Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

A Telomere-to-Telomere Genome of Talaromyces marneffei B33c Reveals Chromosomal Architecture and Cross-Border Connectivity

Xinyu Zhou
Department of Dermatology and Venereology, The First Affiliated Hospital of Guangxi Medical University
Yanqing Zheng
Fourth People's Hospital of Nanning
Zhiwen Jiang
Department of Dermatology and Venereology, The First Affiliated Hospital of Guangxi Medical University
Saroj Karki
Department of Dermatology and Venereology, The First Affiliated Hospital of Guangxi Medical University
Kaisu Pan
Department of Dermatology and Venereology, The First Affiliated Hospital of Guangxi Medical University
Dongyan Zheng
Department of Dermatology and Venereology, The First Affiliated Hospital of Guangxi Medical University
Xiuying Li
Department of Dermatology and Venereology, The First Affiliated Hospital of Guangxi Medical University
Lan Huang
Department of Dermatology and Venereology, The First Affiliated Hospital of Guangxi Medical University
Guoqun Liu
Department of Dermatology and Venereology, The First Affiliated Hospital of Guangxi Medical University
JinJin Ling
Department of Dermatology and Venereology, The First Affiliated Hospital of Guangxi Medical University
Clement K M Tsui
Infectious Diseases Research Laboratory, National Centre for Infectious Diseases
Cunwei Cao
caocunwei@yeah.net

Department of Dermatology and Venereology, The First Affiliated Hospital of Guangxi Medical University

Research Article

Keywords: Talaromyces marneffei, talaromycosis, telomere-to-telomere genome, phylogenomics, comparative genomics

Posted Date: May 23rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9282981/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at Mycopathologia on October 22nd, 2026. See the published version at https://doi.org/10.1007/s11046-026-01105-1.

Abstract

Background

Talaromyces marneffei is a thermally dimorphic fungal pathogen endemic to Southeast Asia and southern China. However, fragmented reference genomes have limited detailed investigations of its chromosomal architecture, structural variation, and evolutionary history.

Methods

We integrated PacBio HiFi, Oxford Nanopore, and Hi-C sequencing data to generate a telomere-to-telomere (T2T), gap-free genome assembly of strain B33c, a landmark clinical isolate collected in 1984. Comparative genomic, phylogenomic, and orthogroup analyses were then performed to characterize chromosomal features, evolutionary relationships, and gene-content variation.

Results

The 28.1-Mb assembly resolved all eight chromosomes and enabled comprehensive delineation of inferred centromeric domains and telomere-proximal regions. Complete telomeric repeats were identified at both ends of every chromosome, and chromosome-scale Hi-C contact maps supported a gap-free, chromosome-complete reference. Using 3,314 single-copy orthologs, we reconstructed a time-calibrated phylogeny that revealed a deep intraspecific split at approximately 6.8 Mya, separating *T. marneffei* into two major lineages. B33c clustered closely with Vietnamese isolates and showed shallow divergence (~ 0.3–0.4 Mya), suggesting cross-border genetic connectivity between northern Vietnam and southern China. Comparative genomics demonstrated broadly conserved macrosynteny, with localized repeat-associated structural variation showing a recurrent chromosome-biased distribution. Orthogroup profiling further supported a “conserved core–flexible periphery” genomic strategy, with marked accessory-genome differences between the environmental isolate ATCC 18224 and the clinical isolate B33c.

Conclusion

This T2T, gap-free genome of *T. marneffei* B33c provides a high-quality chromosomal reference for investigating structural plasticity, phylogeography, and pathogenic evolution in this neglected tropical pathogen.

Introduction

Talaromyces marneffei (formerly *Penicillium marneffei*) is a thermally dimorphic fungal pathogen that grows as a filamentous mold in the environment and switches to a yeast form at 37°C in the host, causing talaromycosis,a life–threatening systemic infection endemic to tropical Asia. Since the first human case was documented in 1973[1], the HIV/AIDS pandemic has significantly increased the incidence of *T. marneffei* infections[2], transforming talaromycosis from sporadic cases into one of the major opportunistic mycoses among HIV–positive individuals in Southeast Asia and southern China[3, 4]. By 2022, over 280,000 cases had been reported across 30 countries[5], and ecological niche modeling suggests that the geographic range may further expand under factors such as climate change and human migration[6]. In China, the first confirmed case was reported in Guangxi in 1984[7, 8], followed by spread from southern endemic foci to multiple provinces nationwide[3, 9], contributing to a substantial disease burden[10].

Talaromycosis is also increasingly reported in HIV–negative patients with acquired immunosuppression (e.g., organ transplantation, autoimmune disorders, or malignancies[11]). Despite standard therapies using amphotericin B and triazoles, mortality remains high, often exceeding 30%[5], while emerging antifungal resistance further complicates clinical management[12, 13]. Recognizing its clinical and public–health importance, the World Health Organization (WHO) listed *T. marneffei* as a Medium Priority Group pathogen in its first-ever fungal priority pathogens list in 2022, emphasizing the urgent need for research into its epidemiology, evolution, and pathogenesis.

High–quality, contiguous genomic resources are indispensable for elucidating the evolution and pathogenicity of fungal pathogens. To date, several *T. marneffei* genomes and genome-enabled studies—including PM1, ATCC 18224, GZ8H79, and others—have provided foundational insights into the species' biology[14–19]. However, existing assemblies remain fragmented at the contig or scaffold level, leaving repetitive, structurally complex regions—such as centromeres and telomere-proximal regions—largely unresolved. These regions are often associated with structural variation and rapid adaptation but remain difficult to resolve without telomere–to–telomere (T2T) continuity. Furthermore, published genomes mostly represent isolates from the past decade, creating a critical gap in our understanding of the genomic landscape of early clinical isolates and hindering the reconstruction of long–term evolutionary and epidemiological trajectories.

To address these gaps, we generated the first telomere-to-telomere (T2T), gap–free reference genome of *T. marneffei* strain B33c, an early clinical isolate obtained in China in 1984. This eight-chromosome assembly provides a high-resolution foundation for comprehensive structural and functional annotation. Using this complete reference, we performed systematic analyses of chromosomal organization, synteny, and gene-family evolution across representative species of the genus *Talaromyces*. Our findings provide new insights into cross-border genetic connectivity and isolate-associated genomic variation, and establish a foundational reference for future investigations into the population history and pathogenicity of this priority fungal pathogen.

Materials and Methods

Strain Collection and DNA Extraction

The *T. marneffei* strain B33c was originally isolated in 1984 from a patient with a clinically confirmed case of talaromycosis at the First Affiliated Hospital of Guangxi Medical University. The archived isolate was kindly provided to us by Professor Weida Liu and is currently maintained at the Guangxi Key Laboratory of Mycosis Prevention and Treatment, Guangxi Medical University, Nanning, Guangxi, China, for this study. This study was approved by the Ethics Committee of the First Affiliated Hospital of Guangxi Medical University (Approval No. 2024-E507-01); the isolate was de-identified, informed consent was waived, and no identifiable patient information was used.

The strain was revived on potato dextrose agar (PDA; Thermo Fisher Scientific, Waltham, MA, USA) at 25°C for 5–7 days. Following single-colony purification and verification of typical morphology and absence of contamination, mycelia were subcultured in potato dextrose broth (PDB) at 25°C with shaking until sufficient biomass was obtained. Mycelia were harvested by filtration, flash-frozen in liquid nitrogen, and ground to a fine powder using a pre-chilled mortar and pestle. Total genomic DNA (gDNA) was extracted using the E.Z.N.A.™ Fungal DNA Mini Kit (Omega Bio–Tek, Norcross, GA, USA) according to the manufacturer's protocol. DNA integrity was assessed by 1% (w/v) agarose gel electrophoresis. DNA purity was evaluated using a NanoDrop 2000 spectrophotometer, and DNA concentration was determined using a Qubit 4 fluorometer with the dsDNA HS Assay Kit (Thermo Fisher Scientific). Only DNA samples meeting purity criteria (A260/A280 = 1.8–2.0) and showing minimal degradation on agarose gels were used for subsequent library construction and sequencing.

Genome Sequencing

A multi-platform sequencing strategy integrating PacBio HiFi, Oxford Nanopore long reads, and Hi-C sequencing was employed to generate a gap-free, telomere-to-telomere (T2T) genome assembly. Library construction and sequencing were performed by Shanghai Lingen Biotechnology Co., Ltd. (Shanghai, China).

For PacBio sequencing, SMRTbell libraries were prepared using the SMRTbell Express Template Prep Kit v2.0 (Pacific Biosciences, Menlo Park, CA, USA) and sequenced on a Sequel IIe system. Circular consensus sequencing (CCS) reads were generated using SMRT Link v12.0 (Pacific Biosciences) with a minimum read quality (RQ) of 0.99. For Oxford Nanopore sequencing, high-molecular-weight DNA was size-selected to enrich long fragments, and libraries were prepared using the SQK–LSK109 ligation kit (Oxford Nanopore Technologies, Oxford, UK). Sequencing was performed on the PromethION platform. Basecalling was conducted with Guppy v5.0.11, and only reads with a mean quality score (Q) ≥ 7 were retained for downstream analysis. For Hi-C sequencing, chromatin was cross-linked using 1% formaldehyde, and libraries were constructed using the Arima–HiC + kit (Arima Genomics, San Diego, CA, USA) with MboI as the restriction enzyme. Hi-C libraries were sequenced on a DNBSEQ–T7 platform (MGI Tech, Shenzhen, China) in 150-bp paired-end mode. Detailed sequencing statistics, including total yield and estimated coverage for each platform, are provided in Table S1.

