. pinophilus* followed by *T. marneffei* and *T. stipitatus*. The mitochondrial genome of *T. pinophilus* is smaller than *T. trachyspermus* as compared to *T. marneffei* and *T. stipitatus*, therefore

the proportional query coverage is comparatively smaller but with higher similarity. This indicates it is closer to *T. pinophilus* than to *T. marneffei*.

Table 6 Description of mitochondrial genomes of closely related Fungal species to *Talaromyces trachyspermus* 4014

Sr. No.	Name (Accession no)	Length	GC%	tRNA	rRNA
1	*Aspergillus clavatus* NRRL 1 (JQ354999.1)	35056	24.96	26	2
2	*Aspergillus flavus* TX-A-1-1-1 (MZ714581)	29220	26.14	26	2
3	*Aspergillus fumigatus* (MW560465.1)	31751	25.38	26	2
4	*Aspergillus nidulans* FGSC A4 (JQ435097.1)	33227	24.94	26	2
5	*Aspergillus oryzae* ATCC:44194 (KY352472.1)	29202	26.15	26	2
6	*Aspergillus parasiticus* (NC_041445.1)	29141	26.16	26	2
7	*Aspergillus terricola* (NC_058594.1)	28689	26.34	26	2
8	*Aspergillus tubingensis* WU-2223L (LC545447.1)	32393	26.45	26	2
9	*Penicillium canescens* (NC_059911.1)	28301	24.84	25	2
10	*Penicillium chrysogenum* Wisconsin 54-1255 (AM920464.1)	31790	24.94	25	3
11	*Penicillium nordicum* UASWS BFE487 (KR952336.1)	28520	25.43	27	1
12	*Penicillium solitum* 20-01 (JN696111.1)	28601	25.47	27	2
13	*Penicillium* sp. ShG4C (KX931017.1)	26725	25.16	27	2
14	*Talaromyces marneffei* (AY347307.1)	35438	24.63	28	2
15	*Talaromyces marneffei* PM18 (KU761329.1)	35420	24.62	28	2
16	*Talaromyces marneffei* PM26 (KU761332.1)	35434	24.61	28	2
17	*Talaromyces pinophilus* 1-95 (CP017352.1)	31729	24.84	25	2
18	*Talaromyces stipitatus* ATCC 10500 (JQ354994.1)	36351	24.92	27	2
19	*Talaromyces trachyspermus* 4014 (CM035349.2)	38945	25.31	42	2

Phylogenetic Analysis of Mitochondrial Genome

The mitochondrial genomes of 18 species (**Table 6**) comprised of *Aspergillus*, *Penicillium,* and *Talaromyces* species*,* with a similarity in the range 50% - 73% to *T. trachyspermus* 4014 mitochondrial genome in BLASTN analysis, were selected for phylogenetic analysis. The BLASTN distance tree is calculated from the pairwise sequence alignment of query and database sequences by the Fast Minimum Evolution method of the FASTME tool [55] with a maximum sequence difference of 0.75. The tree clustered species in 3 clades, *Aspergillus* with 8 species, *Penicillium* with 5 species, and *Talaromyces* with 6 species. These selected 18 genomes with *T. trachyspermus* were aligned using the ClustalW [56] program and clustered by ML method with Bootstrap method 500 replicates in the MEGA v11.0.10 [57] software. The generated phylogram (**Fig 3**) shows the clustering of *T. trachyspermus* with *T. pinophilus* linked with the cluster of *A. tubingensis* and *P.*

chrysogenum. This clade is separated from the species-specific clusters of the genera, *Aspergillus, Penicillium,* and *Talaromyces*.

