

Research Article

Open Access

Genome-wide Analysis of Alternative Splicing in *Aspergillus fumigatus* in Response to Different Carbon Sources

Kolade Adeyemo, Xiangjia Jack Min ✉

Department of Chemical and Biological Sciences, Youngstown State University, OH 44555, USA

✉ Corresponding author: xmin@ysu.eduComputational Molecular Biology, 2026, Vol.16, No.1 doi: [10.5376/cmb.2026.16.0001](https://doi.org/10.5376/cmb.2026.16.0001)

Received: 26 Nov., 2025

Accepted: 19 Jan., 2026

Published: 29 Jan., 2026

Copyright © 2026 Adeyemo and Min, This is an open access article published under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Preferred citation for this article:

Adeyemo K., and Min X.J.J., 2026, Genome-wide analysis of alternative splicing in *Aspergillus fumigatus* in response to different carbon sources, Computational Molecular Biology, 16(1): 1-10 (doi: [10.5376/cmb.2026.16.0001](https://doi.org/10.5376/cmb.2026.16.0001))

Abstract *Aspergillus fumigatus* is a saprophytic filamentous fungus that has evolved into an opportunistic pathogen, particularly affecting immunocompromised individuals. RNA sequencing (RNA-seq) experiments have been conducted to investigate the differential gene expression in wild type and *cotA* mutant of *A. fumigatus* in response to different carbon sources. We used the RNA-seq data generated by Martin-Vincete et al. (2024) to investigate alternative splicing (AS) landscape in *A. fumigatus* growing in different carbon sources including glucose, acetate, and ethanol. Comparison of AS profiles between the wild-type strain and *cotA* mutant in each carbon source identified strain specific and carbon source specific AS events. Combining all RNA-seq mapping data we identified a total of 341 exon skipping (ES), 1,404 alternative donor site (AltD), 3,024 alternative acceptor site (AltA), 10,312 intron retention (IR), and 12,898 other splicing events. Among 9,857 reference gene models, 2,823 gene models with pre-RNA transcripts were alternatively spliced, including 86 carbohydrate active enzymes, thus the AS rate was estimated to be about 28.6% in *A. fumigatus*. A total of 53,270 RNA transcripts were assembled, based on RNA-seq mapping information, and were functionally annotated. Strain and carbon source specific differentially expressed transcripts were identified. The collected data including RNA-seq mapping information, assembled transcripts, and identified AS events provide a foundation for further investigation of the gene regulations in *A. fumigatus*.

Keywords *Aspergillus fumigatus*; Alternative splicing; Gene expression; RNA-sequencing; Carbon source; Carbohydrate active enzymes; Protein family

1 Introduction

Aspergillus fumigatus is a saprophytic fungus and primarily found in soil and decaying organic materials in nature. It plays an essential role in carbon and nitrogen recycling and is one of the most common *Aspergillus* species causing diseases in humans with immunodeficiency (Nierman et al., 2005). Martin-Vicente et al. (2024) reported that a morphogenetic kinase, named *CotA*, regulates pathogenic growth in response to carbon source diversity in *A. fumigatus*. The wild-type strain produces two *CotA* kinase protein isoforms, long and short, which are required for tissue invasive hyphal morphogenesis. The *cotA* mutant was selected from a protein kinase disruption library, which was generated in the A1163 (CEA10) wild-type genetic background through coupling of CRISPR/Cas9-based gene targeting with a miniaturized protoplast transformation technique (Souza et al., 2021). The *cotA* mutant was severely growth restricted under all de-repressing conditions. Though *cotA* mRNA expression level in the mutant remained to be stable in different carbon sources tested, however, *cotA* mutant could not express the long protein isoform and expressed the short protein isoform only grew in minimal media, containing 1% glucose or 1% sucrose, but not in 1% acetate or ethanol (Martin-Vicente et al. 2024).