Genome Assembly and Quality Assessment

De novo genome assembly was performed using Hifiasm v0.16.1[20] with PacBio HiFi reads as the primary input. Oxford Nanopore long reads were used to support assembly validation and to assist manual resolution of repetitive and structurally complex regions. Hi–C reads were aligned to the draft B33c assembly using BWA–MEM (MEM algorithm) [21] with default parameters unless otherwise stated, and only read pairs with a mapping quality score (MAPQ) ≥ 30 were retained for scaffolding. Hi–C contact maps were generated with Juicer v1.6[22], and chromosome–scale scaffolding was conducted with ALLHiC v0.9[23] to anchor and orient contigs into eight chromosomes. Manual inspection and correction of scaffold ordering and orientation were performed using Juicebox v1.11[24]. During manual curation, unanchored and/or redundant sequences were removed based on low Hi–C support and redundancy against anchored chromosomal sequences.

Telomeric repeats were identified using quarTeT v1.2.2[25] by searching for the canonical (TTAGGG)_n motif at chromosome termini. Putative centromeric regions were inferred by scanning GC content in 500 bp windows using BEDTools v2.30.0[26]; extended AT–rich, GC–depleted segments were identified as candidate centromeres. TE enrichment in candidate regions was assessed based on RepeatMasker annotations (RepeatModeler-derived library; see Section 2.4)[27, 29], and the positions were further validated by cross–referencing with interaction domains in Hi–C contact maps. Assembly completeness was assessed using BUSCO v5.3[28] with the fungi_odb10 dataset. Hi–C contact maps were further examined to confirm chromosome anchoring and orientation, characterized by continuous intrachromosomal interaction diagonals and distinct inter–chromosomal boundaries.

Repeat Elements and Non–Coding RNA Annotation

Repetitive elements were annotated using a combination of de novo prediction and homology–based searches. A custom repeat library was generated using RepeatModeler v2.0[27] and merged with curated repeat libraries and RepeatMasker ProteinMask entries to include both novel and known transposable elements. The resulting library was applied using RepeatMasker v4.1.0[29] for genome–wide annotation and masking of interspersed repeats. Tandem repeats were detected with Tandem Repeats Finder (TRF) v4.09.1[30] using default parameters. Redundant or overlapping repeat annotations were removed based on genomic coordinates to obtain a non–redundant repeat set.

Non–coding RNAs (ncRNAs) were predicted from the repeat–masked genome using multiple complementary tools. Transfer RNAs (tRNAs) were identified with tRNAscan–SE v2.0.7[31], ribosomal RNAs (rRNAs) with RNAmmer v1.2[32], and other ncRNA families (including sRNAs, snRNAs, and miRNA–like RNAs) were annotated using Infernal v1.1.4[33] against the Rfam v15.0 database.

Gene Structure and Functional Annotation

Ab initio gene prediction was performed using AUGUSTUS v3.4.0[34], trained on available *T. marneffei* genome annotations (ATCC 18224 and PM1, see Table 2 for accession numbers and training files, including GFF and cDNA sequences) to capture species–specific exon–intron structures and splice–site signals. The trained model was then applied to the B33c genome, and predicted protein sequences were extracted from GFF outputs using gffread v0.12.7[35].

Functional annotation was performed by integrating sequence similarity searches and conserved domain analyses. Predicted proteins were searched against multiple databases, including the NCBI non–redundant (Nr), Swiss–Prot, COG, GO, PHI–base[36], DFVF[37], Pfam, and KEGG, using BLAST+ v2.12.0[38] with an E–value cutoff of 1e − 5; the top-scoring (best) hit per database was retained. Functional classification was assigned with eggNOG–mapper v2.1.9[39] for COG and GO terms, while InterProScan v5.57–90.0[40] and Pfam v35.0[41] were used to identify conserved protein domains. KEGG pathway assignments were generated using KofamScan v1.3.0 (KofamKOALA)[42] under default HMM profile thresholds.

Phylogenetic Reconstruction and Divergence Time Estimation

Genomes from seven *T. marneffei* isolates and six related *Talaromyces* species were included in the phylogenomic analysis. The *T. marneffei* dataset comprised B33c (this study), ATCC 18224, PM1, TM4, GZ8H79, and two Vietnamese isolates (11CN–03–130 and 11CN–20–091) (accession numbers are provided in Table 2). Related species included *T. funiculosus*, *T. stollii*, *T. ruber*, *T. borbonicus*, *T. islandicus*, and *T. proteolyticus* (GenBank accessions: GCA_004299765.1, GCA_037040825.1, GCA_022813215.1, GCA_002916415.1, GCA_000985935.1, and GCA_021365285.1).

Orthologous gene families were identified using OrthoFinder v2.2.7[43]. Single–copy orthologs were aligned with MUSCLE v3.8.31[44], and poorly aligned regions were removed using trimAl v1.4.rev22[45] (–automated1). The trimmed alignments were concatenated to construct a maximum–likelihood phylogeny using RAxML–NG v1.1.0[46], with substitution models selected by ModelTest–NG v0.1.7[47]. Node support was assessed using 1000 bootstrap replicates in RAxML–NG.

Divergence times were estimated using MCMCTree in PAML v4.0[48] under a relaxed molecular clock model, using the maximum-likelihood topology inferred above as the guide tree and calibration points obtained from the TimeTree (v5; accessed on 1 September 2025). Convergence was assessed by examining likelihood traces and posterior distributions. Two independent MCMC runs (burn–in = 20,000; total samples = 200,000; sampling frequency = 10) achieved an effective sample size (ESS) > 200 for all parameters.

Synteny and Structural Variation Analysis

To evaluate chromosomal conservation and rearrangements, predicted proteins were compared using BLASTP v2.12.0[38] (E–value cutoff of 1e − 10), with homologous gene pairs and collinear blocks (requiring ≥ 10 consecutive genes and allowing up to six intervening genes) identified via MCScanX v1.1[49]. Using B33c as the reference genome, whole–genome alignments against other *T. marneffei* isolates and related *Talaromyces* species were performed with Minimap2 v2.24 [50] using an asm preset. Structural variations (SVs), including inversions and translocations, were subsequently identified via SyRI v1.6 [51] with default parameters and visualized as ribbon plots using plotsr v1.1.1 [52].

Shared and Species–Specific Gene Family Analysis

To characterize conserved and lineage–specific gene families, redundant protein sequences were clustered using CD–HIT v4.5.7[53] (identity ≥ 0.9; alignment coverage ≥ 0.8). The resulting non–redundant protein sets were analyzed with OrthoFinder v2.2.7[43] (–M msa). Core and species–specific gene families were defined based on orthogroup presence–absence profiles. Functional annotation of orthogroups was performed using eggNOG–mapper v2.1.9 [39] for Gene Ontology (GO) categories and KOBAS v3.0[54] for KEGG pathway enrichment. These annotations were used to summarize the functional composition of core and species–specific gene families. For strain–specific gene sets identified within *T. marneffei*, enrichment analyses were conducted using eggNOG–mapper–derived GO terms and KEGG pathway, with all annotated genes in the corresponding dataset used as the background. Statistical significance was assessed using a hypergeometric test followed by Benjamini–Hochberg false discovery rate (FDR) correction; categories with FDR < 0.05 were considered significantly enriched. Enrichment analysis and visualization were performed using the clusterProfiler R package[55].

Results

Genome Assembly and Chromosomal Landmarks

The initial de novo assembly of *T. marneffei* strain B33c yielded 28 contigs, totaling 29,107,112 bp, with a contig N50 of 3,326,235 bp. This draft assembly was further refined through Hi-C-guided scaffolding and rigorous manual curation. Consequently, 96.7% of the total sequence was successfully anchored and oriented into eight chromosomes, with lengths ranging from 1,989,838 bp to 6,430,980 bp. The final assembly spanned 28,134,402 bp (scaffold N50 = 3,696,281 bp) with no gaps (N content = 0%), representing a high-quality, telomere-to-telomere (T2T) gap-free genome (Table 1). Unanchored sequences were excluded as redundant or low-confidence after manual curation.

Table 1
Chromosome-level assembly statistics of the telomere-to-telomere (T2T) genome of *T. marneffei* strain B33c.

