Unraveling the dynamic transcriptomic changes during the dimorphic transition of *Talaromyces marneffei* through time-course analysis

Minghao Du[1†], Changyu Tao[2†], Xueyan Hu[3], Yun Zhang[3],
Jun Kan[3], Juan Wang[1] and Ence Yang[1,3]*

[1]Department of Microbiology and Infectious Disease Center, School of Basic Medical Sciences, Peking University Health Science Center, Beijing, China, [2]Department of Human Anatomy, Histology and Embryology, School of Basic Medical Sciences, Peking University Health Science Center, Beijing, China, [3]Department of Medical Bioinformatics, School of Basic Medical Sciences, Peking University Health Science Center, Beijing, China

Introduction: Systemic dimorphic fungi pose a significant public health challenge, causing over one million new infections annually. The dimorphic transition between saprophytic mycelia and pathogenic yeasts is strongly associated with the pathogenesis of dimorphic fungi. However, despite the dynamic nature of dimorphic transition, the current omics studies focused on dimorphic transition primarily employ static strategies, partly due to the lack of suitable dynamic analytical methods.

Methods: We conducted time-course transcriptional profiling during the dimorphic transition of *Talaromyces marneffei*, a model organism for thermally dimorphic fungi. To capture non-uniform and nonlinear transcriptional changes, we developed DyGAM-NS (dynamic optimized generalized additive model with natural cubic smoothing). The performance of DyGAM-NS was evaluated by comparison with seven other commonly used time-course analysis methods. Based on dimorphic transition induced genes (DTIGs) identified by DyGAM-NS, cluster analysis was utilized to discern distinct gene expression patterns throughout dimorphic transitions of *T. marneffei*. Simultaneously, a gene expression regulatory network was constructed to probe pivotal regulatory elements governing the dimorphic transitions.

Results: By using DyGAM-NS, model, we identified 5,223 DTIGs of *T. marneffei*. Notably, the DyGAM-NS model showcases performance on par with or superior to other commonly used models, achieving the highest F1 score in our assessment. Moreover, the DyGAM-NS model also demonstrates potential in predicting gene expression levels throughout temporal processes. The cluster analysis of DTIGs suggests divergent gene expression patterns between mycelium-to-yeast and yeast-to-mycelium transitions, indicating the asymmetrical nature of two transition directions. Additionally, leveraging the identified DTIGs, we constructed a regulatory network for the dimorphic transition and identified two zinc finger-containing transcription factors that potentially regulate dimorphic transition in *T. marneffei*.

Discussion: Our study elucidates the dynamic transcriptional profile changes during the dimorphic transition of *T. marneffei*. Furthermore, it offers a novel perspective for unraveling the underlying mechanisms of fungal dimorphism, emphasizing the importance of dynamic analytical methods in understanding complex biological processes.

Introduction

Systemic dimorphic fungi are significant human pathogens that cause millions of new infections annually worldwide (Klein and Tebbets, 2007). The ability of these fungi to transition between unicellular yeast and multicellular mycelial forms, known as dimorphic transition, enables them to adapt from environmental saprophytes to pathogens within hosts. In details, the mycelium-to-yeast (M-to-Y) transition is crucial for their pathogenicity as the yeast form evades the host immune system, while the yeast-to-mycelium (Y-to-M) transition is important for maintaining an environmental reservoir (Nemecek et al., 2006; Gauthier, 2017). As the highly association between fungal dimorphic transition and their pathogenicity, elucidating the genetic mechanisms of dimorphic transition will not only shed light on understanding the pathogenic mechanisms of human fungal pathogens, but also provide theoretical guidance for molecular-based anti-fungal therapy.

The rapid advancement of high-throughput sequencing technologies, especially in the field of transcriptomics, has spurred numerous investigations aiming to employ more efficient genomic techniques for a comprehensive understanding of the regulatory mechanisms underlying fungal dimorphic transition. However, many studies have predominantly relied on static transcriptomic analysis approaches, focusing on contrasting transcriptomic differences between the mycelial and yeast phases to identify regulatory genes involved in the dimorphic transition process. Given the complex regulatory network governing fungal dimorphic transition, such static transcriptomic studies often fall short in capturing the full spectrum of genes that actively regulate the transition. While some studies have made efforts to elucidate the regulatory mechanisms by utilizing transcriptomic data obtained at different time points during the dimorphic transition, the limited number of sampling time points poses constraints on the comprehensive analysis of the intricate transcriptional changes that occur throughout the transition. Consequently, there is an urgent demand for innovative research strategies that can enable efficient and accurate deciphering of fungal dimorphic transition.

Temporal transcriptome analysis is a commonly employed approach for investigating dynamic gene expression changes in complex biological processes (Bar-Joseph et al., 2012). It allows for the identification of transient gene expression patterns, response times, and gene regulatory relationships, thus holding significant potential for unraveling the regulatory mechanisms underlying dimorphic transition. However, the proper interpretation of time-course transcriptome data during dimorphic transition remains challenging. In recent years, various methods, such as polynomial regression (Nueda et al., 2014), Gaussian process regression (Äijö et al., 2014), autoregressive models (Leng et al., 2015), and natural cubic spline smoothing (Michna et al., 2016), have been developed to address the complexities associated with time-course transcriptome analysis. However, the dynamics of gene expression can differ greatly across species, particularly in microorganisms like fungi, where gene expression changes exhibit non-uniform and nonlinear patterns. This variability limits the effectiveness of existing models in capturing the complex temporal dependencies inherent in dimorphic transition. Therefore, it is essential to develop precise and adaptable methodologies that can effectively analyze time-course transcriptome data and uncover the dynamic gene expression changes occurring during dimorphic transition.

Talaromyces marneffei (formerly *Penicillium marneffei*) is a thermally dimorphic pathogenic fungus that undergoes a switch from a mycelial form at ambient temperature (25°C) to a yeast form at the host temperature (37°C). This temperature-mediated dimorphic transition is intricately linked to the pathogenesis of *T. marneffei* (Boyce and Andrianopoulos, 2015). Once inside the host, *T. marneffei* transforms into yeast cells and adeptly navigates the harsh environment of macrophages by modulating metabolic pathways, adjusting nitrogen source preferences, and expressing peroxidases to counter nutrient deprivation and oxidative stress (Boyce and Andrianopoulos, 2015; Pruksaphon et al., 2022). This enables the fungus to thrive and multiply within macrophages. Subsequently, as macrophages migrate within the host, *T. marneffei* achieves systemic dissemination, leading to disseminated systemic infection. Given its propensity to infect immunocompromised individuals, such as HIV-positive patients, *T. marneffei*'s lower biosafety level and its susceptibility to dimorphic transition induction under laboratory conditions render it an exemplary model organism for investigating fungal dimorphism.

In this study, we employed a DyGAM-NS (dynamic optimized generalized additive model with natural cubic smoothing) approach to analyze the time-course transcriptome data during the dimorphic transition of *T. marneffei*. The DyGAM-NS model combines the flexibility of generalized additive models, which enable capturing non-linear and complex changes, with dynamic optimization using natural cubic spline smoothing. By dynamically optimizing the model parameters for each gene, we identified the best-fitted models and determined the dimorphic transition induced genes (DTIGs) in *T. marneffei*. Cluster analysis of these DTIGs revealed dynamic expression patterns, providing insights into the temporal characteristics of the transcriptome during the dimorphic transition. Furthermore, by integrating the DTIGs with the transcription factors of *T. marneffei*, we constructed a gene expression regulatory network for the dimorphic transition, uncovering potential regulators involved in dimorphic transition of *T. marneffei*. In summary, the DyGAM-NS model not only enables efficient and accurate deciphering of the regulatory mechanisms underlying the dimorphic transition of *T. marneffei*, but also provides a powerful tool for unraveling complex biological processes in other microorganisms.

Materials and methods

Strains media, and culture conditions

Talaromyces marneffei strain PM1, isolated from a talaromycosis patient without HIV infection in Hong Kong (Yang et al., 2014), served as the basis for our research. Previously, we generated *madsA* overexpression and knockout mutant strains (Yang et al., 2014; Wang et al., 2018) derived from the wild-type PM1. All strains were maintained at a temperature of 25°C on Sabouraud Dextrose Agar (SDA) (Becton, Dickinson and Company, USA), which was also utilized for preparing conidial

inocula. Sabouraud Dextrose Broth (SDB) (Becton, Dickinson and Company, USA) was employed for liquid cultures of the wild-type strain. To ensure long-term storage, mycelia from each strain were suspended in 25% (w/v) sterile glycerol and subsequently frozen at −80°C.

Whole-genome sequencing of *T. marneffei* PM1

Initially, a conidia suspension of PM1, containing ∼1 × 10⁶ conidia, was introduced into a 50 ml SDB culture. This culture was incubated at 37°C with shaking at 200 rpm for 7 days. Subsequently, 5 ml of the culture was transferred to a fresh 45 ml SDB and incubated overnight at 37°C. The resulting yeast cells were harvested through centrifugation at 3000 × g and washed twice with 1 × phosphate-buffered saline (PBS). Genomic DNA extraction was carried out using the Epicenter MasterPure Yeast DNA Purification Kit (Cat. MPY80020) according to the manufacturer instructions, with a minor modification: RNA was digested using RNase A, RNase T1, and RNase I instead of solely RNase A. The quantity and quality of the extracted genomic DNA were evaluated using the Qubit 4.0 fluorometer (Thermo Scientific, USA) and NanoDrop spectrophotometer (Thermo Scientific, USA). Whole genome sequencing was performed by utilizing the Illumina HiSeq X platform (350 bp insert size with 150 bp paired-end) and the PacBio Sequel platform (10 kb inserts library).

