International Journal of Biomedical Science and Engineering
2026, Vol. 14, No. 3, pp. 68–77
https://doi.org/10.11648/j.ijbse.20261403.12

SciencePG
Science Publishing Group

Research Article

Systematic Genome-wide Identification and Analysis of Alternative Splicing in the Pathogenic Fungus *Talaromyces marneffei*

Xiangjia Min*, Tanya Pai Dhungat, Theoni Kasamias, Chester Cooper

Department of Chemical and Biological Sciences, Youngstown State University, Youngstown, USA

Abstract

Talaromyces marneffei, a thermally dimorphic saprophytic fungus, causes talaromycosis in immunocompromised patients. Differential gene expressions determine the morphology of this species at different growing temperatures. It is known that alternative splicing (AS) is common in eukaryotes, including fungal species, however, there is not a systematic AS analysis yet in *T. marneffei*. We collected over 2.3 billion paired RNA-sequencing reads and mapped these reads to the reference genome. A total of 36,769 AS events, including 3,512 alternative acceptor sites, 2,216 alternative donor site, 219 exon skipping, 12,442 intron retention, and 18,380 other events were identified by the genome-wide mapping analysis. These AS events were identified from 1,734 protein-coding gene models and 752 newly identified genomic loci, including 46 genes encoding carbohydrate active enzymes. The AS rate was estimated to be ~19.8%. A total of 46,249 unique RNA transcripts were assembled and 45,296 polypeptides were predicted and functionally annotated. Preliminary analysis using data collected from samples grown at 25 ℃ and 37 ℃ or in dimorphic transitions by switching temperatures identified treatment specific events, suggesting that AS may play some roles in establishing specific morphological characteristics in this species. The data collected in the work, including RNA-seq data mapping information, assembled RNA transcripts, identified AS events, and new genomic loci, provide a solid resource for further investigation of the gene regulations in the dimorphism of *T. marneffei*.

Keywords

Alternative Splicing, *Talaromyces Marneffei*, Fungus, RNA-Sequencing, Dimorphism, Carbohydrate Active Enzyme, Protein Family

1. Introduction

Talaromyces marneffei, formerly known as *Penicillium marneffei*, was renamed in 2011 based on several DNA marker sequences [1]. It was classified as an opportunistic pathogen as it causes infections, called talaromycosis, in immune-compromised individuals, particularly, HIV-positive persons [2]. *T. marneffei* is a thermally dimorphic saprophytic fungus which has a multicellular hyphal form growing around 25 ℃ or has a single celled yeast form around 37 ℃ [3]. The biology of *T. marneffei* and associated disease taloromycosis was recently comprehensively reviewed [2, 4].

Research efforts have been made to investigate the molecular mechanisms of the dimorphic switching in this species

*Correspondence: Xiangjia Min (xmin@ysu.edu)

Received: 27 June 2026; **Accepted:** 24 July 2026; **Published:** 26 August 2026

and more than 30 genes have been identified involved in regulation of *T. marneffei* growth and morphogenesis, including areA, abaA, msgA, hgrA, madsA, and drkA particularly involved in dimorphic transition [2]. Along with the applications of RNA-sequencing (RNA-seq) technologies, recently RNA-seq data have been generated in *T. marneffei* and dynamic transcriptomic changes during dimorphic transitions have been investigated [5-12]. Yang et al. (2013) measured global gene expression profiles of mycelial cells grown at 25 ℃ and yeast cells at 37 ℃ for a week [5]. A total of 2,861 differentially expressed genes were identified, including 1,447 over-expressed and 1,414 under-expressed in yeast cells at 37 ℃, compared with mycelial cells at 25 ℃ [5]. Further, Yang et al. (2014) collected samples in temperature switch experiments with yeast or mycelial growth plates were transferred to 25 ℃ or 37 ℃ for 6 hours, respectively, for RNA sequencing [6]. They classified 2,718 genes with variable expression levels in switching experiments and these genes were clustered into 14 different groups. Du et al. (2024) recently reported a large scale of gene expression profiling during the dimorphic transition of *T. marneffei* through time-course analysis [7]. They identified 5,223 dimorphic transition induced genes. In addition to the publications mentioned above on global gene expression studies related to dimorphic transitions, Lau et al. (2018) reported the effect of *T. marneffei* partitivirus-1 (TmPV1) on the gene expression in *T. marneffei*, with potential virulence factors up-regulated and RNA interference related genes down-regulated [8]. Wangsanut et al. (2024) compared gene expression of wildtype, the acuM deletion strain, and the acuK deletion strain growing at 25 ℃, demonstrated the role of AcuK and AcuM as globlal regulators controlling multiple cellular adaptations to harsh environmental conditions in *T. marneffei* [9]. Dan et al. (2024) reported that global gene expression profiling of *T. marneffei* grew at 37 ℃ and prediction and functional analysis of secreted proteins for discovery of biomarkers [10]. All these works through exploring RNA-seq data have significantly increased our understanding of gene regulations of *T. marneffei* growth and dimorphic development.