Number of chromosomes	Total genome size (bp)	N50 length (bp)	N90 length (bp)	GC content (%)	Longest chromosome (bp)	Shortest chromosome (bp)
8	28,134,402	3,696,281	1,998,691	46.75%	6,430,980	1,989,838

Assembly completeness was assessed using BUSCO (fungi_odb10), which recovered 752 of 758 expected orthologs (99.2% complete; 98.9% single–copy), indicating a high level of gene-space recovery (Fig. 1A). Hi–C contact maps (200–kb bins) confirmed accurate chromosomal anchoring and orientation, characterized by continuous intrachromosomal interaction diagonals and distinct inter–chromosomal boundaries across the eight chromosomes (Fig. 1B). Canonical telomeric repeats (TTAGGG)n were detected at both termini of all eight chromosomes, supporting a gap–free, telomere–to–telomere (T2T) assembly (Fig. 1C). Putative centromeres were identified in AT–rich, GC–depleted regions enriched with DNA transposons (e.g., TcMar–Fot1, hAT–Charlie) and retrotransposons (e.g., LINE/Tad1). Notably, satellite DNA signals were specifically detected on chromosomes 2 and 8 (Fig. 1C and D). These centromeric domains were associated with strong Hi-C insulation boundaries, providing additional support for their chromosomal localization.

Notes: (A) BUSCO assessment (fungi_odb10) indicating high gene-set completeness (99.2%). (B) Whole-genome Hi-C contact map at 200-kb resolution, displaying strong intrachromosomal interaction signals and clear chromosome boundaries, supporting the gap-free assembly of eight chromosomes. (C) Putative centromeric domains inferred from AT-rich/GC-depleted tracts with transposable-element enrichment and supported by Hi-C insulation boundaries across all chromosomes. (D) Chromosome ideograms showing genome-wide tracks of transposable elements and satellite/tandem-repeat signals.

Compared with previously released *T. marneffei* assemblies (Table 2), which generally show high BUSCO completeness (98.5–99.2%) yet remain non–T2T and/or multi-scaffold (8–13 sequences), the B33c genome substantially improves assembly contiguity and structural integrity. Clinical strains such as PM1 (13 sequences) and GZ8H79 (10 sequences), as well as the widely used environmental reference ATCC 18224 (11 sequences; assembled in 2024), remain fragmented. Notably, a Vietnamese assembly annotated as “chromosome-level” in NCBI (11CN–03–130) still comprises nine scaffolds. In contrast, the B33c assembly represents the first telomere-to-telomere (T2T), gap-free chromosomal reference, spanning 28,134,402 bp across eight fully resolved chromosomes with 99.2% BUSCO completeness and a low duplication rate of 0.3%.

Table 2
Comparison of genome assembly quality across *T. marneffei* strains.

Strain	NCBI Assembly ID	Origin (Location/Host)	Year (Isolation/ Assembly)	Assembly size (bp)	No. of sequences	N50leghth (bp)	Assembly level	BUSCO completeness (%)[2]	BUSC dupli (%)
B33c	—	Guangxi, China / Human	1984/2025	28,134,402	8	3,696,281	Complete (T2T,gap-free)	99.2%	0.3%
PM1	GCA_000750115.2	HongKong, China / Human	1992/2023	29,022,615	13	3,334,497	Contig[1]	99.0%	0.7%
TM4	GCA_003971505.1	Guangxi, China / Human	2014/2018	28,307,147	8	3,708,933	Complete[1]	98.8%	0.5%
GZ8H79	GCA_013122295.1	Guangzhou, China / Human	2007/2020	28,376,234	10	3,658,854	Contig[1]	98.9%	0.4%
11CN-03-130	GCA_009650675.1	Vietnam, / Human	2019/2019	28,216,733	9	3,743,714	Chromosome[1]	99.2%	0.3%
11CN-20-091	GCA_009556855.1	Vietnam, / Human	2019/2019	28,198,338	8	3,704,010	Complete[1]	98.9%	0.4%
ATCC18224	GCA_046127475.1	Vietnam, / Bamboo rat	1956/2024	28,872,128	11	4,186,281	Contig[1]	98.5%	1.8%

Notes: [1]Assembly level was defined according to the NCBI Assembly database annotation. [2]BUSCO assessment was performed using BUSCO v5.3.2 with th *fungi_odb10* database.

Repeat Elements and ncRNA Annotation

A total of 3,975 interspersed repeats, spanning 1,459,378 bp (accounting for 5.18% of the genome), were identified in the B33c assembly (Table S2). The repetitive landscape was dominated by DNA transposons (1.47%), LTR retrotransposons (1.18%), and LINE elements (0.96%), complemented by trace amounts of SINEs and rolling–circle (RC) transposons (both < 0.02%). Unclassified elements represented 1.56% of the assembly. Circos visualization revealed a heterogeneous chromosomal distribution of transposable elements (TEs), with pronounced accumulation in centromeric regions and occasional telomere-proximal accumulations, whereas gene-rich regions were largely depleted (Fig. S1). Tandem Repeat Finder (TRF) identified 4,685 tandem repeats (1.08% of the genome), comprising non-satellite tandem repeats (0.61%), minisatellites (0.38%), and microsatellites (0.09%). Genome-wide TRF profiles exhibited discrete peaks, with particularly robust signals on chromosomes 2 and 8, aligning with the identified centromeric landmarks (Fig. 1C and D). Post-masking annotation yielded 140 non-coding RNAs (ncRNAs), including 98 tRNAs (mean length 93 bp), 26 rRNAs (mean length 309 bp), 13 snRNAs, and 3 sRNAs. Consistent with other filamentous ascomycetes, no canonical microRNAs were annotated (Table S2).

Gene Structure and Functional Annotation

The *T. marneffei* B33c genome encodes 9,635 protein-coding genes spanning 17,011,444 bp (60.46% of the assembly), with 30,803 exons and 21,168 introns, corresponding to an overall gene density of ~ 342 genes per Mb (Table S3). Functional annotation assigned putative functions to 99.78% (9,614/9,635) of genes based on at least one public database hit (Table S4), leaving only 21 genes unannotated. COG, GO, and KEGG profiles were dominated by core metabolism (amino acid, carbohydrate, and energy metabolism), protein turnover, signal transduction, and secondary-metabolite biosynthesis, with general-function prediction and unknown-function classes also prominent (Fig. S2A–C).

Phylogenetic Relationships and Divergence Time Estimation

Phylogenetic relationships among *Talaromyces* taxa—comprising seven *T. marneffei* isolates and six closely related species—were reconstructed based on 3,314 single-copy orthologs, with divergence times inferred using a Bayesian relaxed molecular clock (Fig. 2). The resulting phylogeny was well resolved, with strong support for major nodes.

- Notes: Maximum-likelihood phylogeny inferred from 3,314 single-copy orthologs with node support shown as bootstrap values; divergence times were estimated with MCMCTree using TimeTree-derived calibrations and are shown with 95% HPD intervals.

Using TimeTree-derived calibrations in MCMCTree, the earliest diversification within *Talaromyces* was estimated at ~ 15.9 million years ago (Mya) (95% HPD: 1.7–54.1 Mya), with wide credible intervals reflecting calibration uncertainty. The divergence between the early-branching relative *T. borbonicus* and the remaining taxa was estimated at ~ 12.6 Mya, followed by the split of *T. marneffei* from its closest sister clade (consisting of *T. funiculosus*, *T. stollii*, and *T. ruber*) at ~ 10.0 Mya.

Within *T. marneffei*, two distinct lineages diverged at ~ 6.8 Mya (95% HPD: 0.7–23.3 Mya). One lineage includes the clinical isolate B33c, the reference strain ATCC 18224, and the Vietnamese isolates (11CN–03–130 and 11CN–20–091). Notably, B33c showed shallow divergence from the Vietnamese isolates (~ 0.3–0.4 Mya), while ATCC 18224 diverged earlier at ~ 1.0 Mya. The second lineage, comprising isolates GZ8H79, TM4, and PM1, exhibited very recent diversification (< 0.1 Mya), suggesting limited divergence among these isolates.

Synteny and Structural Variation within and across Species

Comparative synteny analyses anchored on the *T. marneffei* B33c reference revealed a gradient of chromosomal conservation across representative *Talaromyces* species, ranging from near-linear macrosynteny in *T. funiculosus* to extensive rearrangements in *T. ruber* (Fig. 3A–C). *T. funiculosus* showed the highest structural conservation, whereas *T. stollii* and *T. ruber* exhibited progressively more frequent inversions and translocations; more distantly related species (*T. borbonicus*, *T. islandicus*, and *T. proteolyticus*) displayed weak macrosynteny and fragmented collinearity (Fig. S3A–C).

Among *T. marneffei* isolates, the overall macrosyntenic framework was largely preserved, but rearrangements showed a recurrent, chromosome-biased distribution, most prominently on Chr2 with additional events on Chr3/Chr4/Chr6 (and occasionally Chr7) (Fig. 3D–F; Fig. S3D–L). Inferred SVs were dominated by inversions, with fewer translocation-consistent patterns. Within the first lineage, ATCC 18224 remained broadly collinear with B33c but carried chromosome-biased differences, while the Vietnamese isolates showed lineage-specific inversion/translocation patterns, including a Chr2–Chr6 translocation-consistent pattern in 11CN–20–091 (Fig. 3D; Fig. S3D, S3F, S3J). Within the second lineage, rearrangements were isolate-specific: GZ8H79 was largely collinear with B33c, TM4 showed intermediate restructuring dominated by Chr2, and PM1 exhibited the most extensive remodeling with fragmented syntenic blocks and pervasive rearrangements (Fig. 3F; Fig. S3G–H, S3I, S3K–L).