Alternative splicing (AS) is a fundamental regulatory process that enables the production of two or more mRNA isoforms from a single intron-containing gene in eukaryotes. This mechanism significantly contributes to transcriptomic and proteomic diversities in eukaryotic organisms, including fungi (Chaudhary et al., 2019; Min et al., 2025). In recent years, there has been growing recognition of AS being a dynamic regulating process of gene expression, enabling rapid adaptation to environmental stimuli - an attribute particularly relevant to fungal pathogens like *A. fumigatus*, since it must navigate complex host environments to establish infection (Martin-Vicente et al., 2024). In eukaryotic genes, the presence of introns and exons allows for the modulation of

gene expression through varied exon inclusion or exclusion during mRNA maturation. The basic forms of AS events include exon skipping (ES), intron retention (IR), alternative donor sites (AltD), alternative acceptor sites (AltA), and mutually exclusive exons (MEX). Since MEX only accounts for a small percentage in plants and fungi, it is often included in “complex” or “other” categories in the analysis. The “complex” category consists of AS events formed by more than one basic event, for example, an event having both AltD and AltA. Among basic types of AS events, ES is the most dominant splicing type in animals, while IR is most common in plants and fungi (Chaudhary et al., 2019; Min et al., 2025).

Recent surveys reveal AS events in fungi ranged from 0.2% in non-pathogenic yeast *Saccharomyces cerevisiae* to 38.44% in *Shiraia bamlusicola* and 38.82% in *Histoplasma capsulatum* (Fang et al., 2020; Liu et al., 2020; Muzafar et al., 2021). Our most recent work by mapping 303 RNA-sequencing (RNA-seq) samples which were collected from multiple published projects to the genome of *A. niger* estimated about 50% of genes alternatively spliced (Min et al., 2025). There is a lack of genome-wide AS analysis in *A. fumigatus*. Since AS might be related to the virulence of fungal pathogens (Grützmann et al., 2014), we carried out a genome-wide identification and analysis of AS events using the RNA-seq data generated by Martin-Vicente et al. (2024), aiming to generate a catalog of genes subjecting to AS in *A. fumigatus*. Further comparing the AS events and expression levels at transcripts level in both the wild type and *cotA* mutant strain in response to different carbon sources is expected to provide a deeper understanding of the regulation of hyphal growth and pathogenesis in this species.

2 Materials and Methods

2.1 *A. fumigatus* genome sequences and RNA-seq datasets

A. fumigatus reference genome sequences with annotation GFF (Gene feature format) file (RefSeq accession:GCF_000002655.1, Genome assembly ASM265v1, strain Af293) 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_000002655.1/) (Niernan et al., 2005). The reason for using this assembly is because it has more detailed annotation information, which is useful for downstream analysis as the reference for this species. The RNA-seq data was down-loaded from the NCBI SRA database (<https://www.ncbi.nlm.nih.gov/sra/docs/srdownload/>) using SRA Toolkit. The RNA-seq data (BioProject: PRJNA1147328) were generated and analyzed for differential expressions of genes in the wild type and *cotA* mutant by Martin-Vicente et al. (2024). The dataset contains a total of 18 samples collected from both the wild type and *cotA* mutant strain growing in minimal media containing one of the following carbon substrates: 1% glucose, 1% potassium acetate, or 1% ethanol, with three replicates, respectively. The details of growth conditions and sample preparation can be found in Martin-Vicente et al. (2024).

2.2 RNA-seq reads mapping to the genome, AS identification and analysis of transcripts expression

The RNA-seq reads were mapped to the reference genome sequences using hisat2 (version 2.2.1) with default parameters (Kim et al., 2019). The RNA-seq alignment bam file together with annotation GFF file were used as input for Cufflinks (v2.2.1) (Trapnell et al., 2010). The transcript GTF (Gene Transfer Format) files generated from each RNA-seq dataset after running Cufflinks were merged using cuffcompare script within the Cufflinks package for each treatment. The GTF files generated from merged RNA-seq GTF files in each treatment were further merged using cuffcompare script for each strain, and lastly, the GTF files from both strains were merged to obtain the final GTF file. The standalone Astalavista (version 3.2) and our in-house scripts were used for AS event classification and genomic loci analysis (Foissac and Sammeth, 2007; Min et al., 2025).

The quantification of transcripts and identification of differentially expressed transcripts between treatments were carried out using cuffdiff script with default parameters, i. e., the Q-value < 0.05 was used as a cutoff. Q-value is the false discovery rate (FDR) adjusted P-value to determine significance. The analysis and generation of Venn diagrams were carried out using InteractiVenn program (Heberle et al., 2015).

2.3 Functional annotation of transcripts and data availability

The sequences of assembled transcripts were retrieved using gtf_to_fasta tool in the TopHat2 package (Kim et al. 2013) based on the GTF file generated by cuffcompare program after merging all GTF files. These transcripts

were functionally annotated including BLASTX against UniProt Swiss-Prot database, protein sequence (open reading frame, ORF) prediction using OrfPredictor, protein family (Pfam) search (E-value $\leq 1e-10$), and comparison with reference gene models (minimum aligned length 50bp and >95% identity) (Min et al., 2005). The transcript sequences, detailed information of AS events, and other supplementary files are available at our bioinformatics site (<http://bioinformatics.ysu.edu/publication/data/Afumigatus/>).