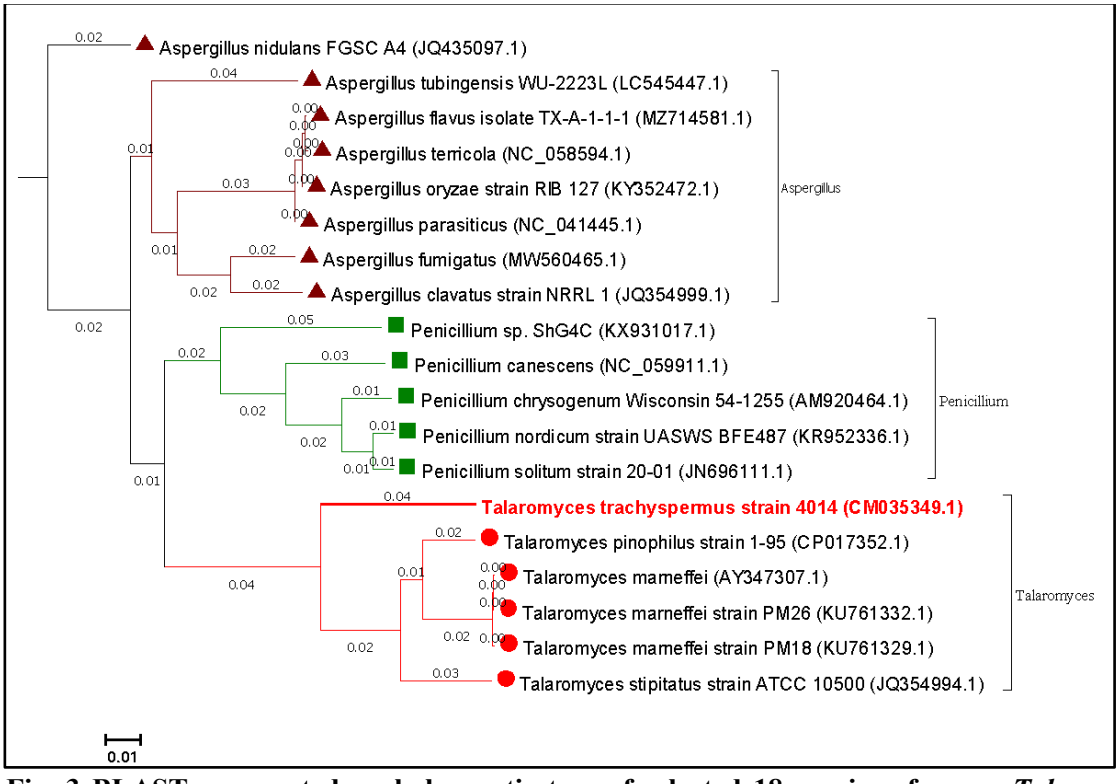


Fig. 3 BLASTn generated a phylogenetic tree of selected 18 species of genus *Talaromyces* (Red circle), *Aspergillus* (Brown triangle), and *Penicillium* (Green square) with *T. trachyspermus* strain 4014 that has query coverage between 50% to 75% and similarity/identity between 82% to 90% in BLASTN analysis. (Note: 500 Bootstrap replicates for the ML method)

Phylogenetic relationship of *Talaromyces trachyspermus* 4014

The phylogenetic relationship was predicted for isolated *T. trachyspermus* strain 4014 based on the whole mitochondrial genome sequence and other similar (identity >80% and query coverage >40%) complete mitochondrial genome sequences that were retrieved by BLAST homology sequence searching. The phylogeny reconstruction based on the mitochondrial genome was calculated using MEGA6 software by UPGMA method and Kimura-2 parameter model. A test of phylogeny was done using the Bootstrap method with 1000 replicates.

DISCUSSION

The isolation and characterization of *Talaromyces trachyspermus* 4014 from the medicinal plant *Withania somnifera* unveils a treasure trove of biotechnological potentials, particularly in agriculture and environmental sustainability. This study presents a comprehensive analysis of the genomic and functional traits of *T. trachyspermus* 4014, revealing its promising role as a plant growth-promoting fungus (PGPF) and a biocontrol agent against phytopathogens.

The genome of *T. trachyspermus* 4014, assembled and annotated with high confidence, uncovers a multitude of genes associated with plant growth promotion and biocontrol. The presence of genes coding for indole-3-acetic acid (IAA) synthesis, siderophore production, and various hydrolytic enzymes underscores the multifaceted mechanisms by which this fungus can enhance plant health. The ability to produce IAA, a key phytohormone,

suggests that *T. trachyspermus* 4014 can stimulate root development, improving nutrient uptake and plant growth [58].