De novo genome assembly and evaluation of quality

Subreads were obtained from the PacBio sequencing data using the SMRT analysis pipeline v5.1.0 (Jayakumar and Sakakibara, 2019). The Canu v1.6 algorithm (Koren et al., 2017) was utilized to generate a draft genome through an overlapping layout consensus (OLC) assembly approach, employing PacBio subreads. In this process, a genome size of 30 Mb was specified, as previously reported that the genome size of *T. marneffei* is about 30 Mb (Yang et al., 2014), and the longest 40× of subreads were selected as seed reads to generate error-corrected long reads. The draft assembly underwent further processing, including the filtration of bubble sequences and the identification of the mitochondria sequence by comparing it to the known *T. marneffei* sequence using BLASTN v2.12.0+ (Camacho et al., 2009). To enhance the accuracy of the assembly, Arrow v2.0.2 (Jayakumar and Sakakibara, 2019) was employed based on PacBio reads, while Pilon v1.24 (Walker et al., 2014) was used based on Illumina reads for additional polishing. The completeness of the assembly was assessed by performing BUSCO v4 analysis (Manni et al., 2021) against the Eurotiales database.

Temporal RNA preparation and sequencing

In the initial phase of the experiment, conidia of strain PM1 were inoculated onto SDA plates and cultivated at two

different temperatures, 25°C and 37°C, for a duration of 1 week. Subsequently, the SDA plates incubated at 25°C were transferred to 37°C, while those cultivated at 37°C were transferred to 25°C. This process facilitated the growth of colonies representing the M-to-Y or Y-to-M transition growth forms, respectively. Total RNA samples were extracted at specific time points after the temperature switch, namely at 0, 1, 3, 6, 9, 12, 24, 36, 48, and 72 hours. In the case of the *madsA* overexpression and knockout strains, total RNA samples were extracted at 0 and 48 h for the M-to-Y transition, and at 0 and 6 h for the Y-to-M transition. For each time point, three independent biological replicates were used to extract total RNA using the E.Z.N.A. fungal RNA kit (Omega Bio-Tek). The quantification of RNA was performed using a Qubit 4.0 fluorometer (Thermo Scientific, USA). A total of 84 strand-specific libraries, with RNA integrity numbers (RIN) exceeding 6.5, were constructed and sequenced on an Illumina HiSeq X platform, resulting in ∼20 million 150-bp paired-end reads for each sample. Prior to analysis, the reads were trimmed using Trimmomatic v0.38 (Bolger et al., 2014) based on quality assessment carried out by FastQC v0.11.5 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/).

Annotation of repeat sequences and protein-coding genes

Repeat regions in the assembly were masked using RepeatMaker v4.0.7 (http://repeatmasker.org) with a *T. marneffei*-specific repeat library generated by RepeatModeler v4.0.7 (http://repeatmasker.org/RepeatModeler/). Based on the annotation results obtained from RepeatMasker, we identified the presence of tandem repeat sequences (with a repeat fragment length of 6 bp) at the ends of each genomic sequence to ascertain telomeric sequences in *T. marneffei*. To locate centromeres, we scanned the genome sequences using 50 kb windows, and identified regions characterized by low GC content and high transposon proportion as potential centromeric regions. Gene annotations were conducted using three distinct types of evidence. Firstly, the assembly was aligned to a protein database that combined the UniProt/Swiss-Prot protein database and all sequences of *Talaromyces marneffei* from the NCBI protein database, using Exonrate v2.2.0 (Slater and Birney, 2005). Additionally, temporal RNA-seq reads of *T. marneffei* PM1 from various time points during the dimorphic transition were *de novo* assembled with Trinity v2.3.2 (Grabherr et al., 2011). The resulting Trinity assembly was then passed through the PASA v2.5.2 pipeline (Haas, 2003) to generate transcript evidence. Furthermore, *ab initio* gene predictions were carried out using FGENESH (Solovyev et al., 2006) and BRAKER2 v2.1.6 (Brůna et al., 2021). For FGENESH, the genome-specific parameters of *Penicillium* were employed, while BRAKER2 utilized RNA-seq reads to enhance the accuracy of gene predictions. The EVidenceModeler (EVM) v1.1.1 (Haas et al., 2008) was employed to integrate the three types of evidence into consensus gene structures, assigning a weight of 10 to transcript alignments, 5 to protein alignments, and 1 to *ab initio* gene predictions. The final gene sets were subjected to a BLASTP v2.12.0+ (Camacho et al.,

2009) search against the NCBI non-redundant (NR) protein database. The domains of these genes were annotated using InterProScan v5.56-89.0 (Jones et al., 2014), which utilizes publicly available databases. Functional annotations were then performed using BLAST2GO v5.2.5 (Conesa and Götz, 2008), incorporating the results obtained from both BLASTP and InterProScan analyses.

RNA-seq data analysis

RNA-seq short reads were aligned to the annotated genomes using STAR v2.7.9a (Dobin et al., 2013) to obtain read mappings. The read count for each gene was calculated using FeatureCount v2.0.3 (Liao et al., 2014), specifying a library by using "-s 2". The expression level of genes was quantified using TPM (transcripts per million). Genes with a mean TPM >1 at least once during the dimorphic transition were selected for subsequent analysis. The raw counts were then normalized using the median of ratios method implemented in DESeq2 (Love et al., 2014). For the time-course RNA-seq data, principal component analysis (PCA) was performed using the prcomp function in R v4.2.2. Visualization of the PCA results was achieved using the ggplot2 package. The heatmap depicting gene expression levels during the dimorphic transition was generated using the pheatmap package. In the case of RNA-seq data from *madsA* overexpression and knockout strains, differentially expressed genes were identified using the DESeq2 method, applying a false discovery rate (FDR) threshold of <0.05 and a fold change threshold >2.

Identification of dimorphic transition induced genes

We utilized a dynamic optimized generalized additive model with natural cubic smoothing (DyGAM-NS) approach to capture the non-linear relationship between gene expression profiles and time during the dimorphic transition (Equation 1).

$$Y_{gene_i} \sim GAM(t) + \varepsilon_{gene_i} \tag{1}$$

To account for the discrete nature and overdispersion of read counts, we employed a negative binomial (NB) distribution with gene-specific means and dispersion, estimated using DESeq2 (Equation 2).

$$\begin{cases} Y_{gene_i} \sim NB\left(\mu_{gene_i}, \theta_{gene_i}\right) \\ \mu_{gene_i} = E\left(Y_{gene_i}\right) \end{cases} \tag{2}$$

The flexibility of the DyGAM-NS method allowed us to adjust the number and placement of knots (Equation 3). To determine the optimal parameters for each gene, we evaluated models with all possible combinations of knots using the corrected Akaike information criterion (AICc).

$$GAM(t) = \begin{cases} GAM_1(t),\ 0 \leq t < t_1 \\ GAM_2(t - t_1),\ t_1 \leq t < t_2 \\ \cdots \\ GAM_n(t - t_{n-1}),\ t_{n-1} \leq t \leq t_n \end{cases} \tag{3}$$

The DyGAM-NS model was fitted using the gam function in the mgcv package (Wood, 2011), and AICc values were calculated using the AICc function in the AICcmodavg package. Finally, we filtered the dimorphic transition induced genes (DTIGs) based on criteria of FDR <0.05, deviance explained >0.7. The code for fitting the DyGAM-NS model can be downloaded from GitHub (https://github.com/yangence/DyGAM-NS).

Predicting gene expression levels during dimorphic transitions using DyGAM-NS model

Initially, we extracted expression data of all DTIGs identified during the M-to-Y and Y-to-M transitions. For each DTIG, a set of data points at an intermediate time point (1, 3, 6, 9, 12, 24, 36, 48 h) was randomly removed. Subsequently, leveraging data from the remaining 9 time points, we fitted the DyGAM-NS model for the responding gene. Then, the removed time point was provided to the DyGAM-NS model to yield its predicted gene expression level at that time point. Employing the model's estimated standard error, we derived the 95% confidence interval of the predicted expression level. Ultimately, by contrasting the original gene expression level with the confidence interval, we evaluated the predictive performance of the DyGAM-NS model.

Clustering gene expression patterns of DTIGs

In order to select more credible DTIGs, we filtered genes with expression changes exceeding a 2-fold threshold during the dimorphic transitions, designating them as highly variable DTIGs for subsequent clustering analysis. To analyze the expression pattern of DTIGs, we employed the trimmed *k*-means method from the tclust package (Fritz et al., 2012). This method utilized the fitted values of the DyGAM-NS model for each gene, where the fitted values were standardized to have a zero mean and unit variance across samples for each gene. The optimal number of clusters was determined based on the within sum of squares. The R package clusterProfiler v4.6.0 was employed, based on functional annotations specific to *T. marneffei*, for Gene Ontology (GO) functional enrichment analysis of genes within diverse expression patterns.

Construction of gene regulatory networks

Gene regulatory networks were constructed using GRNboost2 v0.1.6 (Moerman et al., 2019) based on the time-course expression data sets. The list of transcription factors (TFs) used as input for GRNboost2 was identified through the InterPro database (IPR) and compiled from previous studies (Shelest, 2008). The final gene regulatory network was obtained by retaining edges with the top 5% cumulative weight. The betweenness centrality of each gene was calculated using the igraph package in R. Visualization of the gene regulatory network was performed using the NetworkD3 package in R. The motifs associated with each regulatory gene in the network were identified using MEME (Bailey et al., 2009), utilizing the upstream 500-bp regions of the corresponding target genes.

Data availability

The PacBio and Illumina sequencing data used for genome assembly have been deposited in the Genome Sequence Archive under the accession PRJCA002718. The reference genome sequences of *T. marneffei* PM1 have been deposited at DDBJ/ENA/GenBank under the accession JPOX02000000. The sequencing dataset for time-course RNA-seq of *T. marneffei* PM1 during M-to-Y and Y-to-M transitions have been deposited in the Sequence Read Archive under the accession PRJNA970557. The RNA-seq dataset used for identification of high variable genes during dimorphic transitions are available for download from the SRA under the accession PRJNA431116.