Intron-containing genes in eukaryotic organisms can be transcribed to generate multiple RNA transcripts through a process of alternative splicing (AS). Since AS alters the mRNA sequences, different isoform sequences may encode different functional proteins or may become non-coding. Numerous studies have shown that AS plays important biological roles in regulation of biological development, adaptations and evolution to the changing environments through increasing both the diversities of transcriptome and proteome [13, 14]. A survey reveals AS events in fungi ranged from 0.2% in non-pathogenic yeast Saccharomyces cerevisiae to 38.44% of expressed transcripts in *Shiraia bamlusicola* [15-17]. However, our recent study revealed that about 50% of expressed genes and up to 55.8% of intron containing genes subjected to AS in *Aspergillus niger* [18]. As mentioned above, research works have been published in *T. marneffei* using RNA-seq data for investigating global gene expression changes, however, only Yang et al. (2013) mentioned that there were 2,908 genes (26.3%) expressed more than one isoform through AS but detailed elaboration was not provided [5]. In considering the importance of *T. marneffei* as a human pathogen, collecting information on genes subjecting AS is necessary for understanding of the gene functional regulation of dimorphic transitions in the process of infections. In this work we carry out a genome-wide identification and analysis of AS events in *T. marneffei* by integrating a large amount of RNA-seq data obtained from multiple published projects. The aim is to compile a catalog of genes subjecting to AS and their associated isoform sequences. Such a collection and analysis of these genes subjected to the regulation of AS at different developmental stages and dimorphic transitions may provide information for further characterizing the biological functions of the products of these genes and for finding medical solutions for treatments of the diseases in patients caused by *T. marneffei* infection.

2. Materials and Methods

2.1. Reference Genome Sequences and RNA-seq Datasets

T. marneffei reference genome sequences with annotated genomic GFF (General Feature Format) file (Genome assembly ASM955685v1) and other related files were downloaded from the genome database of the National Center for Biotechnology Information (NCBI, https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_009556855.1/). The genome assembly was generated from isolate 11CN-20-091 with a total size of 28.2 Mb and 10,066 genes including 9,994 protein coding genes annotated [19].

The RNA-seq data were downloaded from the NCBI SRA database (https://www.ncbi.nlm.nih.gov/sra/docs/sradownload/) using SRA Toolkit (version 3.1.1). The RNA-seq datasets were selected from 8 projects with a total of 94 samples (Table 1). The detailed treatments of samples from which RNA-seq data were generated can be found from respective references [5-12] (Table 1).

2.2. RNA-seq Reads Mapping and AS Events Identification

The RNA-seq reads were mapped to the reference genome sequences using HISAT2 (v2.2.1) with default parameters [20]. HISAT2 is a fast and sensitive alignment program for mapping next-generation sequencing reads to a single reference genome. The mapping file generated by HISAT2 was further processed using Samtools and StringTie tool for generating the transcripts GTF (Gene Transfer Format) file following the protocol described by Pertea et al. (2016) [21]. The individual sample "transcripts.gtf" files were further

merged using StringTie merge function for samples in each project and for the whole work. The default parameters were used with "-G <reference.gff>" for providing a reference annotation to guide the merge and include known isoforms and "-i" for keeping merged transcripts with retained introns, as the intron retention type is known to be a prevalent AS event type in fungal genes. Astalavista was used for AS event classification [22]. Basic AS events are generally classified as exon skipping (ES), alternative donor site (AltD), alternative accepter site (AltA), intron retention (IR) and mutually exclusive exons (MXE). Other "complex" events can be formed with two or more basic AS events, such as an event containing both AltA and AltD.

AS events were analyzed in selected samples for identifying morphology specific and morphology transition specific AS events using in-house tools. The Venn diagram was generated using R4.2.1.

Table 1. *RNA-seq data sources and associated references.*

Projects	Samples	Length (bp)	Treatments	References
PRJNA1041082	3	150	Wildtype (ATCC18224), acuK or acuM deletion strains at 25 ℃	[9]
PRJNA1108869	3	150	Wildtype (ATCC18224) at 37 ℃	[10]
PRJNA1163092	4	150	Wildtype (ATCC18224) at 37 ℃ yeast form and 25 ℃ filamentous	[12]
PRJNA212740	4	90	Strain PM1 mycelia at 25 ℃ or yeast cells at 37 ℃	[5]
PRJNA251718	4	90	Strain PM1 dimorphic transitions by switching (25 or 37 ℃)	[6]
PRJNA353903	8	101	Strain PM1 TmPV1- free or infected isolates of *T. marneffei*	[8]
PRJNA431116	8	100	Strain PM1 and the madsA deletion nutant dimorphic transitions at 25 ℃ or at 37 ℃	[11]
PRJNA970557	60	150	Strain PM1 time-course profiling during dimorphic transitions	[7]

2.3. Functional Annotation of Transcripts and Data Availability

The assembled RNA transcript sequences were retrieved using gtf_to_fasta tool in the TopHat package based on the GTF file generated by StringTie tool after merging all GTF files. These transcripts were functionally annotated, including protein sequence prediction, Translated BLAST (Basic local alignment search tool; BLASTX) against UniProt/Swiss-Prot database (cutoff E-value 1e-5), protein family (Pfam) search (cutoff E-value 1e-5), and mapping their genomic locations with reference gene models [23]. The Pfams of carbohydrate-active enzymes (CAZymes) were analyzed based on the information of the CAZyme classification (http://www.cazy.org/Home.html) [24]. The transcript sequences, detailed information of AS events, and supplementary files are available at our bioinformatics site (http://bioinformatics.ysu.edu/publication/data/Tmarneffei/).