Moreover, the production of siderophores indicates that *T. trachyspermus* 4014 can effectively sequester iron from the environment, a critical factor in outcompeting pathogenic microbes in the rhizosphere [59]. The array of hydrolytic enzymes, including protease, cellulase, chitinase, amylase, lipase, and phosphatase, highlights the potential of *T. trachyspermus* 4014 to degrade various substrates, thereby improving soil health and fertility [60].

The antagonistic activity of *T. trachyspermus* 4014 against several plant pathogenic fungi, demonstrated through dual culture assays, further solidifies its role as a biocontrol agent. The inhibition zones observed in these assays suggest that *T. trachyspermus* 4014 produces antifungal compounds that can suppress the growth of pathogens [61]. This finding is particularly significant in the context of sustainable agriculture, where the reliance on chemical pesticides is being reduced in favor of biological control methods [62].

In addition to its agricultural applications, the genomic insights into *T. trachyspermus* 4014 open avenues for exploring its potential in bioremediation. The enzymes produced by this fungus could be harnessed to degrade environmental pollutants, contributing to the restoration of contaminated ecosystems [63]. The submission of the whole genome sequence to the NCBI genome database ensures that this valuable genetic information is accessible for further research and biotechnological exploitation.

The isolation of *T. trachyspermus* 4014 from *Withania somnifera* also raises intriguing questions about the ecological interactions between endophytic fungi and their host plants. The symbiotic relationship between *T. trachyspermus* 4014 and *W. somnifera* may confer mutual benefits, with the plant providing a niche for the fungus, while the fungus enhances the plant's growth and defense mechanisms [64]. This study sets the stage for future investigations into the molecular dialogues that underpin such symbiotic associations. With limited studies on fungal-supplementation on various plant leaves/extracts, our efforts on long-read sequencing may provide plausible insights into understanding undetermined avenues such as horizontal gene transfer (HGT) events. adverse effects of the hosts besides providing information on cellular machinery for biomass-degrading enzymes. We aim to provide a pan genome-sequence basis for identifying strategies employing this fungus as a microbial cell factory in discerning candidate secondary metabolites [65].

The mitochondrial genome-based phylogenetic analysis clustered our strain with *Talaromcyes* species. The complete mitochondrial genome sequence of *T. Trachyspermus* 4014 may be the input data source for the studies related to fungal evolution [66] and species identification [67] based on the mitochondrial genome.

In conclusion, *Talaromyces trachyspermus* 4014 emerges as a versatile fungus with significant implications for agriculture, environmental sustainability, and beyond. Its ability to promote plant growth, combat phytopathogens, and potentially degrade pollutants positions it as a valuable asset in the quest for sustainable solutions to global challenges. Further research into the molecular mechanisms driving these beneficial traits will undoubtedly unlock new dimensions of its biotechnological potential

Genomics Data Access

The whole genome sequence data comprising 15 scaffolds are submitted to the NCBI genome assembly database comprised of with central genomic accession number: JACGXQ010000001..JACGXQ010000004 and JACGXQ010000006..JACGXQ010000015 (https://www.ncbi.nlm.nih.gov/nuccore/JACGXQ000000000) and mitochondrial genome accession number: CM035349 (JACGXQ010000005) (https://www.ncbi.nlm.nih.gov/nuccore/CM035349). The work is registered under BioProject with ID: PRJNA650128 (https://www.ncbi.nlm.nih.gov/bioproject/650128) and fungal strain under BioSample database

with ID: SAMN15692527 (https://www.ncbi.nlm.nih.gov/biosample/SAMN15692527). The nanopore read sequence data can be accessed from the Sequence Read Archive (SRA) database with the accession number SRR25836654 (https://trace.ncbi.nlm.nih.gov/Traces/?run=SRR25836654). The annotation summary, annotation files in GenBank and tabular formats, protein sequence file, nucleotide sequence file, metabolic pathway analysis files and mitochondrial genome annotation files are uploaded to Zenodo (https://doi.org/10.5281/zenodo.18703994) [68].