Results

Improving the reference genome of *Talaromyces marneffei* strain PM1 to near-chromosome level

To improve the quality of the reference genome for *Talaromyces marneffei* strain PM1 (Yang et al., 2014), we conducted high-coverage whole-genome sequencing using a combination of PacBio SMRT (3.3 Gb, 115×) and Illumina (6.0 Gb, 210×) sequencing technologies. Firstly, we employed PacBio data to construct the initial draft assembly, resulting in 26 contigs. Subsequently, we identified and eliminated 12 bubble sequences and one mitochondria sequence from the draft assembly. Finally, we polished the genome sequences using both PacBio raw data and high-coverage Illumina data. The final assembly consisted of 13 contiguous sequences, with a total size of 29.0 Mb, an N50 of 3.3 Mb, and an L95 of 8 (Table 1). Based on the Eurotiales database ($n = 4{,}191$), we successfully identified 4,067 complete BUSCOs (97.3%), indicating a significant improvement in the completeness and continuity of the PM1 genome sequences. Furthermore, we detected 16 telomere sequences (TCCTAA) and 8 centromere regions in the final assembly (Figure 1), consistent with previous findings (Gifford and Cooper, 2009; Cuomo et al., 2020), thereby suggesting that the genome sequences have achieved near-chromosomal level.
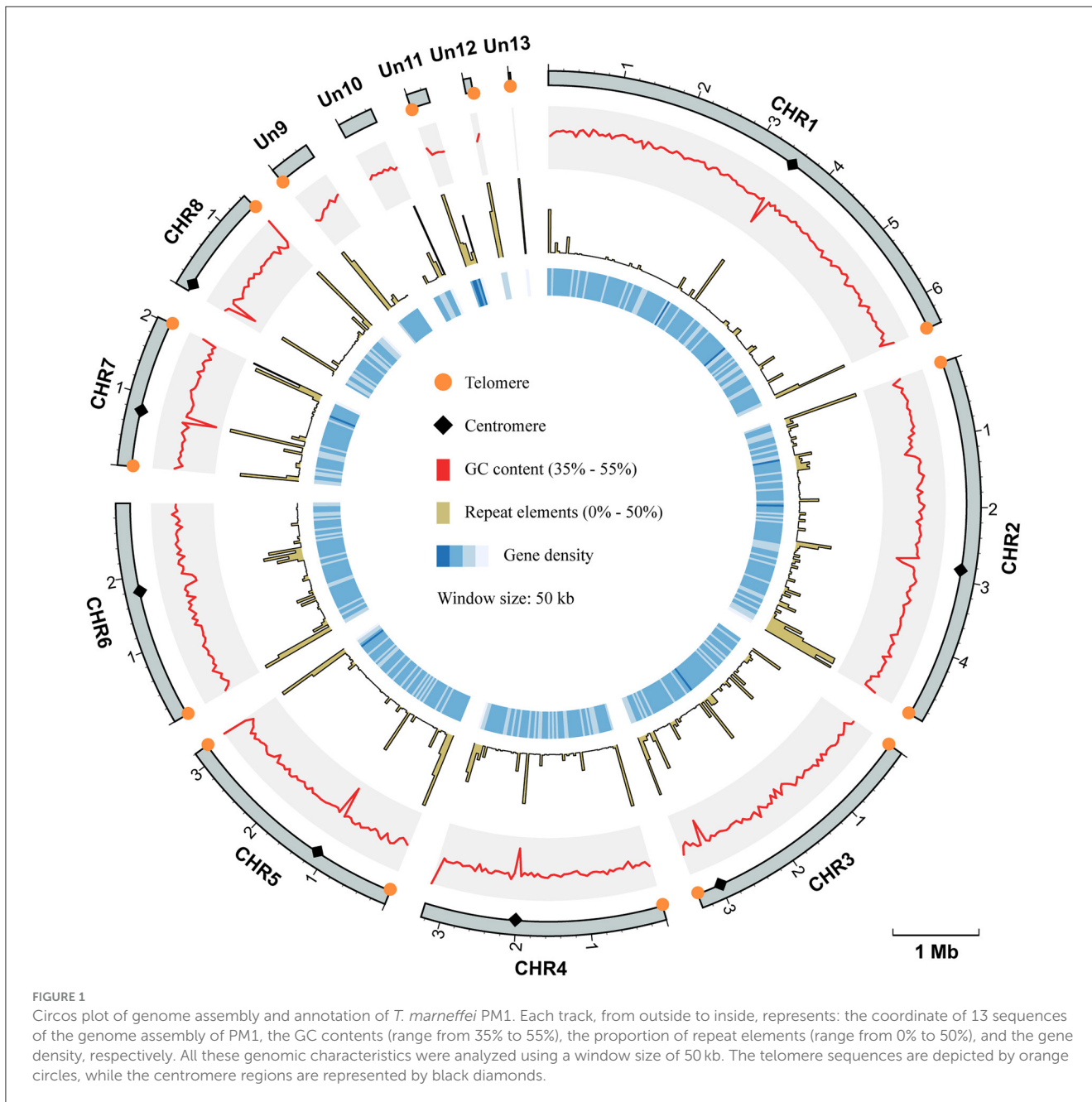
TABLE 1 Summary of *T. marneffei* PM1 genome assembly and annotation.

	T. marneffei PM1	Previously reported PM1
Genome assembly		
Genome size (bp)	29,019,915	28,345,621
Contigs	13	216
N50 (bp)	3,334,497	678,243
L95	8	112
GC content (%)	46.7	46.7
Complete BUSCOs	4,076 (97.3%)	3,778 (90.1%)
Single-copy	4,067	2,973
Duplicated	9	805
Fragmented BUSCOs	10 (0.2%)	56
Missing BUSCOs	105 (2.5%)	357
Genome annotation		
Protein-coding genes	10,378	9,480
No. of transcripts	16,683	13,042
Mean gene length (bp)	2,400	1,816
Mean no. of exons	3.6	3.6
Gene density (Mb^{-1})	357.9	334.4
Genes with IPR domain	8,053 (77.6%)	/
Genes with GO term	5,822 (56.1%)	/

Illustrating transcriptional dynamics of dimorphic transition by using time-course RNA-seq data
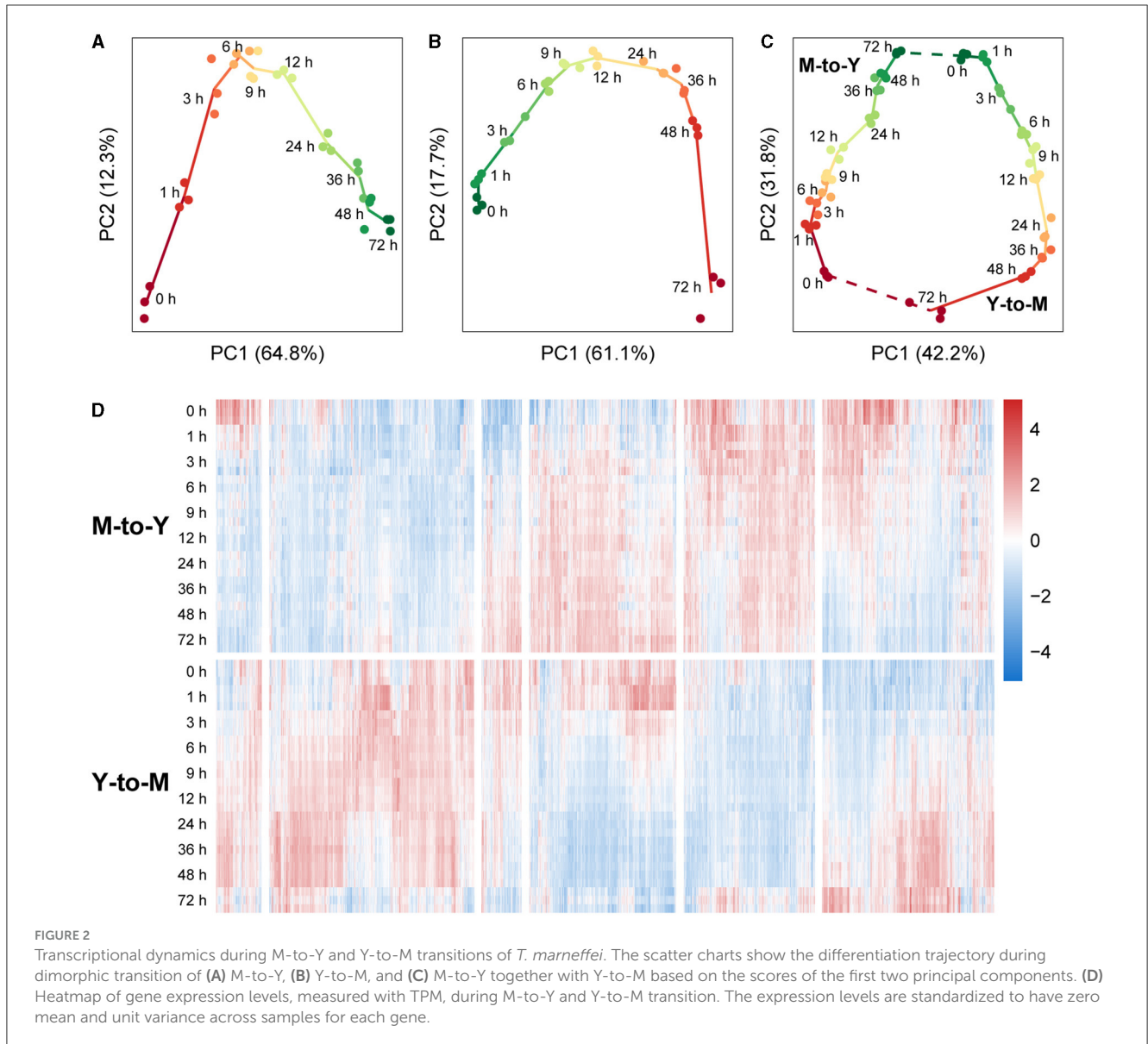
To explore the dynamic gene activity during the dimorphic transition of *T. marneffei*, we conducted RNA-seq analysis on samples collected at various stages of the time-course, encompassing both the mycelium-to-yeast (M-to-Y) and yeast-to-mycelium (Y-to-M) transitions. Recognizing the non-uniform nature of the dimorphic transition, we strategically selected 10 time points distributed unevenly over the 0–72 h period (0, 1, 3, 6, 9, 12, 24, 36, 48, 72 h) for both transitions. With three biological replicates for each time point, a total of 60 samples were subjected to RNA-seq, yielding an average of 24 million read pairs per sample.