3. Results

3.1. Mapping RNA-seq Data to the Genome

We collected and then mapped a total of 94 RNA-seq datasets to the reference genome of *T. marneffei*. The accession numbers and detailed mapping information of all samples data can be found in supplementary Table 1. The mapping information for each project was summarized in Table 2. The overall mapping rates varied in different projects including >90% mapping rates in 7 projects and only 70.3% in PRJNA353903. Combining all data of 8 projects revealed that a total of 2.3 billion paired reads were collected with 2.0 billion reads (~87.3.0%) being mapped uniquely concordantly to the genome and about 2.9% reads mapped to more than one genomic locus. The overall mapping rate including both mapped paired mates and single mates in all combined RNA-seq data was 92.1% (Table 2).

Table 2. *Summary of mapping rates of RNA-seq data of T. marneffei.*

Project ID	Total Paired Reads (millions)	Mapping Rate (%)
PRJNA1041082	63.9	95.7
PRJNA1108869	67.5	95.1
PRJNA1163092	89.0	91.0
PRJNA212740	53.7	96.1
PRJNA251718	53.7	93.6
PRJNA353903	274.8	70.3
PRJNA431116	256.2	96.2
PRJNA970557	1458.2	95.0
Total	2316.9	92.1

3.2. Identification of AS Events and Associated Genomic Loci or Genes

RNA-seq data were mapped individually for each sample and then the mapped Bam files were merged using StringTie program in each project and all Bam files were merged for generation of the final assembly of the transcripts for annotation. AS events were identified in each project and in all merged data using Astalavista software and the results were shown in Table 3. The number of total AS events was highly positively correlated with the total mapped reads or mates in each project (correlation coefficient R = 0.987) (Supplement Figure 1). However, the length of reads and sequencing depth of transcriptomes may also affect the number of AS events identified in each project. The results of each category of basic AS events in *T. marneffei* are consistent with data reported in other fungal species. That is, among basic AS events, IR (34.9 – 50.0%) is the major type of AS events, followed by AltA (9.2 – 16.9%) and AltD (6.2 – 8.9%), and ES (0.7 – 1.0%) represented the least AS category (Table 3). The “others” category includes MXEs and other complex events which have two or more basic AS categories resulting in an AS event. Combining all RNA-seq data collected in current work we identified a total of 36,769 AS events, including 3,512 AltA, 2,216 AltD, 219 ES, 12,422 IR, and 18,380 other events. It is noticed that relatively there was a higher percentage of “other” events in all combined data, as the more data are used the more complex events can be identified. These AS events were identified from a total of 31,311 assembled transcripts which were generated from 4,380 genomics loci assembled by StringTie.

Based on RNA-seq mapping of the merged data, there were a total of 8,819 genomic loci assembled by StringTie program, thus, the estimated AS rate was 49.7% in *T. marneffei*. However, this rate was over estimated. There is a known "super loci" problem in RNA-seq assembly when using tools like Cufflinks or StringTie. It refers to the erroneous merging of distinct, neighboring genes into a single large transcript or genomic locus. This oftenoccurs in compact genomes or regions with high gene density, where multiple, separate genes are mistakenly connected due to sequencing artifacts, read-through transcripts, or low abundance “transfrags” bridging transcripts that span intergenic regions. While StringTie generally outperforms Cufflinks by better resolving splice junctions and producing fewer super-transcripts, it can still generate "super-loci" that span multiple reference genes. To avoid the “super loci” problem, we then used annotated gene models of the reference genome to estimate the AS rate, i. e., only the alternatively spliced sites located within transcripts of gene models were used. Among a total of 9,994 protein coding genes annotated in the reference genome, 1,734 genes were identified alternatively spliced, thus, the AS rate among protein coding genes was ~17.4%. The list of alternatively spliced reference protein coding genes was available (see information in Data Availability).

We performed a comparison of StringTie assembled loci with annotated genomic loci in the reference genome based on the genomic locations. There were 4,512 StringTie loci having one to one match with reference gene models and 2,113 StringTie loci having two or more matches with reference gene models (Table 4). Two “super loci” were MSTRG.7101 and MSTRG.2548 matched 29 and 25 gene models, respectively. These StringTie assembled genomic loci require manual examination when the data are used for a particular research work. We also identified 2,554 novel genomic loci assembled by StringTie based on RNA-seq data mapping but not overlapping with any annotated gene models. Among 2,554 new loci assembled by StringTie, 752 loci generated alternatively spliced transcripts. Combining the number of AS gene models (1,734 out of 9,994) with newly identified genomic loci, thus, the genome-wide AS rate was estimated to be ~19.8%. The functional significance of these novel genomic

loci was described in the section of functional analysis of transcripts.