Notes: (A–C) Dot plots showing chromosomal conservation between *T. marneffei* B33c (x-axis) and *T. funiculosus*, *T. stollii*, and *T. ruber* (y-axis). Red and blue dots/segments indicate forward and reverse alignments, respectively (see legend).(D–F) Linear ribbon plots illustrating intraspecific macrosynteny between B33c and ATCC 18224, 11CN–03–130, and GZ8H79; ribbon colors indicate collinear blocks, inversions, and translocations (see legend).

Core and Species–Specific Gene Families within and across Species

Orthogroup clustering across seven *Talaromyces* species identified 2,810 core gene families shared by all taxa, forming a conserved genomic foundation of the genus (Fig. 4A). These core families were significantly enriched (FDR < 0.05) for essential metabolic and translational processes (Fig. 4D; Fig. S4A–B). In contrast, *T. marneffei* contained 1,215 species-specific gene families, enriched for functions related to transcriptional regulation, DNA replication/repair, stress response, and secondary-metabolite biosynthesis (Fig. 4E; Fig. S4C–D).

Notes: (A) Flower diagram showing genus-level core orthogroups shared by seven *Talaromyces* species (2,810) and *T. marneffei*–specific orthogroups (1,215).(B) Flower diagram showing shared (6,945) and isolate-specific orthogroups among seven *T. marneffei* isolates.(C) Venn diagram comparing orthogroups between the clinical isolate B33c and the environmental isolate ATCC 18224, highlighting shared and isolate-specific orthogroups.(D) COG functional classification of genus-level core orthogroups.(E) COG functional classification of *T. marneffei*–specific orthogroups.

Discussion

Recent telomere-to-telomere (T2T) genome assemblies in fungal pathogens have improved our understanding of chromosomal organization and genome evolution by resolving repetitive and structurally complex regions—such as centromeres and telomere-proximal regions—that were previously difficult to assemble and annotate[56, 57]. In the present study, we generated the first gap–free, T2T reference genome for *T. marneffei* (strain B33c), achieving full chromosomal continuity alongside high–confidence structural and functional annotations. This T2T framework provides a high–resolution foundation for the integrated analysis of genome architecture, evolutionary history, and phylogeographic structure in this medically significant, thermally dimorphic pathogen.

The final assembly spans 28.1 Mb across eight chromosomes and encodes 9,635 protein–coding genes, a count consistent with previous assemblies[14–18]. However, this study achieves substantially improved chromosomal completeness by resolving formerly recalcitrant telomeric, centromeric, and repeat–rich regions. In this T2T genome, we identified putative centromeres (14–23.5 kb; mean 19.5 kb) that are heavily enriched in DNA transposons and LTR retrotransposons. Notably, satellite DNA signals were detected on chromosomes 2 and 8. As centromeres are increasingly recognized as hotspots for karyotype dynamics and structural innovation [58–60], our ability to fully resolve them provides a critical baseline for understanding *T. marneffei* lineage diversification.

Beyond centromeres, our genome–wide landscape analysis reveals a moderate overall repeat content (~ 6.3%, including 5.18% interspersed repeats) and high gene density (~ 342 genes per Mb), characteristic of a compact ascomycete architecture [61]. Although the total transposable element (TE) content is relatively low (~ 5.18%), their non–random enrichment near centromeric and telomere-proximal regions may be functionally relevant. For instance, a previously reported ~ 740-kb segmental duplication in the environmental isolate ATCC 18224—likely mediated by Ty1/Copia LTR retrotransposons—highlights how even a streamlined TE repertoire can catalyze substantial structural variation (SV) [14].

There is compelling evidence that fungal genomes are extensively transcribed, yielding a diverse array of non–coding RNAs (ncRNAs) critical for gene regulation, stress adaptation, and pathogenicity[62]. The high contiguity of the B33c T2T assembly facilitated a high–confidence annotation of the ncRNA landscape, in which we identified 140 ncRNAs, including tRNAs, rRNAs, snRNAs, and sRNAs. Notably, no canonical microRNAs were annotated, aligning with the sparse microRNA profiles observed in many filamentous ascomycetes. In other fungal pathogens, such as *Cryptococcus neoformans* and *Candida auris*, ncRNAs have been implicated in gene regulation and stress-associated processes[63, 64]. Our T2T-based annotation serves as a valuable resource for future functional studies on how the non-coding transcriptome modulates life cycle and virulence.

Previous phylogenetic studies of *Talaromyces* have largely relied on multilocus markers [65, 66], which often lack sufficient resolution to decipher complex evolutionary relationships. In this study, we reconstructed a genome–scale, time–calibrated phylogeny of *T. marneffei* based on thousands of single–copy orthologs, providing a genome-scale evolutionary timescale for this pathogen. At the genus level, diversification was estimated to have commenced ~ 15.9 Mya, broadly consistent with the diversification patterns reported for the *Aspergillaceae* [67, 68]. Following speciation, *T. marneffei* underwent early intraspecific diversification, with major lineages diverging approximately 6.8 Mya.

Notably, our phylogenetic analysis reveals that while ATCC 18224 occupies an early–diverging position, the B33c strain clusters closely with Vietnamese isolates. This contrast suggests that the shallow divergence (~ 0.3–0.4 Mya) between B33c and the Vietnamese lineage may reflect a relatively recent cross-border genetic link. This inference is further supported by ecological niche modeling, which identifies high habitat suitability for primary *Rhizomys* hosts across contiguous monsoon-driven tropical zones, a pattern compatible with recent climate-driven risk assessments in the region[6, 69]. Collectively, these lines of evidence support a possible “One Health” interpretation of cross-border genetic connectivity, linked to shared reservoir habitats and regional ecological continuity, as further suggested by geographic risk factors identified in recent regional cohorts [70, 71]. In contrast, recent southern China isolates (e.g., PM1 from Hong Kong, TM4 from Guangxi, and GZ8H79 from Guangzhou) form a distinct cluster, supporting geographically structured sublineages within *T. marneffei* and suggesting regional diversification shaped by ecological factors and host interactions. Similar population studies of Vietnamese isolates have also highlighted substantial genetic diversity across regions [19].

To connect this evolutionary framework with genome–scale features, we examined chromosomal synteny and structural variation across *Talaromyces* using the B33c T2T reference as an anchor. At the interspecific level, *T. marneffei* remains highly collinear with *T. funiculosus*, whereas the *T. stollii–T. ruber* lineage shows pronounced fragmentation, inversions, and translocations. More distantly related species *(T. borbonicus*, *T. islandicus*, and *T. proteolyticus*) exhibit weak or highly fragmented macrosynteny, reflecting extensive chromosomal reorganization during divergence. Placed within the divergence–time framework, these patterns imply heterogeneous rates of karyotype evolution across the genus. Similar genome remodeling has been documented in other filamentous fungi, where structural variant landscapes directly influence gene function[72] and contribute to the divergence of chemical and pathogenic traits between close relatives[73]. Such large–scale rearrangements can reposition secondary–metabolite and virulence–associated loci, potentially driving the ecological differentiation and unique pathogenic traits observed in *T. marneffei*.

At the intraspecific level, overall chromosomal architecture is largely conserved, yet structural variation persists and is non–randomly distributed, frequently involving repeat–rich pericentromeric and telomere-proximal domains, as suggested by their spatial coincidence with repeat-enriched regions in the T2T assembly. Comparable dynamics have been reported in opportunistic fungal pathogens[74], where these regions often serve as hubs of genome plasticity, gene–family diversification, and adaptive evolution. In *T. marneffei*, the accumulation of structural variation within repeat–enriched domains is consistent with repeat–mediated recombination and localized genome instability contributing to microevolutionary differentiation among closely related lineages.

The “conserved core–flexible periphery” pattern observed at the structural level is also reflected by gene–family evolution. Orthogroup analyses across *Talaromyces* species and *T. marneffei* isolates reveal a conserved genomic core underpinning essential processes (metabolism, translation, energy production), alongside a more flexible accessory repertoire associated with lineage–specific innovation. Across species, *T. marneffei* retains a large set of species–enriched families linked to transcriptional regulation, DNA repair, stress response, and secondary–metabolite biosynthesis—paralleling trends observed in *Aspergillus*[75] and supporting the idea that lineage–specific genes can contribute to ecological adaptation and pathogenic potential. Within *T. marneffei* specifically, this accessory flexibility is evidenced by the pronounced differences between the environmental isolate ATCC 18224 and the clinical isolate B33c. While ATCC 18224 retains an expanded accessory repertoire enriched for nutrient acquisition and metal–ion homeostasis—reflecting fluctuating pressures in natural reservoirs—B33c exhibits a comparatively streamlined accessory complement, suggesting evolutionary specialization toward host–associated conditions. Similar patterns of lineage-specific accessory genome variation have been reported in other fungal pathogens, such as

Aspergillus species[76], *Cryptococcus gattii*[77], and *Candida albicans*[78]. Notably, broader sampling remains essential to determine the universality of these patterns across diverse environmental and clinical populations.