Initially, these transcriptome data were employed to facilitate the genome annotation of PM1, resulting in the identification of 10,378 protein-coding genes (Table 1). Among these genes, 77.6% of the proteins were annotated with at least one InterPro domain, and 56.1% were assigned at least one Gene Ontology (GO) term. Subsequently, the RNA-seq data were mapped to the PM1 reference genome, and uniquely mapped read pairs were utilized to estimate the expression levels of protein-coding genes using the transcripts per million (TPM) method. To mitigate the impact of transcriptional noise arising from low-expression genes, we filtered out 9,541 genes with a mean TPM <1 in at least one time point during either the M-to-Y or Y-to-M transition.

FIGURE 1
Circos plot of genome assembly and annotation of *T. marneffei* PM1. Each track, from outside to inside, represents: the coordinate of 13 sequences of the genome assembly of PM1, the GC contents (range from 35% to 55%), the proportion of repeat elements (range from 0% to 50%), and the gene density, respectively. All these genomic characteristics were analyzed using a window size of 50 kb. The telomere sequences are depicted by orange circles, while the centromere regions are represented by black diamonds.

An initial analysis of the transcriptomic dynamics during the dimorphic transition was performed using principal component analysis (PCA). In the M-to-Y transition, PC1 accounted for 64.8% of the variance between samples. Arranging the samples in the order of sampling time revealed a distinct trajectory on the PC1-PC2 plane, indicating clear differentiation among the sample groups (Figure 2A). Notably, during the early stage of the M-to-Y transition (0–3 h), the distances between sample groups were notably larger than those at other time points, suggesting a rapid transcriptional change during the initial phase of the M-to-Y transition. Similarly, in the Y-to-M transition, PC1 explained 61.1% of the variance between samples. A discernible trajectory was observed on the PC1-PC2 plane among the sample groups in the order of transition time (Figure 2B). The inter-group distances

indicated that the magnitude of transcriptional changes during the early stage of the Y-to-M transition was not as pronounced as in the M-to-Y transition. Importantly, when all samples were subjected to PCA, a circular differentiation trajectory was formed between sample groups, aligning with the reversible cyclic nature of the dimorphic transition (Figure 2C). On the PC1-PC2 plane in the PCA of all samples, PC1 effectively distinguished between the M-to-Y and Y-to-M transitions, indicating distinct transcriptional regulation patterns during the dimorphic transition. Heatmap analysis based on gene expression levels across all samples further demonstrated that a considerable proportion of genes exhibited differential expression patterns during the dimorphic transition (Figure 2D), highlighting the complexity and nonlinearity of transcriptional dynamics during this process.

FIGURE 2
Transcriptional dynamics during M-to-Y and Y-to-M transitions of *T. marneffei*. The scatter charts show the differentiation trajectory during dimorphic transition of **(A)** M-to-Y, **(B)** Y-to-M, and **(C)** M-to-Y together with Y-to-M based on the scores of the first two principal components. **(D)** Heatmap of gene expression levels, measured with TPM, during M-to-Y and Y-to-M transition. The expression levels are standardized to have zero mean and unit variance across samples for each gene.

Identification of dimorphic transition induced genes by using generalized additive model and natural cubic spline

In this investigation, we employed an unevenly spaced sampling strategy that better aligns with the biological changes observed during the dimorphic transition of *T. marneffei*. To effectively capture the non-uniform and non-linear dynamics of the time-course data, we proposed a dynamic fitting method, referred to as dynamic optimized generalized additive model with natural cubic smoothing (DyGAM-NS), to examine the dynamic alterations in gene expression induced by the dimorphic transition. By incorporating natural cubic splines as smoothing terms, DyGAM-NS allows for non-uniform and customizable partition knots, thereby accommodating the changing characteristics of gene expression during the dimorphic transition. The read counts for each gene during transition was modeled using

a negative binomial distribution with gene-specific dispersion parameters. To explore the optimal parameters for the DyGAM-NS model, we randomly selected 100 genes without replacement and fitted their expression changes over time using various NS partition numbers, NS knots, and dispersion coefficients. Model performance was evaluated using the corrected Akaike information criterion (AICc). The results demonstrated that gene-specific optimization of NS partition numbers, NS knots, and dispersion coefficients enhanced the performance of the DyGAM-NS model (Supplementary Figure S1). Consequently, for each gene, we exhaustively explored combinations of NS smoothing term partition numbers and knots and selected the optimal parameter combination based on AICc. Compared to the default parameters, most genes (92.3% in M-to-Y and 96.4% in Y-to-M) exhibited improved model performance after parameter optimization (Supplementary Figures S2A, B). Summarizing the optimal model parameters for each gene revealed that most of genes had an

optimal partition number of 5 (Supplementary Figures S2C, D), indicating the complex and diverse changes in the transcriptome during the dimorphic transition. In the M-to-Y transition, there was a noticeable presence of genes with partition knots at 1 and 3 h compared to the Y-to-M transition (Supplementary Figure S2E, F), further suggesting the more rapid early transcriptome changes in the M-to-Y transition.

Using DyGAM-NS model, we identified 6,167 and 8,085 DTIGs in M-to-Y and Y-to-M transitions, respectively. Due to the lack of known data on true DTIGs of *T. marneffei*, direct evaluation of the precision and recall of the DyGAM-NS model is challenging. Thus, we alternatively estimated precision and recall using two different approaches. Given that genes involved in regulating the dimorphic transition should display significant expression changes during the process, DTIGs that do not exhibit differential expression between any two time points in the time-course data are likely false positives. Applying this criterion, we estimated the precision of M-to-Y DTIGs to be 96.1% (5,926/6,167) and Y-to-M DTIGs to be 96.1% (7,770/8085) (Table 2). To assess the recall of the model, we identified highly variable genes (HVGs) based on the biological coefficient of variation (BCV) during the dimorphic transition. Genes involved in regulating the dimorphic transition are expected to have higher BCV in temporal data compared to a static state. We employed a chi-square test to identify genes with significantly higher BCV (Brennecke et al., 2013) (FDR < 0.05) during the dimorphic transition than in a static state, considering them as HVGs. Using the HVG dataset as the criterion to evaluate the DyGAM-NS model, we estimated the recall of M-to-Y to be 98.1% (152/155) and Y-to-M to be 98.2% (280/285) (Table 2). Finally, based on the estimated precision and recall, the estimated F_1-scores of the DyGAM-NS model in the M-to-Y and Y-to-M transitions were both 97.3% (Table 2).

To comprehensively assess the effectiveness of the DyGAM-NS model in identifying DTIGs, we conducted a comparison with seven commonly employed time-course analysis methods (McCarthy et al., 2012; Law et al., 2014; Nueda et al., 2014; Leng et al., 2015; Fischer et al., 2018) (Table 2). Comparison results revealed that the DyGAM-NS model demonstrates performance on par with other methods in identifying DTIGs. The estimated F_1-score of DyGAM-NS slightly surpasses that of the other seven methods. Furthermore, comparing the DTIGs identified by DyGAM-NS with those of the other seven methods, DTIGs identified by the DyGAM-NS model can be validated by at least two other methods in both M-to-Y and Y-to-M transitions, suggesting the robustness of our approach.

Of particular interest is the DyGAM-NS model's potential in predicting gene expression levels of DTIGs at various time points during dimorphic transition. To validate this, for all identified DTIGs genes, we randomly excluded samples from any one set of time points during the dimorphic transition. We then utilized the expression data from the remaining 9 time points to fit the DyGAM-NS model. Next, we used the newly fitted model to predict the gene expression levels for the randomly excluded time points and compared them with the original data to assess their accuracy. The results indicate that during the M-to-Y transition process, 51.5% (3174/6167) of DTIGs' actual expression levels fall within the 95% confidence interval of the predicted values.

Similarly, during the Y-to-M transition process, 51.9% (4194/8085) of DTIGs' actual expression levels fall within the 95% confidence interval of the predicted values. In comparison to other methods, DyGAM-NS not only exhibits comparable DTIG identification capabilities but also demonstrates potential in predicting temporal gene expression levels.

Clustering of DTIGs indicate that different gene expression pattern for M-to-Y and Y-to-M transitions of *T. marneffei*

To explore the dynamic changes of DTIGs during the dimorphic transition of *T. marneffei*, we further applied stringent filtering criteria (fold change of expression levels during dimorphic transitions > 2), resulting in the identification of 3,084 and 4,156 highly variable DTIGs in M-to-Y and Y-to-M transitions, respectively. By using a modified trimmed *k*-means clustering algorithm (Fritz et al., 2012), we discovered 11 distinct expression patterns in both M-to-Y and Y-to-M transitions. In the M-to-Y transition, we discovered 11 distinct expression patterns (Figure 3), encompassing three primary types of changes: 4 up-regulated patterns, 5 down-regulated patterns, and 2 pulse-like patterns. Among the up-regulated expression patterns, we observed critical genes involved in the survival of *T. marneffei* within host macrophages, such as iron affinity-related genes (Pasricha et al., 2016) (*ftrA*, *ftrC*, and *sidA*) in pattern M-to-Y-up-1, *pepA* (Payne et al., 2019) in pattern M-to-Y-up-3, and cytochrome P450 monoxygenases (Boyce et al., 2018) (*simA*) and the tyrosine metabolic regulator (Boyce et al., 2015) (*hmgR*) in pattern M-to-Y-up-4. These findings suggest significant transcriptional changes occurring during the early stages of the M-to-Y transition, facilitating rapid adaptation to the harsh intracellular environment. Concerning the down-regulated expression patterns, we identified genes associated with conidiation and pigment synthesis (Borneman et al., 2002a; Todd et al., 2003; Woo et al., 2010, 2012; Sapmak et al., 2015) (*alb1*, *pbrB*, *pks11*, *pks12*, *brlA*, and *stuA*), which are closely linked to the morphological characteristics of the mycelia. Additionally, the pulse-like expression patterns exhibited an immediate transient decrease in gene expression at the early stage of the M-to-Y transition, potentially indicating a transient response regulation mechanism during this transition. However, the underlying mechanism governing these genes remains unknown and warrants further investigation.