Table 3. *Alternative splicing events identified in each RNA-seq project in T. marneffei.*

Projects	AltA (%)	AltD (%)	ExonS (%)	IntronR (%)	Others (%)	Total
PRJNA1041082	542 (16.9)	264 (8.2)	28 (0.9)	1566 (48.8)	808 (25.2)	3208
PRJNA1108869	599 (13.7)	392 (8.9)	46 (1.0)	2079 (47.4)	1267 (28.9)	4383
PRJNA1163092	660 (15.9)	369 (8.9)	33 (0.8)	1963 (47.2)	1132 (27.2)	4157
PRJNA212740	693 (14.0)	415 (8.4)	47 (1.0)	2472 (50.0)	1318 (26.7)	4945
PRJNA251718	825 (13.7)	493 (8.2)	49 (0.8)	2966 (49.2)	1696 (28.1)	6029
PRJNA353903	722 (9.2)	485 (6.2)	51 (0.7)	3691 (47.2)	2865 (36.7)	7814
PRJNA431116	1133 (12.4)	709 (7.7)	66 (0.7)	3996 (43.6)	3262 (35.6)	9166
PRJNA970557	2772 (11.2)	1749 (7.1)	180 (0.7)	8662 (34.9)	11433 (46.1)	24796
All merged	3512 (9.6)	2216 (6.0)	219 (0.6)	12442 (33.8)	18380 (50.0)	36769

3.3. Functional Annotation of Transcripts Assembled by StringTie

A total of 46,249 RNA transcript sequences, with an average length of 3,461 bp, were assembled by StringTie tool (Table 4). The set of transcripts were functionally annotated using BLASTX search against UniProt/Swiss-Prot database. A total of 45,296 protein sequences, with an average size of 440 amino acids, were predicted using OrfPredictor from the set of transcripts. The predicted protein sequences were further mapped to protein families (Pfam) (Table 4). A total of 28,053 predicted proteins had Pfam matches (Table 4). While 7,378 of 9,994 (74.3%) gene model proteins in the reference genome had Pfam matches, 1,404 of 1734 (81.0%) gene model proteins coded from alternatively spliced genomic loci had Pfam matches. We then used protein sequences of gene models annotated by NCBI reference genome project for evaluating the impact of AS on the functionalities of the predicted proteins encoded by the isoform transcripts. Within the subset of 1,404 gene models undergoing AS with proteins having Pfam matches, a total of 9,675 isoform transcripts were used for comparing the Pfams of predicted polypeptides. We found 6,326 (65.4%) had Pfams identical to gene model proteins, 2,451 (25.6%) had different Pfams from gene model proteins, and 898 (9.3%) had no Pfam matches.

Since carbohydrate-active enzymes (CAZymes) may play important roles in the evolution of *T. marneffei*, we identified a total of 244 CAZymes from predicted gene models, and among 46 of them were alternatively spliced. In our previous work we identified 123 out of 251 genes in *Aspergillus niger* and 86 out of 267 genes encoding CAZymes in *A. fumigatus* were alternatively spliced (Table 5). The relatively lower AS rate in these genes in *T. marneffei* was likely due to *T. marneffei* were not challenged with variable carbon sources as used in *A. niger* and *A. fumigatus* [18, 27]. The functions of these CAZyme isoforms generated by AS need to be further investigated in the regulation of carbohydrate metabolism in *T. marneffei*.

Table 4. *Basic features of assembled genomic loci and RNA transcripts assembled by StringTie in T. marneffei.*

Features	Counts (%)
Total genomic loci	8819
Loci with one transcript	2807 (31.8%)
Loci with two or more transcripts	6012 (68.2%)
Loci mapped to gene model loci	6265 (71.0%)
Loci not mapped to gene models	2554 (29.0%)
Total unique transcripts	46249

Features	Counts (%)
Average transcript length (bp)	3461
Transcripts mapped to gene model loci	37662 (81.4%)
BLASTX search against Swiss-Prot dataset	31524 (68.2%)
Total predicted ORFs (M20)	45296 (97.9%)
Average ORF length (amino acids)	440
Total ORFs match with Pfam	28053 (61.9%)

*ORF: open reading frame, polypeptide; M20: polypeptides with M as the first amino acid and a minimum length of 20 amino acids.

Within 2,554 novel genomic loci which were not annotated in the reference genome, we retrieved 7,094 transcripts and 6,422 polypeptides were predicted from them. Further, 1,226 transcripts had BLASTX matches and 573 predicted polypeptides had Pfam matches. Thus, some of the newly identified genomic loci may represent real protein coding genes. The biological roles of these newly identified genes need further investigations. The complete list of Pfam classifications for all assembled transcripts, gene model proteins including AS gene models can be found in Supplementary Table 2.

3.4. Dynamic Regulations of AS During Dimorphic Transitions

To analyze AS variations in *T. marneffei* with different morphology and in dimorphic transitions grown at different temperature treatments, we used data collected by Yang et al. [5, 6] (PRJNA212740, 2013; PRJNA251718, 2014). The AS events were identified in samples with mycelial form (grown at 25 ℃), yeast form (grown at 37 ℃), mycelia transferred from 25 to 37 ℃ for 6 hours, and yeast form transferred from 37 to 25 ℃ for 6 hours (Supplementary Table 3). Though there were some fluctuations in the total numbers of AS events among different treatments, the proportions of each basic type of AS events remained to be consistent. That is, IR was the most abundant AS type, followed by AA, AD, and ES represented the lowest AS type. However, when we compared the AS events among the treatments using replicates merged data for each treatment, a large proportion (66.2 – 73.5%) of treatment specific AS events was evident in all treatments, demonstrating the dynamic changes of AS in morphological development in responses to growing temperatures (Figure 1; Supplementary Table 3). We also noted that relatively there were more AS events per million mapped RNA-seq mates (fragments) in samples grown at 37 ℃ than 25 ℃ or transferred from 25 to 37 ℃ than from 37 to 25 ℃, suggesting *T. marneffei* grown at 37 ℃ was stressed or disturbed, thus, using the AS mechanism to cope with the stressful conditions. The dynamic changes in AS profiles were further demonstrated when the AS events were analyzed using the samples collected in the time-course experiments during dimorphic transitions [7](Project: PRJNA970557) (Figure 2). The identical AS events and new AS events were identified by comparing AS events at different time points with their previous time points during dimorphic transitions (Figure 2). The trend was clear that there were more AS events for samples transferred to 37 ℃ from 25 ℃ than samples transferred to 25 ℃ from 37 ℃. The dynamic responses of AS during the dimorphic transitions revealed that AS regulation during morphological transitions was intensified. The exact roles of AS in the development of dimorphic transitions in this species still need to be elucidated.