Taken together, this study provides the first T2T genomic framework for *T. marneffei*, enabling integrated analyses of chromosomal structure, structural variation, and gene–family evolution, and supporting an adaptive strategy that combines genomic stability with localized flexibility. Several limitations should be noted. First, although B33c provides a high–quality reference, conclusions about population–wide diversity remain constrained by the limited number of isolates and the scarcity of environmental genomes. Second, while structural variation showed a recurrent, chromosome-biased distribution, its functional relevance and potential links to adaptation or virulence require experimental validation. Future work should expand long–read–based genome resources from diverse geographic regions and ecological niches, and integrate multi–omics approaches (e.g., transcriptomics and functional assays) to connect genomic variation with phenotypes relevant to pathogenicity and clinical outcomes.

Conclusion

We report the first telomere-to-telomere (T2T), gap-free genome assembly of *T. marneffei* (strain B33c), a landmark clinical isolate from China. This 28.1-Mb assembly resolves all eight chromosomes, providing a chromosome-complete reference and improved annotation of repeat-rich pericentromeric and telomere-proximal regions. Phylogenomics suggests cross-border connectivity, reflected by shallow divergence and close clustering between northern Vietnam and southern China, while comparative synteny reveals a conserved macrosyntenic framework with localized, repeat-associated, chromosome-biased structural variation. Orthogroup analyses further highlight a flexible accessory repertoire enriched for stress response and secondary metabolism, with accessory-genome differences between the environmental isolate ATCC 18224 and the clinical isolate B33c. Collectively, this B33c T2T reference provides a foundation for future studies of population structure, structural plasticity, and pathogenic evolution in *T. marneffei*.

Declarations

Author Contributions

Conceptualization, Project Administration, Funding Acquisition, and Supervision, C.C.; Supervision, C.K.M.T.; Investigation, X.Z., Y.Z., and S.K.; Data Curation, X.Z., Y.Z., S.K., Z.J., K.P., D.Y.Z., X.L., L.H., G.L., and J.L.; Formal Analysis and Validation, X.Z.; Visualization, X.Z., Z.J., K.P., D.Y.Z., X.L., L.H., G.L., and J.L.; Writing—Original Draft, X.Z. and S.K.; Writing—Review & Editing, C.C. and C.K.M.T. All authors have read and approved the manuscript.

Funding

This study was supported by the National Key Research and Development Program of China (2022YFC2504803), the Natural Science Foundation of Guangxi Province of China (AB24010134), and the National Natural Science Foundation of China (U25A2088). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Data Availability

The raw sequencing reads generated in this study are publicly available in the CNGB Sequence Archive (CNSA) under accession number CNP0008012. The genome assembly and associated annotation files, including the GFF3 annotation, CDS sequences, and predicted protein sequences, are publicly available in figshare at https://figshare.com/articles/dataset/The_First_Telomere_to_Telomere_Genome_Assembly _of_Talaromyces_marneffei_B33c_Insights_into _Chromosomal_Evolution_Cross_Border _Connectivity_and_Niche_Adaptation/31899742. No VCF files were generated in this study.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

During the preparation of this manuscript, the authors used AI-assisted tools for language editing and grammar correction. The authors reviewed and edited the output and take full responsibility for the content of this publication. We sincerely thank Professor Weida Liu for kindly providing the archived isolate used in this study. We also acknowledge the Chinese Academy of Medical Sciences Collection Center of Pathogen Microorganisms for the long-term preservation and management of the strains.

References

1. DiSalvo AF, Fickling AM, Ajello L. Infection caused by *Penicillium marneffei*: description of first natural infection in man. Am J Clin Pathol. 1973;60:259–263.
2. Vanittanakom N, Cooper CR Jr, Fisher MC, Sirisanthana T. *Penicillium marneffei* infection and recent advances in the epidemiology and molecular biology aspects. Clin Microbiol Rev. 2006;19:95–110.
3. Cao C, Xi L, Chaturvedi V. Talaromycosis (penicilliosis) due to *Talaromyces* (*Penicillium*) *marneffei*: insights into the clinical trends of a major fungal disease 60 years after the discovery of the pathogen. Mycopathologia. 2019;184:709–720.
4. Tsang CC, Lau SKP, Woo PCY. Sixty years from Segretain's description: what have we learned and should learn about the basic mycology of *Talaromyces marneffei*? Mycopathologia. 2019;184:721–729.

5. Wang F, Han R, Chen S. An overlooked and underrated endemic mycosis-talaromycosis and the pathogenic fungus *Talaromyces marneffei*. Clin Microbiol Rev. 2023;36:e0005122.
6. Bao XL, Wei WD, He JH, Zhong XN, Li XP. Potential distribution of *Talaromyces marneffei* infections worldwide based on the Maxent model. Chin Trop Med. 2023;23:10–15. (In Chinese).
7. Li JS, Pan LQ, Wu SX, Su SX, Su SB, Shan LY. Disseminated *penicilliosis marneffei* in China. Report of three cases. Chin Med J (Engl). 1991;104:247–251.
8. Deng Z, Ribas JL, Gibson DW, Connor DH. Infections caused by *Penicillium marneffei* in China and Southeast Asia: review of eighteen published cases and report of four more Chinese cases. Rev Infect Dis. 1988;10:640–652.
9. Hu Y, Zhang J, Li X, Yang Y, Zhang G, Zheng J, Cao C. *Penicillium marneffei* infection: an emerging disease in mainland China. Mycopathologia. 2013;175:57–67.
10. Zhou LH, Jiang YK, Li RY, Huang LP, Pang JW, Tong ZH, Li R, Chen SC, Liu W, Pan WH, et al. Risk-based estimate of human fungal disease burden, China. Emerg Infect Dis. 2020;26:2137–2147.
11. Chan JF, Lau SK, Yuen KY, Woo PC. *Talaromyces* (*Penicillium*) *marneffei* infection in non-HIV-infected patients. Emerg Microbes Infect. 2016;5:e19.
12. Zheng YQ, Li QG, Latge JP, Chen P, Li L, He YZ, Huang JH, Nong JY, Lu DH, Pan PH, et al. Potential risk of cross-resistance to voriconazole in HIV/AIDS patients with *Talaromyces marneffei* infection and the mechanisms of the cross-resistance. J Antimicrob Chemother. 2025;80(4):976–979.
13. Guo P, Chen W, Chen S, Jiang J, Zeng W, Lu S, Lin H, Wang W, Zheng S. The delayed clearance of *Talaromyces marneffei* in blood culture may be associated with higher MIC of voriconazole after antifungal therapy among AIDS patients with talaromycosis. PLoS Negl Trop Dis. 2023;17:e0011201.
14. Luo J, Bian J, Murillo M, Zhang R, Liang J, Li X, Fan J, Wei L, Chen S, Cuomo CA, et al. High-quality genome assembly and comparative analysis reveal extensive genomic variation in *Talaromyces marneffei*. Microb Genom. 2025;11:001400.
15. Yang E, Chow WN, Wang G, Woo PCY, Lau SKP, Yuen KY, Tsang CC, Chan CM, Tse H, Kao RYT. Signature gene expression reveals novel clues to the molecular mechanisms of dimorphic transition in *Penicillium marneffei*. PLoS Genet. 2014;10:e1004662.
16. Cuomo CA, Shea T, Nguyen T, Ashton P, Perfect J, Le T. Complete genome sequences for two *Talaromyces marneffei* clinical isolates from northern and southern Vietnam. Microbiol Resour Announc. 2020;9:e01367-19.
17. Li Y, Luo H, Fan J, Deng J, Wu J, Zhang G, He G. Genomic analysis provides insights into the transmission and pathogenicity of *Talaromyces marneffei*. Fungal Genet Biol. 2019;130:54–61.
18. Yang H, Liu M, Xu N, He Y, Wang Q, Zhang Z, Liang L, Wang L, Zhang J, Gao G. Disseminated *Talaromyces marneffei* infection associated with haemophagocytic syndrome in an HIV-negative patient in northern China: a case report. BMC Infect Dis. 2024;24:63.
19. Sephton-Clark P, Nguyen T, Hoa NT, Ashton P, Van Anh DT, Van Vinh Chau N, Beardsley J, Perfect JR, Cuomo CA, Le T. Impact of pathogen genetics on clinical phenotypes in a population of *Talaromyces marneffei* from Vietnam. Genetics. 2023;224:iyad100.
20. Cheng H, Concepcion GT, Feng X, Zhang H, Li H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. Nat Methods. 2021;18:170–175.
21. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv. 2013;arXiv:1303.3997.
22. Durand NC, Shamim MS, Machol I, Chen SS, St Hilaire BG, Aden JK, Zuckerman L, Fan AY, Dudchenko AD, Aiden EL. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. Cell Syst. 2016;3:95–98.
23. Zhang X, Zhang S, Zhao Q, Ming R, Tang H. Assembly of allele-aware, chromosomal-scale autopolyploid genomes based on Hi-C data. Nat Plants. 2019;5:833–845.
24. Durand NC, Robinson JT, Shamim MS, Machol I, Mesirov JP, Lander ES, Aiden EL. Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. Cell Syst. 2016;3:99–101.
25. Lin Y, Ye C, Li X, Chen Z, Lu J, Ye Y, Yang G, Tang H. quarTeT: a telomere-to-telomere toolkit for gap-free genome assembly and centromeric repeat identification. Hortic Res. 2023;10:uhad127.
26. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010;26:841–842.
27. 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 USA. 2020;117:9451–9457.
28. Manni M, Berkeley MR, Seppey M, Simao 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;38:4647–4654.
29. Chen N. Using RepeatMasker to identify repetitive elements in genomic sequences. Curr Protoc Bioinformatics. 2004;4:4.10.
30. Benson G. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. 1999;27:573–580.
31. Chan PP, Lowe TM. tRNAscan-SE: searching for tRNA genes in genomic sequences. Methods Mol Biol. 2019;1962:1–14.
32. Lagesen K, Hallin P, Rodland EA, Staerfeldt HH, Rognes T, Ussery DW. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res. 2007;35:3100–3108.
33. Nawrocki EP, Eddy SR. Infernal 1.1: 100-fold faster RNA homology searches. Bioinformatics. 2013;29:2933–2935.
34. Stanke M, Waack S. Gene prediction with a hidden Markov model and a new intron submodel. Bioinformatics. 2003;19:ii215-ii225.
35. Pertea G, Pertea M. GFF utilities: GffRead and GffCompare. F1000Res. 2020;9:304.
36. Urban M, Cuzick A, Seager J, Wood V, Rutherford K, Venkatesh SY, De Silva N, Martinez MC, Pedro H, Yates AD, et al. PHI-base in 2022: a curated resource for pathogen-host interactions. Nucleic Acids Res. 2022;50:D940-D947.
37. Lu T, Yao B, Zhang C. DFVF: database of fungal virulence factors. Database (Oxford). 2012;2012:bas032.

38. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. BLAST+: architecture and applications. BMC Bioinformatics. 2009;10:421.
39. Cantalapiedra CP, Hernandez-Plaza A, Letunic I, Bork P, Huerta-Cepas J. eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. Mol Biol Evol. 2021;38:5825–5829.
40. Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, et al. InterProScan 5: genome-scale protein function classification. Bioinformatics. 2014;30:1236–1240.
41. Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, Sonnhammer ELL, Tosatto SCE, Paladin L, Raj S, Richardson LJ, et al. Pfam: the protein families database in 2021. Nucleic Acids Res. 2021;49:D412-D419.
42. Aramaki T, Blanc-Mathieu R, Endo H, Ohkubo K, Kanehisa M, Goto S, Ogata H. KofamKOALA: KEGG ortholog assignment based on profile HMM and adaptive score threshold. Bioinformatics. 2020;36:2251–2252.
43. Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. 2019;20:238.
44. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004;32:1792–1797.
45. Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics. 2009;25:1972–1973.
46. Kozlov AM, Darriba D, Flouri T, Morel B, Stamatakis A. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. Bioinformatics. 2019;35:4453–4455.
47. Darriba D, Posada D, Kozlov AM, Stamatakis A, Morel B, Flouri T. ModelTest-NG: a new and scalable tool for the selection of DNA and protein evolutionary models. Mol Biol Evol. 2020;37:291–294.
48. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 2007;24:1586–1591.
49. Wang Y, Tang H, DeBarry JD, Tan X, Li J, Wang X, Paterson AH. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Res. 2012;40:e49.
50. Li H. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics. 2018;34:3094–3100.
51. Goel M, Habermann B, Schneeberger K. SyRI: synteny and rearrangement identifier of structurally polymorphic genomes. Genome Biol. 2019;20:179.
52. Schwab C, Habermann B, Schneeberger K, Goel M. Plotsr: visualizing structural variations and genome-to-genome comparisons. Bioinformatics. 2022;38:3257–3259.
53. Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics. 2012;28:3150–3152.
54. Bu D, Luo H, Huo P, Wang Z, Zhang S, He Z, Wu Y, Zhao L, Liu J, Guo J, et al. KOBAS-i: intelligent prioritization and exploratory visualization of biological functions for gene enrichment analysis. Nucleic Acids Res. 2021;49:W317-W325.
55. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS. 2012;16:284–287.
56. Bowyer P, Currin A, Delneri D, Fraczek MG. Telomere-to-telomere genome sequence of the model mould pathogen *Aspergillus fumigatus*. Nat Commun. 2022;13:5394.
57. Hoyer LL, Freeman BA, Hogan EK, Hernandez AG. Use of a *Candida albicans* SC5314 PacBio HiFi reads dataset to close gaps in the reference genome assembly, reveal a subtelomeric gene family, and produce accurate phased allelic sequences. Front Cell Infect Microbiol. 2024;14:1329438.
58. Coelho MA, David-Palma M, Shea T, Walker LA, Heitman J. Comparative genomics of the closely related fungal genera *Cryptococcus* and *Kwoniella* reveals karyotype dynamics and suggests evolutionary mechanisms of pathogenesis. PLoS Biol. 2024;22:e3002682.
59. Smith KM, Galazka JM, Phatale PA, Connolly LR, Freitag M. Centromeres of filamentous fungi. Chromosome Res. 2012;20:635–656.
60. Barra V, Fachinetti D. The dark side of centromeres: types, causes and consequences of structural abnormalities implicating centromeric DNA. Nat Commun. 2018;9:4340.
61. Galagan JE, Calvo SE, Cuomo C, Ma LJ, Wortman JR, Batzoglou S, Lee SI, Basturkmen M, Spevak CC, Clutterbuck J, et al. Sequencing of *Aspergillus nidulans* and comparative analysis with *A. fumigatus* and *A. oryzae*. Nature. 2005;438:1105–1115.
62. Li L, Luo Y, Li S. Non-coding RNA in fungi-a review. Wei Sheng Wu Xue Bao. 2013;53(8):790–797. (In Chinese).
63. Kalem MC, Panepinto JC. Long non-coding RNAs in *Cryptococcus neoformans*: insights into fungal pathogenesis. Front Cell Infect Microbiol. 2022;12:858317.
64. Gao J, Chow EWL, Wang H, Xu X, Giam CLX, Teo JDW, Thio SK, Kouroutzidis DP, Tan ASM, Koh JJH, et al. LncRNA DINOR is a virulence factor and global regulator of stress responses in *Candida auris*. Nat Microbiol. 2021;6:842–851.
65. 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;70:159–183.
66. Yilmaz N, Visagie CM, Houbraken J, Frisvad JC, Samson RA. Polyphasic taxonomy of the genus *Talaromyces*. Stud Mycol. 2014;78:175–341.
67. Samson RA, Visagie CM, Houbraken J, Hong SB, Hubka V, Klaassen CHW, Perrone G, Seifert KA, Susca A, Tanney JB, et al. Phylogeny, identification and nomenclature of the genus *Aspergillus*. Stud Mycol. 2014;78:141–173.
68. Steenwyk JL, Shen XX, Lind AL, Goldman GH, Rokas A. A robust phylogenomic time tree for biotechnologically and medically important fungi in the genera *Aspergillus* and *Penicillium*. mBio. 2019;10:e00925-19.
69. Xu N, Min X, Wu K, La T, Cao B. Environmental transmission from bamboo rats: Mapping current and future talaromycosis risk in China under climate change. One Health. 2025;21:101264.

70. Cao C, Liang L, Wang W, Luo H, Huang S, Liu D, Xu J, Henk DA, Fisher MC. Common reservoirs for *Penicillium marneffei* infection in humans and rodents, China. Emerg Infect Dis. 2011;17:209–214.
71. Brown L, Jonat B, Ly VT, Tung NLN, Lam PS, Thanh NT, Bich DTN, Thao VP, Thu NTM, Hoa NT, Le T. Multicenter case-control study of behavioral, environmental, and geographic risk factors for talaromycosis, Vietnam. Emerg Infect Dis. 2025;31:1309–1318.
72. Dhakal U, Kim HS, Toomajian C. The landscape and predicted roles of structural variants in *Fusarium graminearum* genomes. G3 (Bethesda). 2024;14:jkae065.
73. Mead ME, Knowles SL, Raja HA, Beattie SR, Kowalski CH, Steenwyk JL, Silva LP, Chiaratto J, Ries LNA, Goldman GH, et al. Characterizing the pathogenic, genomic, and chemical traits of *Aspergillus fischeri*, a close relative of the major human fungal pathogen *Aspergillus fumigatus*. mSphere. 2019;4(1):e00018-19. doi:10.1128/mSphere.00018-19.
74. Diotti R, Esposito M, Shen CH. Telomeric and sub-telomeric structure and implications in fungal opportunistic pathogens. Microorganisms. 2021;9:1405.
75. Kjærbølling I, Vesth TC, Frisvad JC, Nybo JL, Theobald S, Kuo A, Bowyer P, Houbraken J, Lantz H, Barry K, et al. Linking secondary metabolites to gene clusters through genome sequencing of six diverse *Aspergillus* species. Proc Natl Acad Sci USA. 2018;115:E753-E761.
76. Vesth TC, Nybo JL, Theobald S, Frisvad JC, Larsen TO, Nielsen KF, Hoof JB, Brandl J, Tamayo RJ, de Vries RP, et al. Investigation of inter- and intraspecies variation through genome sequencing of *Aspergillus* section Nigri. Nat Genet. 2018;50:1688–1695.
77. Engelthaler DM, Hicks ND, Gillece JD, Roe CC, Schupp JM, Drees KP, Colman RE, Foster JT, Boro M, Keim PS, et al. *Cryptococcus gattii* in North American Pacific Northwest: whole-population genome analysis provides insights into species evolution and dispersal. mBio. 2014;5:e01464-14.
78. Butler G, Rasmussen MD, Lin MF, Santos MA, Sakthikumar S, Munro CA, Pasamontes L, Puri OH, Coleman KA, Gillin LL, et al. Evolution of pathogenicity and sexual reproduction in eight *Candida* genomes. Nature. 2009;459:657–662.