DTIGs associated with the Y-to-M transition can be classified into 11 distinct expression patterns (Figure 4), which can further be grouped into three types of changes: 5 up-regulated patterns, 2 down-regulated patterns, and 4 pulse-like patterns. In the up-regulated expression pattern Y-to-M-up-1, we identified two basic leucine-zipper (bZIP) transcription factors (Nimmanee et al., 2014; Dankai et al., 2016) (*atfA* and *yapA*), known to regulate the response of *T. marneffei* to environmental stress. Another up-regulated pattern (Y-to-M-up-3) contains the C2H2 transcription factor (Bugeja et al., 2013) (*hgrA*), which governs the formation of the hyphal cell wall. Similarly, up-regulated patterns Y-to-M-up-2
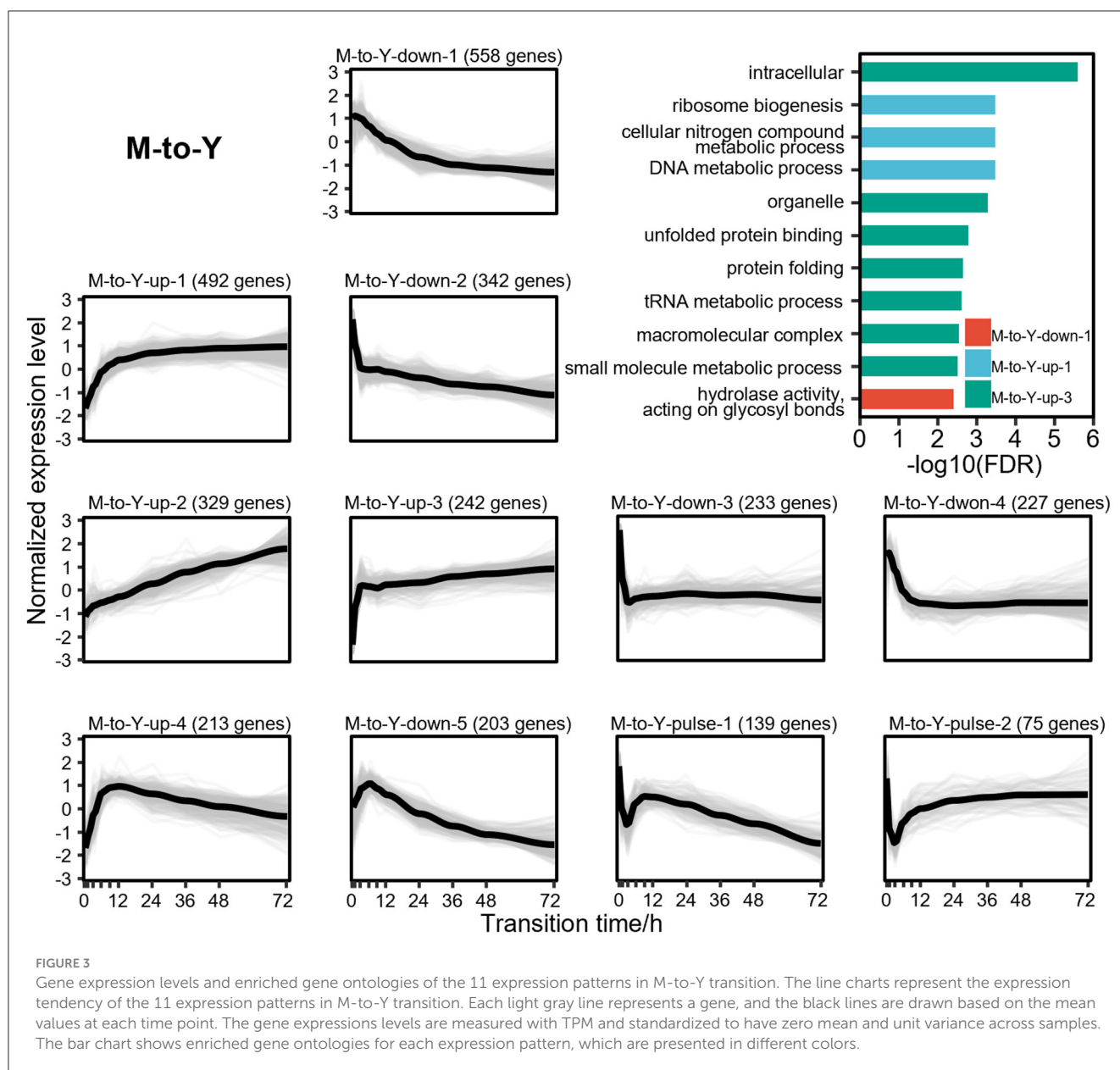
TABLE 2 Performance of DyGAM-NS and other commonly used models in identification of DTIGs.

Model	Tools	DTIGs	Precision	Recall	F_1-score
M-to-Y					
DyGAM-NS	In-house	6,167	96.1%	98.1%	97.1%
AR-HMM	EBSeqHMM	6,134	98.7%	88.3%	93.2%
Polynomial regression	Next maSigPro	2,855	98.4%	79.6%	88.0%
GLM	edgeR	7,527	93.2%	97.1%	95.1%
Linear	limma	9,129	90.4%	100.0%	95.0%
Impulse	ImpulseDE2	9,488	87.5%	100.0%	93.3%
Cubic splines	edgeR	8,279	91.8%	100.0%	95.7%
Autocorrelation	AutoCor	3,513	94.6%	58.4%	72.2%
Y-to-M					
DyGAM-NS	In-house	8,085	96.1%	98.2%	97.1%
AR-HMM	EBSeqHMM	7,356	97.5%	82.0%	89.1%
Polynomial regression	Next maSigPro	5,610	98.5%	95.2%	96.8%
GLM	edgeR	7,803	94.6%	96.7%	95.6%
Linear	limma	9,486	92.5%	96.3%	94.4%
Impulse	ImpulseDE2	9,698	91.1%	99.6%	95.2%
Cubic splines	edgeR	9,302	93.2%	99.3%	96.2%
Autocorrelation	AutoCor	3,769	98.4%	62.9%	76.7%

(Borneman et al., 2002a; Zuber et al., 2003; Woo et al., 2012) (*gasC*, *pks11*, *pks12*, and *stuA*) and Y-to-M-up-4 (Todd et al., 2003; Woo et al., 2010; Sapmak et al., 2015) (*pbrB*, *alb1*, and *brlA*) exhibit genes related to the morphological characteristics of the mycelia, including conidiation and pigment synthesis. Notably, the dynamic changes in the transcriptome during the Y-to-M transition are not as pronounced as those observed in the M-to-Y transition. For instance, the mycelia phase-related genes found in the pattern Y-to-M-up-4 gradually upregulate in the later stage of the Y-to-M transition (Figure 4). In terms of the down-regulated expression patterns, functional enrichment analysis reveals a significant enrichment of translation-related functions, indicating a decreased demand for protein synthesis and related processes during the Y-to-M transition. Moreover, the continuously down-regulated pattern Y-to-M-down-2 exhibits a significant enrichment of mitochondrial-related components and genes associated with the tricarboxylic acid (TCA) cycle. This finding is consistent with the fact that *T. marneffei* relies primarily on the TCA cycle for energy metabolism in the yeast phase, whereas glycolysis is prominent in the mycelia phase (Pasricha et al., 2017). Additionally, we observed genes related to intracellular survival (Pasricha et al., 2016; Payne et al., 2019) (*sidA*, *pepA*, and *ftrC*) in the two down-regulated expression patterns. Within the pulse-like expression patterns, Y-to-M-pulse-1 includes the RFX transcription factor (Bugeja et al., 2010) *rfxA*, which plays a role in regulating cell division. Functional enrichment analysis also indicates a significant enrichment of cell division and cycle-related functions in pattern Y-to-M-pulse-1, suggesting that cell division is inhibited in the early stage of the Y-to-M transition and reactivated in the later stage.

Gene regulation network underlying potential regulators of dimorphic transition of *T. marneffei*

To elucidate the regulatory relationships among DTIGs during the dimorphic transition, we employed the GRNboost2 (Moerman et al., 2019) method to construct a directed gene regulatory network. To enhance the accuracy of the GRNboost2 approach, we utilized 5,223 DTIGs identified by the DyGAM-NS model in both the M-to-Y and Y-to-M transitions as target genes. Additionally, we incorporated 401 candidate transcription factors as prior information, identified through a combination of functional annotation of *T. marneffei* and known fungal transcription factor domains (Shelest, 2008). To ensure the focus on robust regulatory relationships, we retained the top 5% cumulative weight of strong gene-gene regulatory interactions to form the final gene regulatory network. Following the filtering process, the dimorphic transition gene regulatory network comprised 2,243 pairs of robust regulatory relationships involving 1,664 genes, of which 325 were transcription factors (Figure 5A). Notably, within these 1,664 genes, TM060272, TM021494, and TM030196 emerged as the most significant genes based on their betweenness centrality within the network (Figure 5B). TM060272, possessing a Zn(2)-C6 fungal-type DNA-binding domain (IPR001138), displayed up-regulation during the M-to-Y transition but exhibited minimal changes during the Y-to-M transition, indicating its role in M-to-Y transition regulation. Conversely, TM021494, which also contains a zinc finger domain, exhibited up-regulation during the Y-to-M transition while showing little change during the M-to-Y

FIGURE 3
Gene expression levels and enriched gene ontologies of the 11 expression patterns in M–to–Y transition. The line charts represent the expression tendency of the 11 expression patterns in M–to–Y transition. Each light gray line represents a gene, and the black lines are drawn based on the mean values at each time point. The gene expressions levels are measured with TPM and standardized to have zero mean and unit variance across samples. The bar chart shows enriched gene ontologies for each expression pattern, which are presented in different colors.

transition, suggesting its involvement in the regulation of the Y-to-M transition. The TM030196 gene (*abaA*) exhibited an immediate transient increase in expression during the early stage of the M-to-Y transition (0–9 h) but showed minimal changes during the Y-to-M transition, supporting its regulatory role in the M-to-Y transition, consistent with previous research (Borneman et al., 2002b).