CONCLUSIONS

We present the sequencing and assembly of the genome of the fungus, *Talaromyces trachyspermus* which inhibits diverse environments, including soil, marine sponges, food materials, and plant parts. The species was isolated from the leaves of the medicinal plant, *Withania somnifera,* and observed to exhibit plant growth-promoting and pathogen-antagonistic properties. The mitochondrial genome showed a close relation to *T. marneffei* and *T. pinuphilus* concerning size and gene content. While there is room for understanding HGTed events, more uniformity in hybrid assembly could reveal genomic insights associated with habitat diversity, plant-microbe interaction, secondary metabolites synthesis, and siderophore pathways too.

REFERENCES

1. Yilmaz N, López-Quintero CA, Vasco-Palacios AM, Frisvad JC, Theelen B, Boekhout T, Samson RA, Houbraken J. Four novel *Talaromyces* species isolated from leaf litter from Colombian Amazon rain forests. Mycol Prog. 2016; https://doi.org/10.1007/s11557-016-1227-3
2. Fang D and Shi C. Characterization and flocculability of a novel proteoglycan produced by *Talaromyces trachyspermus* OU5. J Biosci Bioeng. 2016; https://doi.org/10.1016/j.jbiosc.2015.05.001
3. Dethoup T, Kumla D, Kijjoa A. Mycocidal activity of crude extracts of marine-derived beneficial fungi against plant pathogenic fungi. J Biopest. 2015; https://doi.org/10.57182/jbiopestic.8.2.107-115
4. Tranquillini R, Scaramuzza N, Berni E. Occurrence and ecological distribution of heat resistant moulds spores (HRMS) in raw materials used by food industry and thermal characterization of two *Talaromyces* isolates. Int J Food Microbiol. 2017; https://doi.org/10.1016/j.ijfoodmicro.2016.11.023
5. Visagie CM, Hirooka Y, Tanney JB, Whitfield E, Mwange K, Meijer M, Amend AS, Seifert KA, Samson RA. Aspergillus, Penicillium and Talaromyces isolated from house dust samples collected around the world. Stud Mycol. 2014; https://doi.org/10.1016/j.simyco.2014.07.002
6. Qin Y, Huang X, Chen H, Liu X, Li Y, Hou J, Li A, Yan X, Chen Y. Burden of *Talaromyces marneffei* infection in people living with HIV/AIDS in Asia during ART era: a systematic review and meta-analysis. BMC Infect Dis. 2020; https://doi.org/10.1186/s12879-020-05260-8
7. Hernández VA, Galleguillos FA, Sagredo N, Machuca Á. A Note on the Dyeing of Wool Fabrics Using Natural Dyes Extracted from Rotten Wood-Inhabiting. Fungi Coatings. 2018; https://doi.org/10.3390/coatings8020077
8. Kim H, You YH, Yoon H, Seo Y, Kim YE, Choo YS, Lee IJ, Shin JH, Kim JG. Culturable fungal endophytes isolated from the roots of coastal plants inhabiting Korean east coast. Mycobiology. 2014; https://doi.org/10.5941/MYCO.2014.42.2.100
9. Sahu S, Prakash A, Shende K. *Talakimromyces trachyspermus*, an endophyte from *Withania somnifera* with plant growth-promoting attributes. J Environ Sustain. 2019; https://doi.org/10.1007/s42398-019-00045-5