Figures

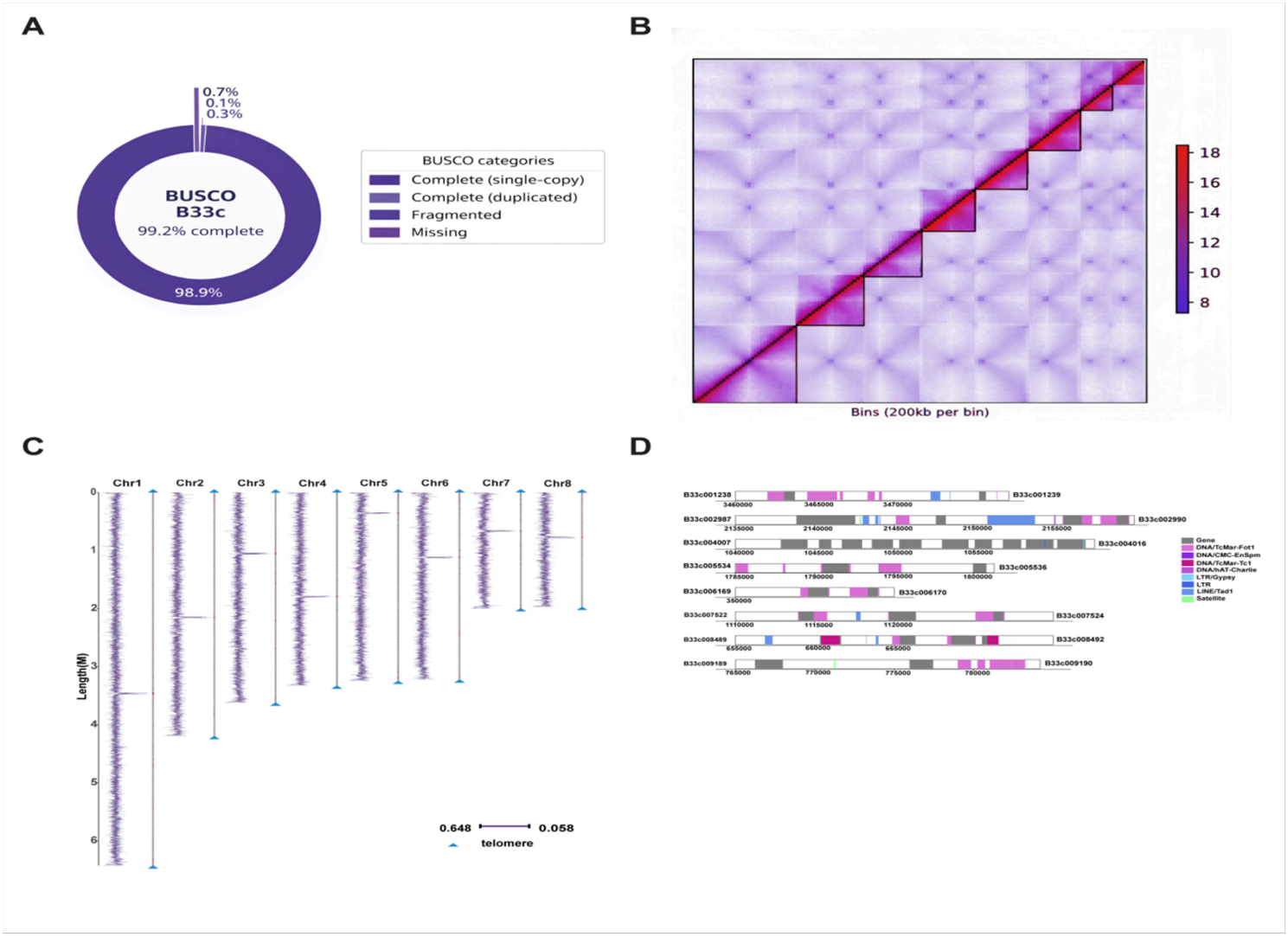


Figure 1

Chromosome-scale validation and landmark annotation of the *T. marneffei* B33c T2T assembly.

Notes：(A) BUSCO assessment (fungi_odb10) indicating high gene-set completeness (99.2%). (B) Whole-genome Hi-C contact map at 200-kb resolution, displaying strong intrachromosomal interaction signals and clear chromosome boundaries, supporting the gap-free assembly of eight chromosomes. (C) Putative centromeric domains inferred from AT-rich/GC-depleted tracts with transposable-element enrichment and supported by Hi-C insulation boundaries across all chromosomes. (D) Chromosome ideograms showing genome-wide tracks of transposable elements and satellite/tandem-repeat signals.

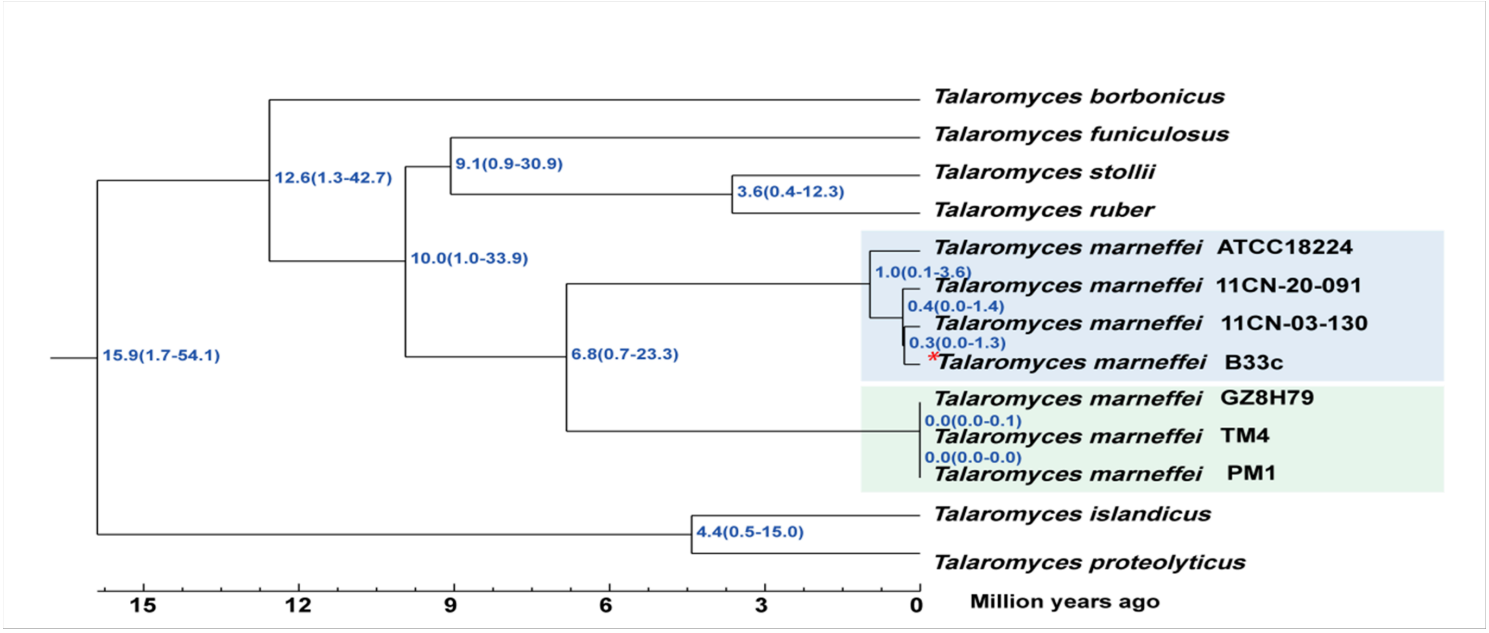


Figure 2

Phylogenomic relationships and divergence time estimation of *T. marneffei* and related *Talaromyces* species.

Notes: Maximum-likelihood phylogeny inferred from 3,314 single-copy orthologs with node support shown as bootstrap values; divergence times were estimated with MCMCTree using TimeTree-derived calibrations and are shown with 95% HPD intervals.