To assess the precision of the gene regulatory network governing the dimorphic transition, we focused on the MADS-box transcription factor *madsA* gene (TM030606), which has been experimentally verified to play a role in regulating the dimorphic transition (Yang et al., 2014; Wang et al., 2018) (Figure 5B). To validate the downstream regulatory genes within the dimorphic transition regulatory network, we employed knockout and overexpression strains of *madsA*. RNA-seq analysis was conducted on these strains at four time points during the dimorphic transition (0 and 48 h for M-to-Y, and 0 and 6 h for

Y-to-M). By integrating the transcriptome data from the wild-type strain at corresponding time points, we identified 1,788 genes regulated by *madsA* during the dimorphic transition, of which 1,066 were identified as DTIGs by DyGAM-NS. Notably, in the gene regulatory network constructed using GRNboost2, over 50% (7/12) of the downstream target genes of *madsA* were validated (Supplementary Figure S4).

Discussion

Dimorphic transition represents a sophisticated adaptive mechanism in thermally dimorphic fungi, enabling their survival and propagation across diverse environments. Understanding the regulatory mechanisms underlying fungal dimorphic transition is a key scientific pursuit, given its close connection to fungal pathogenicity. Our study focused on the model organism of
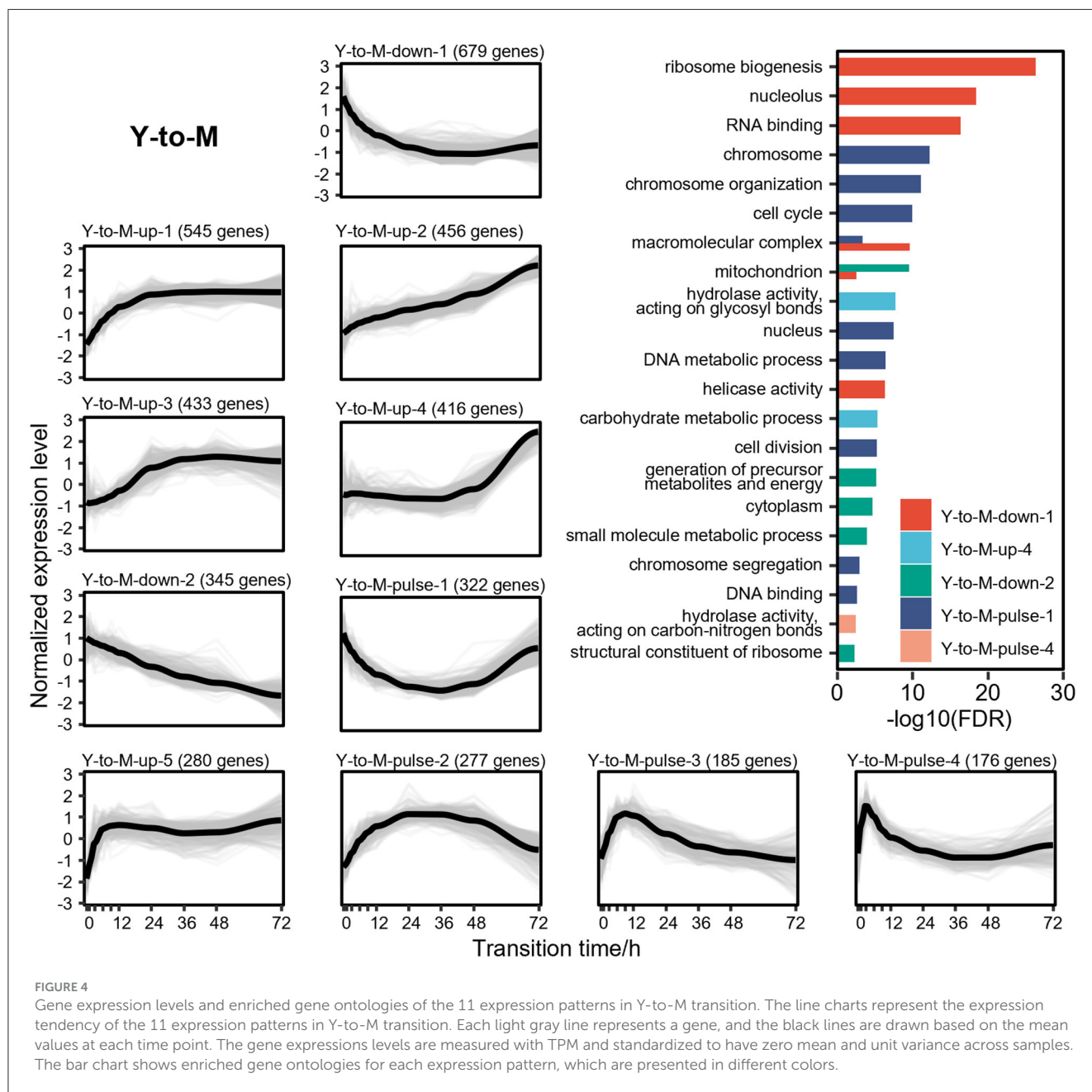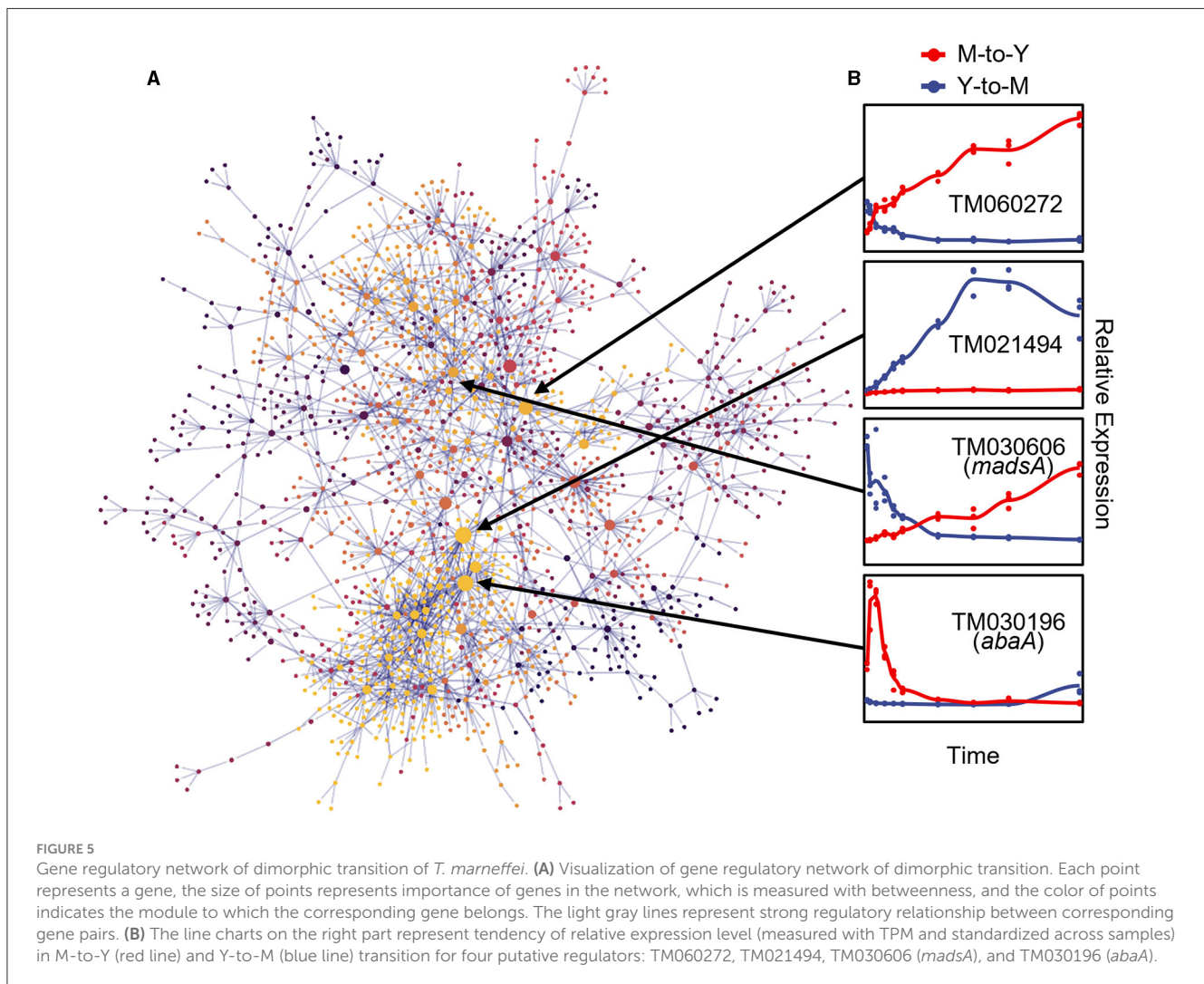
FIGURE 4
Gene expression levels and enriched gene ontologies of the 11 expression patterns in Y-to-M transition. The line charts represent the expression tendency of the 11 expression patterns in Y-to-M transition. Each light gray line represents a gene, and the black lines are drawn based on the mean values at each time point. The gene expressions levels are measured with TPM and standardized to have zero mean and unit variance across samples. The bar chart shows enriched gene ontologies for each expression pattern, which are presented in different colors.