10. Barral B, Chillet M, Doizy A, Grassi M, Ragot L, Léchaudel M, Durand N, Rose LJ, Viljoen A, Schorr-Galindo S. Diversity and Toxigenicity of Fungi that Cause Pineapple Fruitlet Core Rot. Toxins. 2020; https://doi.org/10.3390/toxins12050339
11. Benjamin CR. Ascocarps of *Aspergillus* and *Penicillium*. Mycologia. 1955; https://doi.org/10.1080/00275514.1955.12024485
12. Samson RA, Yilmaz N, Houbraken J, Spierenburg H, Seifert KA, Peterson SW, Varga J, Frisvad JC. Phylogeny and nomenclature of the genus *Talaromyces* and taxa accommodated in *Penicillium* subgenus Biverticillium. Stud Mycol. 2011; https://doi.org/10.3114/sim.2011.70.04
13. McNeill J, Barrie FR, Buck WR, Demoulin V, Greuter W, Hawksworth DL, Herendeen PS, Knapp S, Marhold K, Prado J, Van Reine WFP, Smith GF, Wiersema JH, Turland NJ. International Code of Nomenclature for algae, fungi and plants (Melbourne Code). ARG Gantner Verlag KG; 2012. pp. 154-240.
14. Berbee ML, Yoshimura A, Sugiyama J, Taylor JW. Is *Penicillium* monophyletic? An evaluation of phylogeny in the family Trichocomaceae from 18S, 5.8S and ITS ribosomal DNA sequence data. Stud Mycol. 1995; https://doi.org/10.1080/00275514.1995.12026523
15. Houbraken J, Kocsubé S, Visagie CM, Yilmaz N, Wang XC, Meijer M, Kraak B, Hubka V, Bensch K, Samson RA, Frisvad JC. Classification of *Aspergillus*, *Penicillium*, *Talaromyces* and related genera (Eurotiales): An overview of families, genera, subgenera, sections, series and species. Stud Mycol. 2020; https://doi.org/10.1016/j.simyco.2020.05.002.
16. Frisvad JC, Yilmaz N, Thrane U, Rasmussen KB, Houbraken J, Samson RA. *Talaromyces atroroseus*, a new species efficiently producing industrially relevant red pigments. PLoS One. 2013; https://doi.org/10.1371/journal.pone.0084102
17. Méndez-Líter JA, Nieto-Domínguez M, de Toro BF, Santana AG, Prieto A, Asensio JL, Cañada FJ, de Eugenio LI, Martínez MJ. A glucotolerant β-glucosidase from the fungus *Talaromyces amestolkiae* and its conversion into a glycosynthase for glycosylation of phenolic compounds. Microb Cell Fact. 2020; https://doi.org/10.1186/s12934-020-01386-1
18. Inoue H, Decker SR, Taylor LE, Yano S, Sawayama S. Identification and characterization of core cellulolytic enzymes from *Talaromyces cellulolyticus* (formerly *Acremonium cellulolyticus*) critical for hydrolysis of lignocellulosic biomass. Biotechnol Biofuels. 2014; https://doi.org/10.1186/s13068-014-0151-5
19. Li L, Chen K, Dhungana N, Jang Y, Chaturvedi V, Desmond E. Characterization of Clinical Isolates of *Talaromyces marneffei* and Related Species, California, USA. Emerg Infect Dis. 2019; https://doi.org/10.3201/eid2509.190380
20. Zhai MM, Li J, Jiang CX, Shi YP, Di DL, Crews P, Wu QX. The Bioactive Secondary Metabolites from *Talaromyces* species. Nat Prod Bioprospect. 2016; https://doi.org/10.1007/s13659-015-0081-3
21. Singh LP, Gill SS, Tuteja N. Unraveling the role of fungal symbionts in plant abiotic stress tolerance. Plant Signal Behav. 2011; https://doi.org/10.4161/psb.6.2.14146
22. Mane RS, Vedamurthy AB. The fungal endophytes: source and future prospectus. J Med Plant Stud. 2018; https://www.plantsjournal.com/archives/2018/vol6issue2/PartB/6-2-31-305.pdf
23. Dethoup T, Kumla D, Kijjoa A. Mycocidal activity of crude extracts of marine-derived beneficial fungi against plant pathogenic fungi. J Biopest. 2015; https://doi.org/10.57182/jbiopestic.8.2.107-115
24. Wingett SW and Andrews S. FastQ Screen: A tool for multi-genome mapping and quality control. F1000Research. 2018; https://doi.org/10.12688/f1000research.15931.2
25. Koren S, Walenz BP, Berlin K, Miller JR, Phillippy AM. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res. 2017; https://doi.org/10.1101/gr.215087.116

26. Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EM. 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; https://doi.org/10.1093/molbev/msab199
27. Parra G, Bradnam K, Korf I. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. Bioinformatics. 2007; https://doi.org/10.1093/bioinformatics/btm071
28. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. Bioinformatics. 2013; https://doi.org/10.1093/bioinformatics/btt086
29. Humann JL, Lee T, Ficklin S, Main D. Structural and Functional Annotation of Eukaryotic Genomes with GenSAS. In: Kollmar M. (eds) Gene Prediction. Methods in Molecular Biology 1962:29-51. Humana, New York, NY. 2019; https://doi.org/10.1007/978-1-4939-9173-0_3
30. Simao FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 2015; https://doi.org/10.1093/bioinformatics/btv351
31. Chen N. Using RepeatMasker to identify repetitive elements in genomic sequences. Curr Protoc Bioinformatics. 2004; https://doi.org/10.1002/0471250953.bi0410s05
32. Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, Smit AF. RepeatModeler2 for automated genomic discovery of transposable element families. Proc Natl Acad Sci. 2020; https://doi.org/10.1073/pnas.1921046117
33. Stanke M, Diekhans M, Baertsch R, Haussler D. Using native and syntenically mapped cDNA alignments to improve de novo gene finding. Bioinformatics. 2008; https://doi.org/10.1093/bioinformatics/btn013
34. Lomsadze A, Ter-Hovhannisyan V, Chernoff YO, Borodovsky M. Gene identification in novel eukaryotic genomes by self-training algorithm. Nucleic Acids Res. 2005; https://doi.org/10.1093/nar/gki937
35. Salzberg S, Delcher A, Kasif S, White O. Microbial gene identification using interpolated Markov models. Nucleic Acids Res. 1998; https://doi.org/10.1093/nar/26.2.544
36. Haas BJ, Salzberg SL, Zhu W, Pertea M, Allen JE, Orvis J, White O, Buell CR, Wortman JR. Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. Genome Biol. 2008; https://doi.org/10.1186/gb-2008-9-1-r7
37. Haas BJ, Delcher AL, Mount SM, Wortman JR, Smith RK Jr, Hannick LI, Maiti R, Ronning CM, Rusch DB, Town CD, Salzberg SL, White O. Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies. Nucleic Acids Res. 2003; https://doi.org/10.1093/nar/gkg770
38. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990; https://doi.org/10.1016/S0022-2836(05)80360-2
39. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. Nat Methods. 2015; https://doi.org/10.1038/nmeth.3176
40. Wang M and Kong L. pblat: a multithread blat algorithm speeding up aligning sequences to genomes. BMC Bioinformatics. 2019; https://doi.org/10.1186/s12859-019-2597-8.
41. Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, Pesseat S, Quinn AF, Sangrador-Vegas A, Scheremetjew M, Yong SY, Lopez R, Hunter S. InterProScan 5: genome-scale protein function classification. Bioinformatics. 2014; https://doi.org/10.1093/bioinformatics/btu031
42. El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, Qureshi M, Richardson LJ, Salazar GA, Smart A, Sonnhammer ELL, Hirsh L, Paladin L, Piovesan D, Tosatto SCE, Finn RD (2019) The Pfam protein families database in 2019. Nucleic Acids Res. 2019; https://doi.org/10.1093/nar/gky995