Table 5. *Comparison of alternatively spliced genes encoding carbohydrate metabolism enzymes in different fungal species.*

Pfam ID	Pfam Domain	Pfam Description	*T. Marneffei*		*A. niger*		*A. fumigatus*	
			Total	AS	Total	AS	Total	AS
pfam00704	Glyco_hydro_18	Glycosyl hydrolases family 18	27	6	9	6	16	9
pfam01915	Glyco_hydro_3_C	Glycosyl hydrolase family 3 C-terminal	9	2	13	8	14	4
pfam01055	Glyco_hydro_31	Glycosyl hydrolases family 31	8	3	7	3	7	4
pfam00722	Glyco_hydro_16	Glycosyl hydrolases family 16	8	0	5	1	6	2
pfam04488	Gly_transf_sug	Glycosyltransferase sugar-binding	8	0	9	3	5	0

Pfam ID	Pfam Domain	Pfam Description	*T. Marneffei*		*A. niger*		*A. fumigatus*	
			Total	AS	Total	AS	Total	AS
pfam13641	Glyco_tranf_2_3	Glycosyltransferase like family 2	7	2	4	2	5	1
pfam04616	Glyco_hydro_43	Glycosyl hydrolases family 43	7	1	10	4	16	6
pfam00295	Glyco_hydro_28	Glycosyl hydrolases family 28	7	1	16	5	11	3
pfam03663	Glyco_hydro_76	Glycosyl hydrolase family 76	7	1	11	6	8	2
pfam03659	Glyco_hydro_71	Glycosyl hydrolase family 71	7	1	7	2	8	2
pfam01532	Glyco_hydro_47	Glycosyl hydrolase family 47	7	0	5	3	5	2
pfam00128	Alpha-amylase	Alpha amylase	6	3	14	7	12	4
pfam00982	Glyco_transf_20	Glycosyltransferase family 20	6	3	6	4	7	4
pfam01522	Polysacc_deac_1	Polysaccharide deacetylase	6	1	7	2	6	1
pfam11790	Glyco_hydro_cc	Glycosyl hydrolase catalytic core	6	0	2	0	2	0
pfam07971	Glyco_hydro_92	Glycosyl hydrolase family 92	5	3	5	4	4	2
pfam01793	Glyco_transf_15	Glycolipid 2-alpha-mannosyltransferase	5	1	3	2	3	3
pfam00734	CBM_1	Fungal cellulose binding domain	5	1	1	1	3	2
pfam00150	Cellulase	Glycosyl hydrolase family 5	5	0	4	4	8	4
Total	-	-	244	46	251	123	267	86

*This table only contains a partial list of carbohydrate active enzyme families. The full list of all protein families can be found in Supplementary Table 2.

4. Discussion

Gene expressions in *T. marneffei* have been investigated particularly during the dimorphic transitions [5-7]. However, all the above-mentioned analysis were carried out at the gene level for differential analysis. Certainly, there were important insights obtained in analyzing cellular activities at gene levels.

As recent reports on AS in fungal species revealed that AS was widespread in the process of gene expression among fungal species [16, 18], thus, it would be more accurate to examine the functionalities of gene expression products at the transcript isoform level. However, to quantify transcript isoform expression, we must have a catalog of transcripts available for a given species. This was the aim of our current work. We collected and mapped more than 2.3 billion paired RNA-seq reads to the reference genome. The data were generated from 94 samples including wild-type strains grown at different temperatures or mutants or virus infected sample (Table 1; Table 2). A total of 46,249 RNA transcripts were assembled and functionally annotated including ORF prediction, BLASTX search with Swiss-Prot database and matching to protein families (Table 4).

To our knowledge, this is the first large scale genome-wide meta-analysis of AS events in *T. marneffei*. A total of 36,769 AS events were identified from 31,311 unique transcripts, which were generated from 2,486 genomic loci, including 1,734 protein-coding gene model loci and 752 newly identified genomic loci (Tables 3 and 4). As expected, we found there were 46 genes encoding CAZymes generated AS events since *T. marneffei* and Aspergillus species were phylogenetically close [8, 26]. Our previous analysis identified a large proportion of mRNA transcripts of CAZyme genes alternatively spliced in Aspergillus species [18, 25].