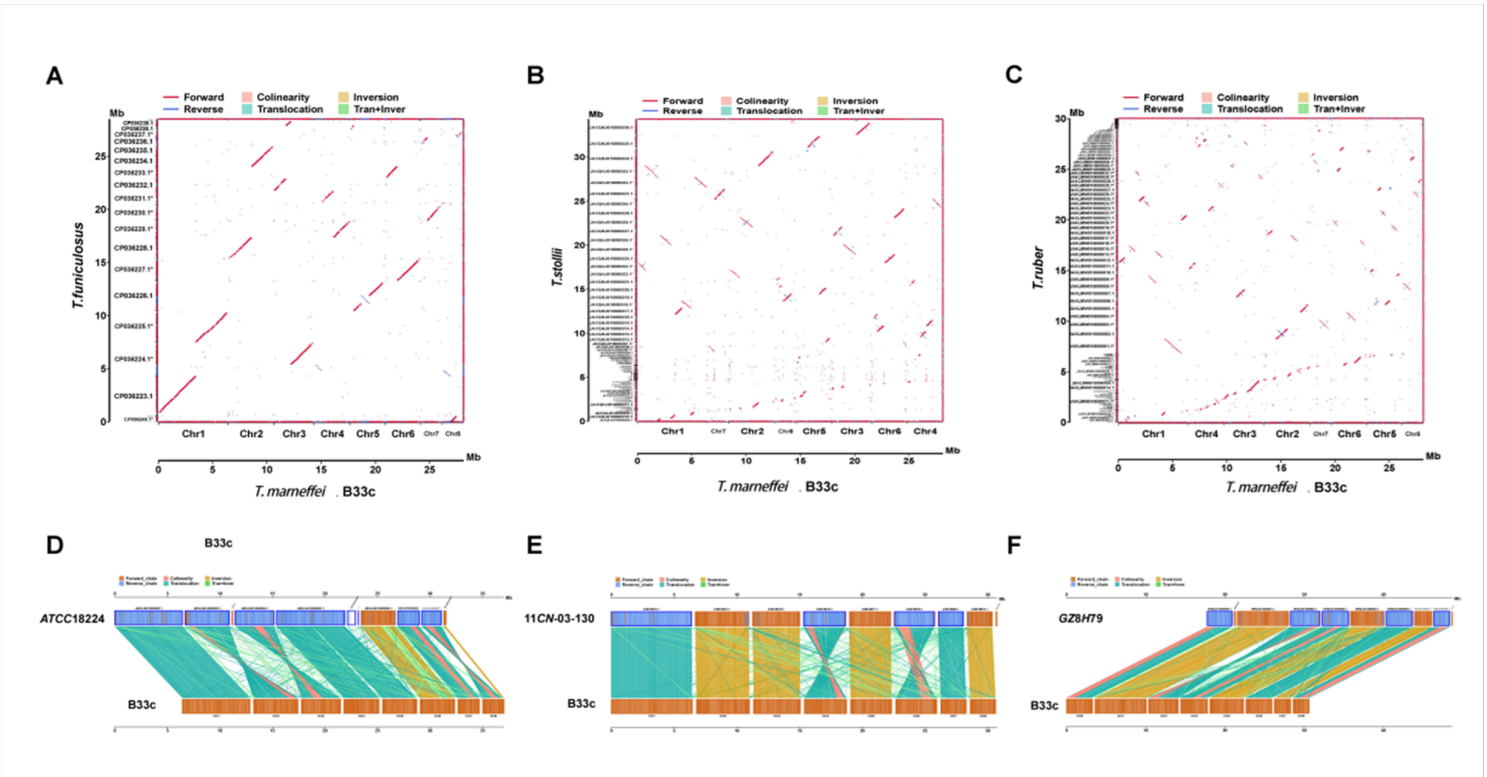


Figure 3

Comparative interspecific and intraspecific synteny and structural variation in *Talaromyces*.

Notes：(A–C) Dot plots showing chromosomal conservation between *T. marneffei* B33c (x-axis) and *T. funiculosus*, *T. stollii*, and *T. ruber* (y-axis). Red and blue dots/segments indicate forward and reverse alignments, respectively (see legend).(D–F) Linear ribbon plots illustrating intraspecific macrosynteny between B33c and ATCC 18224, 11CN–03–130, and GZ8H79; ribbon colors indicate collinear blocks, inversions, and translocations (see legend).

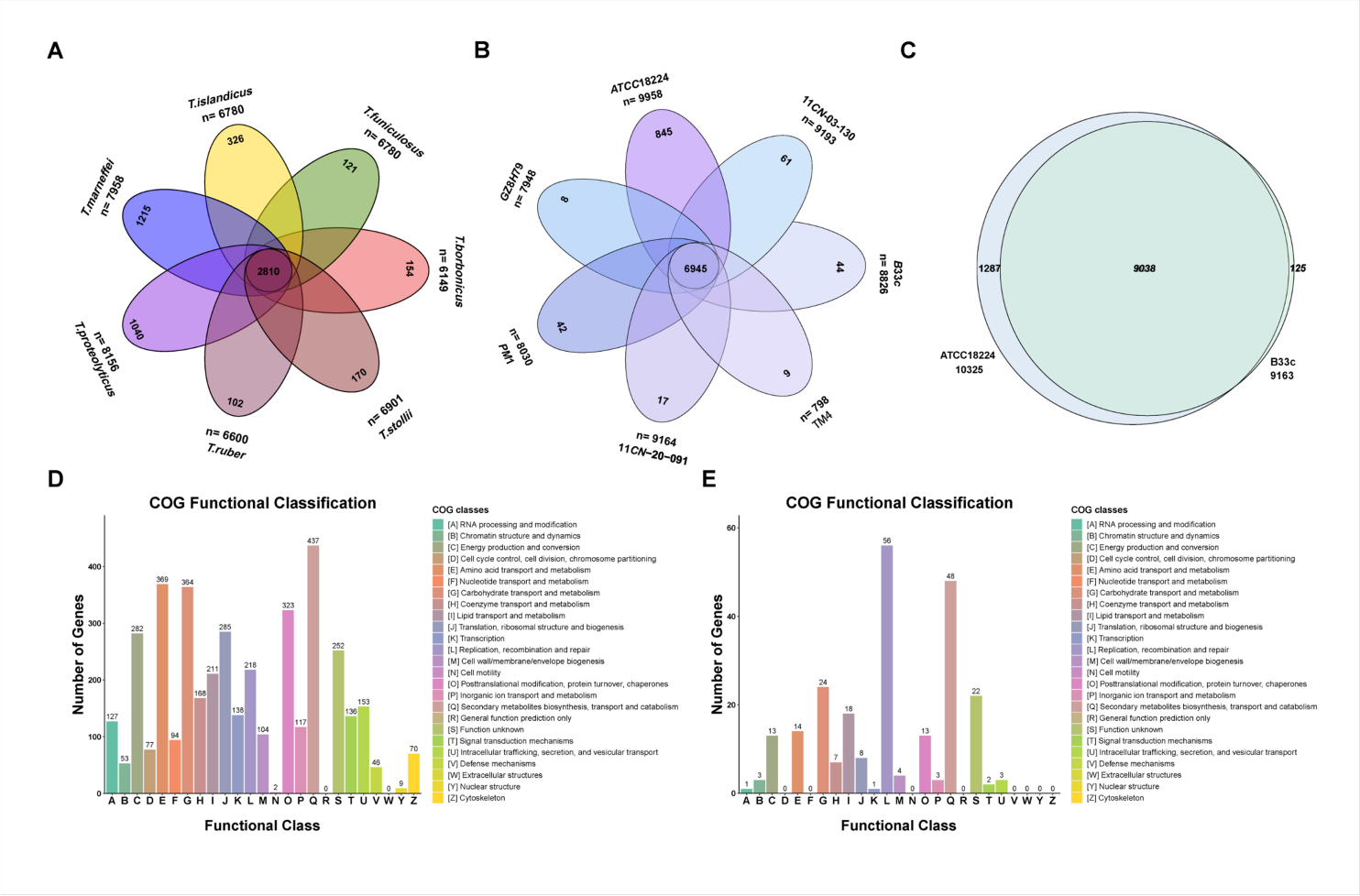


Figure 4

Comparative analysis of core and species-specific gene families and their functional enrichment across *Talaromyces*.

Notes：(A) Flower diagram showing genus-level core orthogroups shared by seven *Talaromyces* species (2,810) and *T. marneffei*–specific orthogroups (1,215).(B) Flower diagram showing shared (6,945) and isolate-specific orthogroups among seven *T. marneffei* isolates.(C) Venn diagram comparing orthogroups between the clinical isolate B33c and the environmental isolate ATCC 18224, highlighting shared and isolate-specific orthogroups.(D) COG functional classification of genus-level core orthogroups.(E) COG functional classification of *T. marneffei*–specific orthogroups.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementalmaterialllegends.docx
- SupplementaryFiguresS1S4..docx
- SupplementaryTablesS1S4.docx