43. Armenteros JJA, Tsirigos KD, Sønderby CK, Nordahl T, Ole Winther P, Brunak S, von Heijne G, Nielsen H (2019) SignalP 5.0 improves signal peptide predictions using deep neural networks. Nat. Biotechnol. 2019; https://doi.org/10.1038/s41587-019-0036-z
44. Lagesen K, Hallin P, Rødland EA, Staerfeldt HH, Rognes T, Ussery DW. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res. 2007; https://doi.org/10.1093/nar/gkm160
45. Chan PP and Lowe TM. tRNAscan-SE: Searching for tRNA Genes in Genomic Sequences. In: Kollmar, M. (eds) Gene Prediction. Methods in Molecular Biology. 20149; https://doi.org/10.1007/978-1-4939-9173-0_1
46. Schwartz S, Zhang Z, Frazer KA, Smit A, Riemer C, Bouck J, Gibbs R, Hardison R, Miller W. PipMaker-A Web Server for Aligning Two Genomic DNA Sequences. Genome Res. 2000; https://doi.org/10.1101/gr.10.4.577
47. Darling ACE, Mau B, Blattner FR, Perna NT. Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res. 2004; https://doi.org/10.1101/gr.2289704
48. Kanehisa M, Sato Y, Morishima K. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. J Mol Biol. 2016; https://doi.org/10.1016/j.jmb.2015.11.006
49. Blin K, Shaw S, Vader L, Szenei J, Reitz ZL, Augustijn HE, Cediel-Becerra JDD, Crécy-Lagard V, Koetsier RA, Williams SE, Cruz-Morales P, Wongwas S, Luchsinger AES, Biermann F, Korenskaia A, Zdouc MM, Meijer D, Terlouw BR, van der Hooft JJ, Ziemert N, Helfrich EJN, Masschelein J, Corre C, Chevrette MG, van Wezel GP, Medema MH, Weber T. antiSMASH 8.0: extended gene cluster detection capabilities and analyses of chemistry, enzymology, and regulation. Nucleic Acids Res. 2025; https://doi.org/10.1093/nar/gkaf334
50. Geib SM, Hall B, DeRego T, Bremer FT, Cannoles K, Sim SB. Genome Annotation Generator: a simple tool for generating and correcting WGS annotation tables for NCBI submission. GigaScience. 2014a; https://doi.org/10.1093/gigascience/giy018
51. Geib SM, Tate R, Hall B, DeRego T. Annie: the ANNotation Information Extractor (Version 1.0) Software. 2014B; http://genomeannotation.github.io/annie Accessed on date: 30-08-2024.
52. Sayers EW, Cavanaugh M, Clark K, Pruitt KD, Sherry ST, Yankie L, Karsch-Mizrachi I. GenBank 2024 Update. Nucleic Acids Res. 2024; https://doi.org/10.1093/nar/gkad903
53. Manni M, Berkeley MR, Seppey M, Zdobnov EM. BUSCO: Assessing genomic data quality and beyond. Curr Protoc Bioinform. 2021; https://doi.org/10.1002/cpz1.323
54. Greiner S, Lehwark P, Bock R. OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. Nucleic Acids Res. 2019; https://doi.org/10.1093/nar/gkz238
55. Desper R and Gascuel O. Theoretical foundation of the balanced minimum evolution method of phylogenetic inference and its relationship to weighted least-squares tree fitting. Mol Biol Evol. 2004; https://doi.org/10.1093/molbev/msh049.
56. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 1994; https://doi.org/10.1093/nar/22.22.4673
57. Tamura K, Stecher G, Kumar S. MEGA11: Molecular Evolutionary Genetics Analysis version 11. Mol Bio Evol. 2021; https://doi.org/10.1093/molbev/msab120
58. Patten CL and Glick BR. Role of *Pseudomonas putida* indoleacetic acid in development of the host plant root system. Appl Environ Microbiol. 2002; https://doi.org/10.1128/AEM.68.8.3795-3801.2002
59. Loper JE and Henkels MD. Utilization of heterologous siderophores enhances levels of iron available to *Pseudomonas putida* in the rhizosphere. Appl Environ Microbiol. 1999; https://doi.org/10.1128/AEM.65.12.5357-5363.1999