Our preliminary analysis of the occurrence of AS events in *T. marneffei* grown at different temperatures or in the process of dimorphic transitions revealed AS were dynamically regulated in response to changing growth conditions, suggesting AS was involved in the process of morphological formation (Figure 1; Figure 2). We computationally evaluated the impact of AS on the protein domains of isoform transcripts in gene models, however, the real impact of AS events on the encoded protein functions needs to be further evaluated using experimental approaches by RT-PCR or cloning the full-length of mRNA transcripts such as used in *A. niger* [27]. As our analysis and other reports show AS in fungal genes is a common process, we suggest that fungal community integrates AS analysis when performing transcriptomic studies for analysis of differentially expressed genes. Recently, Neves-da-Rocha et al. (2026) provided thoughtful insights and critical discussions on the role of AS and its importance

in fungal diversification and adaptability [28]. Our data including the assembled transcript sequences, RNA-seq data mapping files, AS events and newly identified genomic loci are publicly available for the community to experimentally verify the identified isoform sequences and explore their functional novelties of gene products for future investigation in the regulation of developmental processes and applications in the control of the fungal pathogen.

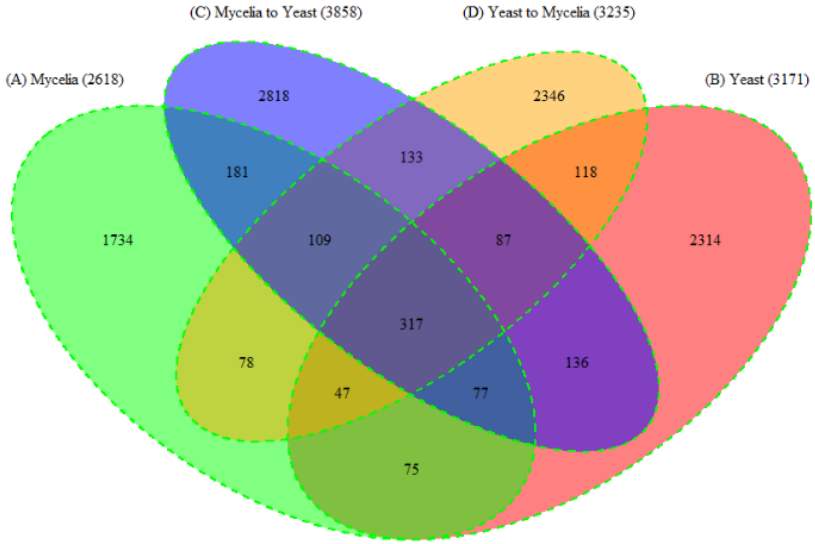

Figure 1. *Comparison of alternative splicing events in T. marneffei growing at different temperatures. (A) Mycelia grown at 25 ℃ for one week; (B) Yeast form grown at 37 ℃ for one week; (C) Mycelia from 25 ℃ were transferred to 37 ℃ for 6 hours; (D) Yeast cells from 37 ℃ were transferred to 25 ℃ for 6 hours.*

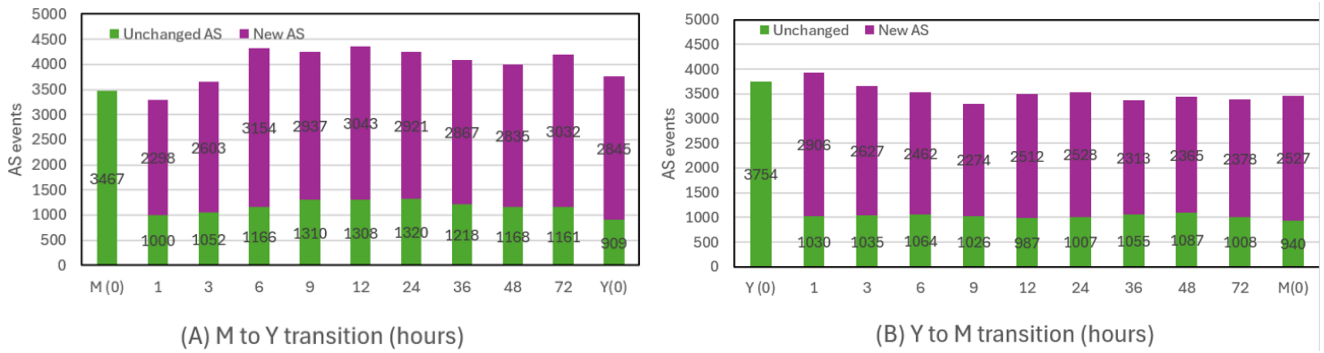

Figure 2. *Dynamic changes in alternative splicing during dimorphic transitions of T. marneffei. (A): Time-course of transition of mycelial cells(M) to yeast cells (Y) by transferring growth plates from 25 ℃ to 37 ℃. (B): Time-course of transition of yeast cells to mycelial cells by transferring growth plates from 37 ℃ to 25 ℃. The unchanged AS events were obtained by comparing the sample at each time point with the sample collected at immediate previous time point.*

5. Conclusions

We systematically collected RNA-seq data from multiple gene expression projects in *T. marneffei* to identify alternatively spliced genes and their associated transcript isoforms. In total 2.3 billion paired RNA-seq reads were collected and 92.1% of the reads were mapped to the reference genome. A total of 36,769 AS events were identified by integrating multiple genome-wide mapping analysis. These AS events were identified from 1,734 protein-coding gene models and 752 newly identified genomic loci, including 46 genes encoding carbohydrate active enzymes. The AS rate was estimated to be ~19.8%. A total of 46,249 unique RNA transcripts were assembled and 45,296 polypeptides were predicted and functionally annotated. The analysis using data collected from samples grown at 25 ℃ and 37 ℃ or in dimorphic transitions by switching temperatures identified treatment specific events, suggesting that AS be involved in establishing specific morphological characteristics. The assembled RNA transcripts, identified AS events, and new identified genomic loci, provide a resource for further investigation of the gene regulations in the dimorphism of *T. marneffei*.