dimorphic fungi, *Talaromyces marneffei*, and demonstrated the significant potential of high-resolution temporal transcriptomes in deciphering the regulatory mechanisms underlying fungal dimorphic transition. On the one hand, temporal transcriptomes provide comprehensive coverage of the entire biological process, surpassing the limitations of static transcriptomic studies. In previous studies, comparisons of transcriptional profiles at different temperatures yielded limited information, with only a fraction of genes associated with dimorphic transition identified. In contrast, our study utilized temporal transcriptional data spanning the first 72 h of M-to-Y and Y-to-M transitions, resulting in the identification of a larger number of dimorphic transition induced genes (DTIGs). Specifically, we identified 3,084 DTIGs in the M-to-Y transition and 4,156 DTIGs in the Y-to-M transition, accounting for 50.3% (5,223 out of 10,378) of all protein-coding genes. This substantial increase in DTIG identification highlights the

importance of high-resolution temporal data in comprehensively identifying regulatory genes related to dimorphic transition. On the other hand, temporal data also enables the observation of gene expression changes throughout the biological process, facilitating the discovery of distinct expression patterns during dimorphic transition. Our differentiation trajectory analysis revealed divergent gene expression patterns between the M-to-Y and Y-to-M transitions. Moreover, we identified numerous transition-specific DTIGs (1,067 for M-to-Y and 2,139 for Y-to-M), further supporting the asymmetrical nature of the two transition directions. Furthermore, the clustering of DTIGs in both the transition from M-to-Y and Y-to-M reveals the presence of numerous distinctive expression patterns throughout the dimorphic transition, including rapid up or down-regulation, continuous up or down-regulation, and impulse-like. Through functional analysis, it becomes evident that diverse gene functions

FIGURE 5
Gene regulatory network of dimorphic transition of *T. marneffei*. **(A)** Visualization of gene regulatory network of dimorphic transition. Each point represents a gene, the size of points represents importance of genes in the network, which is measured with betweenness, and the color of points indicates the module to which the corresponding gene belongs. The light gray lines represent strong regulatory relationship between corresponding gene pairs. **(B)** The line charts on the right part represent tendency of relative expression level (measured with TPM and standardized across samples) in M-to-Y (red line) and Y-to-M (blue line) transition for four putative regulators: TM060272, TM021494, TM030606 (*madsA*), and TM030196 (*abaA*).

are significantly enriched within these distinct expression patterns, suggesting that different biological functions play prominent roles at different stages of dimorphic transition. Consequently, our study underscores the criticality of temporal transcriptomic data as an indispensable tool for dissecting the intricate nature of fungal dimorphic transitions.

Despite the decreasing cost of sequencing technologies, enabling high-resolution time-course transcriptome sequencing, challenges persist in deciphering temporal data for complex biological processes, particularly in microorganisms like fungi. In this study, we developed the DyGAM-NS model, which integrates a flexible generalized additive model with dynamic gene-specific natural cubic spline fitting optimization. We benchmarked the DyGAM-NS model against seven other commonly used time-course analysis methods. To assess its performance, we used differentially expressed gene sets identified through pairwise comparisons at each time point and highly variable gene sets based on biological coefficient of variation, given the absence of a standardized dataset for DTIGs in *T. marneffei*. Through this hybrid estimation approach, the DyGAM-NS model exhibited slightly higher F_1-scores for DTIG identification in both M-to-Y and Y-to-M transitions compared to the other seven methods. Additionally, our analysis involved gene-specific

parameter optimization for each gene, yielding a unique DyGAM-NS model for every DTIG. Leveraging the fitted model, our method can predict gene expression levels at other time points during dimorphic transitions, a capability currently lacking in other methods. Cross-validation using random sampling demonstrated that over 50% of gene expression levels fell within the 95% confidence interval predicted by the DyGAM-NS model in both M-to-Y and Y-to-M transitions. This predictive potential for temporal gene expression levels also supports downstream DTIG expression pattern clustering and gene regulatory network construction. In summary, the DyGAM-NS model, as an innovative approach for identifying dynamic temporal gene regulation, provides fresh insights into unraveling mechanisms of dimorphic transition and other complex biological processes, particularly in fungi and other microorganisms.

In this study, we utilized temporal transcriptomic data to investigate the dimorphic transition of *T. marneffei*, focusing on its distinct gene expression patterns at different stages of dimorphic transition. During the early stage of M-to-Y transition, we observed a substantial number of rapid upregulation genes associated with environmental adaptation, such as *pepA*, *ftrA*, *ftrC*, *sidA*, and *hmgR*. This observation suggests that *T. marneffei* undergoes a rapid adaptive response to the harsh environment

within host macrophages. In contrast, we also identified genes with delayed response patterns, which are linked to characteristics specific to mycelial or yeast phase growth but unrelated to environmental stimuli. For instance, the gene *hgrA*, involved in cell wall regulation, remained unchanged until 48 h into the Y-to-M transition. Functional analysis revealed that although *hgrA* is crucial for maintaining hyphal cell wall integrity, it does not directly participate in stress response, providing valuable insights into its upregulation during the later stage of the Y-to-M transition. Furthermore, we constructed a gene regulatory network for the dimorphic transition of *T. marneffei* using the DyGAM-NS model to identify key transcription factors. Interestingly, two zinc finger-containing transcription factors were found to potentially regulate the M-to-Y and Y-to-M transitions, respectively. These findings present new avenues for future research aimed at unraveling the regulatory mechanisms underlying the dimorphic transition of *T. marneffei*.

Data availability statement

The PacBio and Illumina sequencing data used for genome assembly have been deposited in the Genome Sequence Archive under the accession PRJCA002718. The reference genome sequences of *T. marneffei* PM1 have been deposited at DDBJ/ENA/GenBank under the accession JPOX02000000. The sequencing dataset for time-course RNA-seq of *T. marneffei* PM1 during M-to-Y and Y-to-M transitions have been deposited in the Sequence Read Archive under the accession PRJNA970557. The RNA-seq dataset used for identification of high variable genes during dimorphic transitions are available for download from the SRA under the accession PRJNA431116.

Author contributions

Funding

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2024.1369349/full#supplementary-material

References

Äijö, T., Butty, V., Chen, Z., Salo, V., Tripathi, S., Burge, C. B., et al. (2014). Methods for time series analysis of RNA-seq data with application to human Th17 cell differentiation. *Bioinformatics* 30, i113–i120. doi: 10.1093/bioinformatics/btu274

Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., et al. (2009). MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* 37, W202–W208. doi: 10.1093/nar/gkp335

Bar-Joseph, Z., Gitter, A., and Simon, I. (2012). Studying and modelling dynamic biological processes using time-series gene expression data. *Nat. Rev. Genet.* 13, 552–564. doi: 10.1038/nrg3244

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170

Borneman, A. R., Hynes, M. J., and Andrianopoulos, A. (2002a). A basic helix-loop-helix protein with similarity to the fungal morphological regulators, Phd1p, Efg1p and StuA, controls conidiation but not dimorphic growth in *Penicillium marneffei*. *Mol. Microbiol.* 44, 621–631. doi: 10.1046/j.1365-2958.2002.02906.x

Borneman, A. R., Hynes, M. J., and Andrianopoulos, A. (2002b). The *abaA* homologue of *Penicillium marneffei* participates in two developmental programmes: conidiation and dimorphic growth. *Mol. Microbiol.* 38, 1034–1047. doi: 10.1046/j.1365-2958.2000.02202.x

Boyce, K. J., and Andrianopoulos, A. (2015). Fungal dimorphism: the switch from hyphae to yeast is a specialized morphogenetic adaptation allowing colonization of a host. *FEMS Microbiol. Rev.* 39, 797–811. doi: 10.1093/femsre/fuv035

Boyce, K. J., De Souza, D. P., Dayalan, S., Pasricha, S., Tull, D., McConville, M. J., et al. (2018). *Talaromyces marneffei simA* encodes a fungal cytochrome P450 essential for survival in macrophages. *mSphere* 3, e00056–e00018. doi: 10.1128/mSphere.00056-18

Boyce, K. J., McLauchlan, A., Schreider, L., and Andrianopoulos, A. (2015). Intracellular growth is dependent on tyrosine catabolism in the dimorphic fungal pathogen *Penicillium marneffei*. *PLoS Pathog.* 11:e1004790. doi: 10.1371/journal.ppat.1004790

Brůna, T., Hoff, K. J., Lomsadze, A., Stanke, M., and Borodovsky, M. (2021). BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. *NAR Genom. Bioinform.* 3, 1–11. doi: 10.1093/nargab/lqaa108

Brennecke, P., Anders, S., Kim, J. K., Kołodziejczyk, A. A., Zhang, X., Proserpio, V., et al. (2013). Accounting for technical noise in single-cell RNA-seq experiments. *Nat. Methods* 10, 1093–1095. doi: 10.1038/nmeth.2645

Bugeja, H. E., Hynes, M. J., and Andrianopoulos, A. (2010). The RFX protein rfxA is an essential regulator of growth and morphogenesis in *Penicillium marneffei*. *Eukaryot Cell* 9, 578–591. doi: 10.1128/EC.00226-09

Bugeja, H. E., Hynes, M. J., and Andrianopoulos, A. (2013). HgrA is necessary and sufficient to drive hyphal growth in the dimorphic pathogen *Penicillium marneffei*. *Mol. Microbiol.* 88, 998–1014. doi: 10.1111/mmi.12239

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., et al. (2009). BLAST+: architecture and applications. *BMC Bioinform.* 10:421. doi: 10.1186/1471-2105-10-421

Conesa, A., and Götz, S. (2008). Blast2GO: a comprehensive suite for functional analysis in plant genomics. *Int. J. Plant Genomics* 2008, 1–12. doi: 10.1155/2008/619832