60. Harman GE, Howell CR, Viterbo A, Chet I, Lorito M. Trichoderma species - opportunistic, avirulent plant symbionts. Nat Rev Microbiol. 2004; https://doi.org/10.1038/nrmicro797
61. Whipps JM. Microbial interactions and biocontrol in the rhizosphere. J Exp Bot. 2001; https://doi.org/10.1093/jexbot/52.suppl_1.487.
62. Pal KK and Gardener BM. Biological control of plant pathogens. Plant Health Instr. 2006; https://doi.org/10.1094/PHI-A-2006-1117-02
63. Pointing SB. Feasibility of bioremediation by white-rot fungi. Appl. Microbiol. Biotechnol. 2001; https://doi.org/10.1007/s002530100745
64. Stone JK, Bacon CW, White JF. An Overview of Endophytic Microbes: Endophytism Defined. In: Bacon, C.W. and White, J.F., Eds., Microbial Endophytes, Marcel Dekker, New York; 2000; PP. 3-29
65. Li CX, Zhao S, Zhang T, Xian L, Liao L, Liu J, Feng J. Genome sequencing and analysis of Talaromyces pinophilus provide insights into biotechnological applications. Sci Rep. 2017; https://doi.org/10.1038/s41598-017-00567-0
66. Fonseca PLC, De-Paula RB, Araújo DS, Tomé LMR, Mendes-Pereira T, Rodrigues WFC, Del-Bem LE, Aguiar ERGR, Góes-Neto A. Global Characterization of Fungal Mitogenomes: New Insights on Genomic Diversity and Dynamism of Coding Genes and Accessory Elements. Front Microbiol. 2021; https://doi.org/10.3389/fmicb.2021.787283
67. Kulik T, Ostrowska A, Buśko M, Pasquali M, Beyer M, Stenglein S, Załuski D, Sawicki J, Treder K, Perkowski J. Development of an FgMito assay: a highly sensitive mitochondrial based qPCR assay for quantification of *Fusarium graminearum* sensu stricto. Int J Food Microbiol. 2015; https://doi.org/10.1016/j.ijfoodmicro.2015.06.012.
68. Prakash, A., Sahu, S., Dharne, M., Nair, K., Rajput, V., Samson, R., Medicherla, K. M., Suravajhala, P., & Shende, K. K. (2026). Genome annotations, mRNA sequences, Protein sequences and mitochondrial genome annotation files Talaromyces_trachyspermus 4014 (Version 1) [Data set]. Zenodo. https://doi.org/10.5281/zenodo.18703994

Declarations

Ethics approval and consent to participate: Not applicable

Consent for publication: Not applicable

Availability of data and materials

The wild plant of *Whitania somnifera* (L) was obtained from the nursery of Vindhay Herbal, a Madhya Pradesh State Forest Produce (T & D) Co. Fed., Bhopal in the year 2015 and planted at Department of Microbiology, Barkatullah University, Bhopal. The Plant was identified by Vindhya Herbal and by Prof. Anil Prakash at Department of Microbiology, Barkatullah University. The endophytic fungal isolates were obtained from the foliage tissues. The potential endophytic fungal strain under study, *Talaromyces trachyspermus* was identified at National Fungal Culture Collection of India (NFCCI), Agarkar Research Institute, Pune, India and provided an accession number: 4014. The Biosample was registered at NCBI BioSample database with accession number: SAMN15692527 under the BioProject: PRJNA650128, for submitting the NCBI assembly database (accession number JACGXQ010000001 .. JACGXQ010000015) and read sequences submitted to the SRA database

(accession number SRR25836654). The annotations and analysis files are uploaded to Zenodo (https://doi.org/10.5281/zenodo.18703994) database.

Competing interests: Authors declare that they have no competing interest.

Funding: The present research work did not receive any funding from any funding agencies. The required funding was managed through the internal resources, which had no role in study design, execution, analysis, data interpretation and submission decision.

Authors' contributions: AP is the PI of the project and mentored by KS, SS carried out the culturing of fungal strain. MD, KN, VR and RS, carried out DNA extraction, library preprrations, genome sequencing and basecalling. KKS performed the genome assembly, annotations and analysis. The genome analyses done by PS. All authors contributed to the drafting of manuscript. All authors agreed to the publication of the MS before communication.

Acknowledgements

The authors acknowledge here, the Department of Microbiology, Barkatullah University, Bhopal-462026, M.P., India for providing the laboratory facility for maintaining the fungal culture and preliminary investigations. We also acknowledge The Director, National Chemical Laboratory, Pune, for infrastructure and support. This work was carried out without the aid of funding from the funding agency.

Authors' information

[a]Department of Biotechnology and Bioinformatics Center (SubDIC), Barkatullah University, Bhopal, M.P., India.
[b]NCIM Resource Center, NCL, Pune, M.S., India.
[c]Swami Atmanand Govt. English Medium School, Bilaspur, CG, India.
[d]Birla Institute of Scientific Research, Jaipur, Rajasthan, India.
[e]Amrita School of Biotechnology, Amrita Vishwa Vidyapeetham, Clappana 690525, Kerala, India.
[f]Department of Microbiology, Barkatullah University, Bhopal, M.P., India
[g]Centre for Sustainable Agricultural and Environment, Prof. H.S. Srivastava Foundation for Science and Society, Uttarathia, Lucknow, U.P., India.