Abbreviations

AS	Alternative Splicing
RNA-seq	RNA-sequencing
TmPV1	T. Marneffei partitivirus-1
ES	Exon Skipping
AltD	Alternative Donor Site
AltA	Alternative Accepter Site
IR	Intron Retention
MXE	Mutually Exclusive Exons
SRA	Sequence Read Archive
GFF	General Feature Format
GTF	Gene Transfer Format
Pfams	Protein Family
CAZymes	Carbohydrate-active Enzymes
ORF	Open Reading Frame
BLAST	Basic Local Alignment Search Tool

Supplementary Material

The supplementary material can be accessed at https://doi.org/10.11648/j.ijbse.20261403.12

Acknowledgments

The work was supported by Youngstown State University (YSU) College of STEM dean's reassigned time to XM.

Author Contributions

Xiangjia Min: Conceptualization, Data curation, Formal Analysis, Methodology, Software, Writing – original draft, Writing – review & editing

Tanya Pai Dhungat: Data curation, Formal Analysis

Theoni Kasamias: Data Curation, Formal Analysis

Chester Cooper: Conceptualization, Formal Analysis, Writing – review & editing

Funding

Youngstown State University and the Ohio Supercomputer Center provided computational facilities for supporting data processing of the project.

Data Availability Statement

The data that support the findings of this study can be found at: https://bioinformatics.ysu.edu/publication/data/Tmarneffei/

Conflicts of Interest

The authors declare no conflicts of interest.

References

[1] Samson, R. A., Yilmaz, N., Houbraken, J., Spierenburg, H., Seifert, K. A., Peterson, S. W., Varga, J., Frisvad, J. C. Phylogeny and nomenclature of the genus Talaromyces and taxa accommodated in Penicillium subgenus Biverticillium. Stud Mycol. 2011, 70: 159–183. https://doi.org/10.3114/sim.2011.70.04 66-75

[2] Wang, F., Han, R., Chen, S. An overlooked and underrated endemic mycosis—Talaromycosis and the pathogenic fungus Talaromyces marneffei. Clinical microbiology reviews. 2023. 36(1): e00051-22. https://doi.org/10.1128/cmr.00051-22

[3] Tsang, C. C., Lau, S. K., Woo, P. C. Sixty years from Segretain's description: what have we learned and should learn about the basic mycology of Talaromyces marneffei? Mycopathologia, 2019. 184(6): 721-729. https://doi.org/10.1007/s11046-019-00395-y

[4] Pruksaphon, K., Nosanchuk, J. D., Ratanabanangkoon, K., Youngchim, S. Talaromyces marneffei infection: virulence, intracellular lifestyle and host defense mechanisms. Journal of Fungi. 2022. 8(2): 200. https://doi.org/10.3390/jof8020200

[5] Yang, E., Wang, G., Woo, P. C., Lau, S. K., Chow, W. N., Chong, K. T., Tse, H., Kao, R. Y., Chan, C. M., Che, X., Yuen, K. Y. Unraveling the molecular basis of temperature-dependent genetic regulation in Penicillium marneffei. Eukaryotic Cell. 2013. 12(9): 1214-1224. https://doi.org/10.1128/ec.00159-13

[6] Yang, E., Chow, W. N., Wang, G., Woo, P. C., Lau, S. K., Yuen, K. Y., Lin, X., Cai, J. J. Signature gene expression reveals novel clues to the molecular mechanisms of dimorphic transition in Penicillium marneffei. PLoS genetics. 2014. 10(10): p. e1004662. https://doi.org/10.1371/journal.pgen.1004662

[7] Du, M., Tao, C., Hu, X., Zhang, Y., Kan, J., Wang, J., Yang, E. Unraveling the dynamic transcriptomic changes during the dimorphic transition of Talaromyces marneffei through time-course analysis. Frontiers in Microbiology. 2024. 15: 1369349. https://doi.org/10.3389/fmicb.2024.1369349

[8] Lau, S. K., Lo, G. C., Chow, F. W., Fan, R. Y., Cai, J. J., Yuen, K. Y., Woo, P. C. Novel partitivirus enhances virulence of and causes aberrant gene expression in Talaromyces marneffei. MBio, 2018. 9(3): 10-1128. https://doi.org/10.1128/mbio.00947-18

[9] Wangsanut, T., Amsri, A., Kalawil, T., Sukantamala, P., Jeenkeawpieam, J., Andrianopoulos, A., Pongpom, M. (2024). AcuM and AcuK: The global regulators controlling multiple cellular metabolisms in a dimorphic fungus Talaromyces marneffei. PLOS Neglected Tropical Diseases. 2024. 18(9): e0012145. https://doi.org/10.1371/journal.pntd.0012145

[10] Dan, J., Wei, W., Ou, W., Gao, G., Song, W., Ye, L., Liang, H., Guo, X., Tan, L., Jiang, J. Excavation of Biomarker Candidates for the Diagnosis of Talaromyces marneffei Infection via Genome-Wide Prediction and Functional Annotation of Secreted Proteins. ACS omega. 2024. 9(25): 27093-27103. https://doi.org/10.1021/acsomega.4c00571