Cuomo, C. A., Shea, T., Nguyen, T., Ashton, P., Perfect, J., Le, T., et al. (2020). Complete genome sequences for two *Talaromyces marneffei* clinical isolates from northern and southern Vietnam. *Microbiol Resour. Ann.* 9, e01367–e01319. doi: 10.1128/MRA.01367-19

Dankai, W., Pongpom, M., Youngchim, S., Cooper, C. R., and Vanittanakom, N. (2016). The *yapA* encodes bZIP transcription factor involved in stress tolerance in pathogenic fungus *Talaromyces marneffei*. *PLoS ONE* 11:e0163778. doi: 10.1371/journal.pone.0163778

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., et al. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21. doi: 10.1093/bioinformatics/bts635

Fischer, D. S., Theis, F. J., and Yosef, N. (2018). Impulse model-based differential expression analysis of time course sequencing data. *Nucleic Acids Res.* 22:675. doi: 10.1093/nar/gky675

Fritz, H., García-Escudero, L. A., and Mayo-Iscar, A. (2012). tclust : an R package for a trimming approach to cluster analysis. *J. Stat. Softw.* 47:i12. doi: 10.18637/jss.v047.i12

Gauthier, G. M. (2017). Fungal dimorphism and virulence: molecular mechanisms for temperature adaptation, immune evasion, and *in vivo* survival. *Mediat. Inflamm.* 2017, 1–8. doi: 10.1155/2017/8491383

Gifford, T. D., and Cooper, C. R. (2009). Karyotype determination and gene mapping in two clinical isolates of *Penicillium marneffei*. *Med. Mycol.* 47, 286–295. doi: 10.1080/13693780802291437

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* 29, 644–652. doi: 10.1038/nbt.1883

Haas, B. J. (2003). Improving the *Arabidopsis* genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Res.* 31, 5654–5666. doi: 10.1093/nar/gkg770

Haas, B. J., Salzberg, S. L., Zhu, W., Pertea, M., Allen, J. E., Orvis, J., et al. (2008). Automated eukaryotic gene structure annotation using evidencemodeler and the program to assemble spliced alignments. *Genome Biol.* 9:R7. doi: 10.1186/gb-2008-9-1-r7

Jayakumar, V., and Sakakibara, Y. (2019). Comprehensive evaluation of non-hybrid genome assembly tools for third-generation PacBio long-read sequence data. *Brief Bioinform.* 20, 866–876. doi: 10.1093/bib/bbx147

Jones, P., Binns, D., Chang, H. Y., Fraser, M., Li, W., McAnulla, C., et al. (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30, 1236–1240. doi: 10.1093/bioinformatics/btu031

Klein, B. S., and Tebbets, B. (2007). Dimorphism and virulence in fungi. *Curr. Opin. Microbiol.* 10, 314–319. doi: 10.1016/j.mib.2007.04.002

Koren, S., Walenz, B. P., Berlin, K., Miller, J. R., Bergman, N. H., Phillippy, A. M., et al. (2017). Canu: scalable and accurate long-read assembly via adaptive *k* - mer weighting and repeat separation. *Genome Res.* 27, 722–736. doi: 10.1101/gr.215087.116

Law, C. W., Chen, Y., Shi, W., and Smyth, G. K. (2014). voom: precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol.* 15:R29. doi: 10.1186/gb-2014-15-2-r29

Leng, N., Li, Y., McIntosh, B. E., Nguyen, B. K., Duffin, B., Tian, S., et al. (2015). EBSeq-HMM: a Bayesian approach for identifying gene-expression changes in ordered RNA-seq experiments. *Bioinformatics* 31, 2614–2622. doi: 10.1093/bioinformatics/btv193

Liao, Y., Smyth, G. K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30, 923–930. doi: 10.1093/bioinformatics/btt656

Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15:550. doi: 10.1186/s13059-014-0550-8

Manni, M., Berkeley, M. R., Seppey, M., Simão, F. A., and Zdobnov, E. M. (2021). BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol. Biol. Evol.* 38, 4647–4654. doi: 10.1093/molbev/msab199

McCarthy, D. J., Chen, Y., and Smyth, G. K. (2012). Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res.* 40, 4288–4297. doi: 10.1093/nar/gks042

Michna, A., Braselmann, H., Selmansberger, M., Dietz, A., Hess, J., Gomolka, M., et al. (2016). Natural cubic spline regression modeling followed by dynamic network reconstruction for the identification of radiation-sensitivity gene association networks from time-course transcriptome data. *PLoS ONE* 11:e0160791. doi: 10.1371/journal.pone.0160791

Moerman, T., Aibar Santos, S., Bravo González-Blas, C., Simm, J., Moreau, Y., Aerts, J., et al. (2019). GRNBoost2 and Arboreto: efficient and scalable inference of gene regulatory networks. *Bioinformatics* 35, 2159–2161. doi: 10.1093/bioinformatics/bty916

Nemecek, J. C., Wüthrich, M., and Klein, B. S. (2006). Global control of dimorphism and virulence in fungi. *Science* 312, 583–588. doi: 10.1126/science.1124105

Nimmanee, P., Woo, P. C. Y., Vanittanakom, P., Youngchim, S., and Vanittanakom, N. (2014). Functional analysis of atfA gene to stress response in pathogenic thermal dimorphic fungus *Penicillium marneffei*. *PLoS ONE* 9:e111200. doi: 10.1371/journal.pone.0111200

Nueda, M. J., Tarazona, S., and Conesa, A. (2014). Next maSigPro: updating maSigPro bioconductor package for RNA-seq time series. *Bioinformatics* 30, 2598–2602. doi: 10.1093/bioinformatics/btu333

Pasricha, S., MacRae, J. I., Chua, H. H., Chambers, J., Boyce, K. J., McConville, M. J., et al. (2017). Extensive metabolic remodeling differentiates non-pathogenic and pathogenic growth forms of the dimorphic pathogen *Talaromyces marneffei*. *Front Cell Infect Microbiol.* 7:368. doi: 10.3389/fcimb.2017.00368

Pasricha, S., Schafferer, L., Lindner, H., Joanne Boyce, K., Haas, H., Andrianopoulos, A., et al. (2016). Differentially regulated high-affinity iron assimilation systems support growth of the various cell types in the dimorphic pathogen *Talaromyces marneffei*. *Mol. Microbiol.* 102, 715–737. doi: 10.1111/mmi.13489

Payne, M., Weerasinghe, H., Tedja, I., and Andrianopoulos, A. (2019). A unique aspartyl protease gene expansion in *Talaromyces marneffei* plays a role in growth inside host phagocytes. *Virulence* 10, 277–291. doi: 10.1080/21505594.2019.1593776

Pruksaphon, K., Nosanchuk, J. D., Ratanabanangkoon, K., and Youngchim, S. (2022). *Talaromyces marneffei* infection: virulence, intracellular lifestyle and host defense mechanisms. *JoF* 8:200. doi: 10.3390/jof8020200

Sapmak, A., Boyce, K. J., Andrianopoulos, A., and Vanittanakom, N. (2015). The *pbrB* gene encodes a laccase required for DHN-melanin synthesis in conidia of *Talaromyces (Penicillium) marneffei*. *PLoS ONE* 10:e0122728. doi: 10.1371/journal.pone.0122728

Shelest, E. (2008). Transcription factors in fungi. *FEMS Microbiol. Lett.* 286, 145–151. doi: 10.1111/j.1574-6968.2008.01293.x

Slater, G., and Birney, E. (2005). Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics* 6:31. doi: 10.1186/1471-2105-6-31

Solovyev, V., Kosarev, P., Seledsov, I., and Vorobyev, D. (2006). Automatic annotation of eukaryotic genes, pseudogenes and promoters. *Genome Biol.* 7:S10. doi: 10.1186/gb-2006-7-s1-s10

Todd, R. B., Greenhalgh, J. R., Hynes, M. J., and Andrianopoulos, A. (2003). TupA, the *Penicillium marneffei* Tup1p homologue, represses both yeast and spore development. *Mol. Microbiol.* 48, 85–94. doi: 10.1046/j.1365-2958.2003.03426.x

Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., et al. (2014). Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS ONE* 9:e112963. doi: 10.1371/journal.pone.0112963

Wang, Q., Du, M., Wang, S., Liu, L., Xiao, L., Wang, L., et al. (2018). MADS-box transcription factor *madsA* regulates dimorphic transition, conidiation, and germination of *Talaromyces marneffei*. *Front. Microbiol.* 9:1781. doi: 10.3389/fmicb.2018.01781

Woo, P. C. Y., Lam, C. W., Tam, E. W. T., Leung, C. K. F., Wong, S. S. Y., Lau, S. K. P., et al. (2012). First discovery of two polyketide synthase genes for mitorubrinic acid and mitorubrinol yellow pigment biosynthesis and implications in virulence of *Penicillium marneffei*. *PLoS Neglect. Trop Dis.* 6:e1871. doi: 10.1371/journal.pntd.0001871

Woo, P. C. Y., Tam, E. W. T., Chong, K. T. K., Cai, J. J., Tung, E. T. K., Ngan, A. H. Y., et al. (2010). High diversity of polyketide synthase genes and the melanin biosynthesis gene cluster in *Penicillium marneffei*. *FEBS J.* 277, 3750–3758. doi: 10.1111/j.1742-4658.2010.07776.x

Wood, S. N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *J. R. Statist. Soc. B* 73, 3–36. doi: 10.1111/j.1467-9868.2010.00749.x

Yang, E., Chow, W. N., Wang, G., Woo, P. C. Y., Lau, S. K. P., Yuen, K. Y., et al. (2014). Signature gene expression reveals novel clues to the molecular mechanisms of dimorphic transition in *Penicillium marneffei*. *PLoS Genet.* 10:e1004662. doi: 10.1371/journal.pgen.1004662

Zuber, S., Hynes, M. J., and Andrianopoulos, A. (2003). The G-protein alpha-subunit gasC plays a major role in germination in the dimorphic fungus *Penicillium marneffei*. *Genetics* 164, 487–499. doi: 10.1093/genetics/164.2.487