[11] Wang, Q., Du, M., Wang, S., Liu, L., Xiao, L., Wang, L., Li, T., Zhuang, H. and Yang, E. MADS-box transcription factor madsA regulates dimorphic transition, conidiation, and germination of Talaromyces marneffei. Frontiers in Microbiology. 2018. 9: 1781. https://doi.org/10.3389/fmicb.2018.01781

[12] Luo, J., Bian, J., Murillo, M., Hau, P. T., Feng, Y., Chau, E. C. T., Yan, Y., Ng, L. C., Parsha, A. S., Siu, G. K. H. and Chow, F. W. N. High-quality genome assembly and comparative analysis reveal extensive genomic variation in Talaromyces marneffei. Microbial genomics. 2025. 11(4): 001400. https://doi.org/10.1099/mgen.0.001400

[13] Wright, C. J., Smith, C. W. and Jiggins, C. D. Alternative splicing as a source of phenotypic diversity. Nature Reviews Genetics. 2022. 23(11): 697-710. https://doi.org/10.1038/s41576-022-00514-4

[14] Verta, J. P., Jacobs, A. The role of alternative splicing in adaptation and evolution. Trends in Ecology & Evolution. 2022. 37(4): 299-308. https://doi.org/10.1016/j.tree.2021.11.010

[15] Grützmann, K., Szafranski, K., Pohl, M., Voigt, K., Petzold, A., Schuster, S. Fungal alternative splicing is associated with multicellular complexity and virulence: a genome-wide multi-species study, DNA Res. 2014. 21(1): 27-39. https://doi.org/10.1093/dnares/dst038

[16] Fang, S., Hou, X., Qiu, K., He, R., Feng, X., Liang, X. Theoccurrence and function of alternative splicing in fungi. Fungal Biol. Rev. 2020. 34(4): 178-188. https://doi.org/10.1016/j.fbr.2020.10.001

[17] Liu, X. Y., Fan, L., Gao, J., Shen, X. Y., Hou, C. L. Global identification of alternative splicing in Shiraia bambusicola and analysis of its regulation in hypocrellin biosynthesis, Appl. Microbiol. Biotechnol. 2020. 104: 211-223. https://doi.org/10.1007/s00253-019-10189-3

[18] Min, X., Jones, C., Logan, B., Campean, M., Wekhyan, B., Dasana, C., Patel, A., Mudiyanselage, P., Adeyemo, K. and Yu, F. Genome-Wide Identification and Analysis of Alternative Splicing in Aspergillus niger. Computational Molecular Biology. 2025. 15: 1-12. https://doi.org/10.5376/cmb.2025.15.0001

[19] Cuomo, C. A., Shea, T., Nguyen, T., Ashton, P., Perfect, J. and Le, T. Complete genome sequences for two Talaromyces marneffei clinical isolates from northern and southern Vietnam. Microbiology Resource Announcements. 2020. 9(2): 1110-1128. https://doi.org/10.1128/mra.01367-19

[20] Kim, D., Paggi, J. M., Park, C., Bennett, C. and Salzberg, S. L. 2019, Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat. Biotechnol. 2019. 37: 907–915. https://doi.org/10.1038/s41587-019-0201-4

[21] Pertea, M., Kim, D., Pertea, G. M., Leek, J. T., Salzberg, S. L. Tran-script-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. Nat Protoc. 2016. 11(9): 1650-67. https://doi.org/10.1038/nprot.2016.095

[22] Foissac, S., and Sammeth, M. ASTALAVISTA: dynamic and flexible analysis of alternative splicing events in custom gene datasets, Nucleic Acids Res. 2007. 35: W297–299. https://doi.org/10.1093/nar/gkm311

[23] Min, X. J., Butler, G., Storms, R. and Tsang, A. OrfPredictor: predicting protein-coding regions in EST-derived sequences. Nucleic acids research. 2005. 33: W677-W680. https://doi.org/10.1093/nar/gki394

[24] Drula, E., Garron, M. L., Dogan, S., Lombard, V., Henrissat, B., Terrapon, N. The carbohydrate-active enzyme database: functions and literature. Nucleic Acids Res. 2022. 50: D571–D577. https://doi.org/10.1093/nar/gkab1045

[25] Adeyemo, K. and Min, X. J. Genome-wide Analysis of Alternative Splicing in Aspergillus fumigatus in Response to Different Carbon Sources. Computational Molecular Biology. 2026. 16: 1-10. https://doi.org/10.5376/cmb.2026.16.0001

[26] Asch, D. K., Ziegler, J., Min, X. J. Molecular evolution of genes involved in quinic acid utilization in fungi. Computational Molecular Biology. 2021. 11(5): 1-15. https://doi.org/10.5376/cmb.2021.11.0005

[27] Xu, Y., Dong, F., Wang, R., Ajmal, M., Liu, X., Lin, H., Chen, H. Alternative splicing analysis of lignocellulose-degrading enzyme genes and enzyme variants in Aspergillus niger, Appl. Microbiol. Biotechnol. 2024. 108(1): 1-11. https://doi.org/10.1007/s00253-024-13137-y

[28] Neves-da-Rocha, J., Lopes, M. E., Santos-Saboya, M. J., Azevedo, M. I., Petrucelli, M. F., Rossi, A., Martinez-Rossi, N. M. Omics-based dissection of alternative splicing and the regulatory architecture in fungal genomes. Current Research in Microbial Sciences. 2026. 10: 100555. https://doi.org/10.1016/j.crmicr.2026.100555