3 Results

3.1 Mapping RNA-seq data to the genome of *A. fumigatus*

RNA-seq data were mapped to *A. fumigatus* genome individually for each sample. The mapping rates varied slightly from 94% - 96% (average 94.7%). About 2.7% of reads were mapped to more than one genomic region. Combining all mapping data together approximately 797.5 million reads out of a total of 841.8 million reads were mapped to the genome. Such high mapping rates of RNA-seq data demonstrated the high quality of the dataset. More detailed mapping data can be found in the thesis for a Master of Science by Adeyemo (2025).

3.2 Identification of AS events

3.2.1 AS in different carbon sources

AS events were identified and classified individually for each sample, the results were described by Adeyemo (2015). Then the mapping data were merged for each carbon source treatment in the wild type and *cotA* mutant for AS identification. The results are shown in Table 1. The types of AS include ES, AltD, AltA, IR, and others. Total AS events varied from 5,799 to 6,397 in different treatments. In all treatments, IR remained to be the most common basic type, accounting for 43.7%~47.0%, followed by AltA (15.5%~16.5%), AltD (5.9%~6.1%), and ES (2.6%~2.8%). In both the wild-type and *cotA* mutant strains, glucose treatment had the lowest number of AS events among three carbon sources. The wild-type strain in glucose had lower number of AS than *cotA* mutant in glucose. The increased AS events in the mutant compared with the wild-type strain and non-glucose treatments compared with glucose treatment may represent the responses of strains coping with the unfavorable conditions including both *cotA* mutation and carbon sources.

Table 1 Types of alternative splicing events in each carbon source in the wild type and *cotA* mutant

	ES (%)	AltD (%)	AltA (%)	IR (%)	Others (%)	Total
Wild type						
Acetate	174 (2.7)	390 (6.1)	1016 (15.9)	3004 (47.0)	1813 (28.3)	6397
Ethanol	168 (2.8)	371 (6.2)	932 (15.5)	2716 (45.1)	1838 (30.5)	6025
Glucose	163 (2.8)	355 (6.1)	920 (15.9)	2596 (44.8)	1765 (30.4)	5799
<i>cotA</i> mutant						
Acetate	162 (2.6)	361 (5.9)	1012 (16.5)	2686 (43.7)	1929 (31.4)	6150
Ethanol	173 (2.8)	382 (6.1)	965 (15.5)	2769 (44.4)	1941 (31.2)	6230
Glucose	164 (2.8)	369 (6.3)	931 (15.8)	2578 (43.7)	1861 (31.5)	5903

We then compared the AS events in three carbon sources in the wild type and *cotA* mutant to examine the common, i. e., constitutive, AS events and treatment specific AS events (Figure 1). There were 1,194 (18.7%~20.6%) and 1,201 (19.5%~20.3%) common AS events among three different carbon sources in the wild-type strain and *cotA* mutant strain, respectively. Carbon treatment specific AS events accounted a large proportion (62.6%~65.7%) of total AS events in all carbon treatments in both strains (Figure 1).

3.2.2 Comparing AS in wild-type strain and *cotA* mutant strain

We compared the common AS events and strain specific AS events between the wild-type strain and *cotA* mutant strain growing in three different carbon sources (Table 2 and data in Supplementary files). There were 1,703 conserved AS events between the wild type and *cotA* mutant, 4,096 the wild type specific and 4,200 *cotA* mutant specific AS events when growing in glucose. There were 1,787 conserved AS events between the wild type and *cotA* mutant, 4,610 the wild type specific and 4,363 *cotA* mutant specific AS events when growing in acetate. A similar trend was observed, there were 1,817 conserved AS events between the wild type and *cotA* mutant, 4,208

the wild type specific and 4,413 *cotA* mutant specific AS events when growing in ethanol. We further merged the data for all carbon sources in the wild-type and mutant strains (Table 2). A total of 27,979 AS events including 341 ES, 1,404 AltD, 3,024 AltA, 10,312 IR, and 12,898 other events were identified from the data of the merged 18 samples. These AS events were identified from a total of 3,610 genomic loci involving 31,374 unique transcripts. The information of these AS events including the genomic sites and functional annotation of transcripts can be found in Supplementary files.

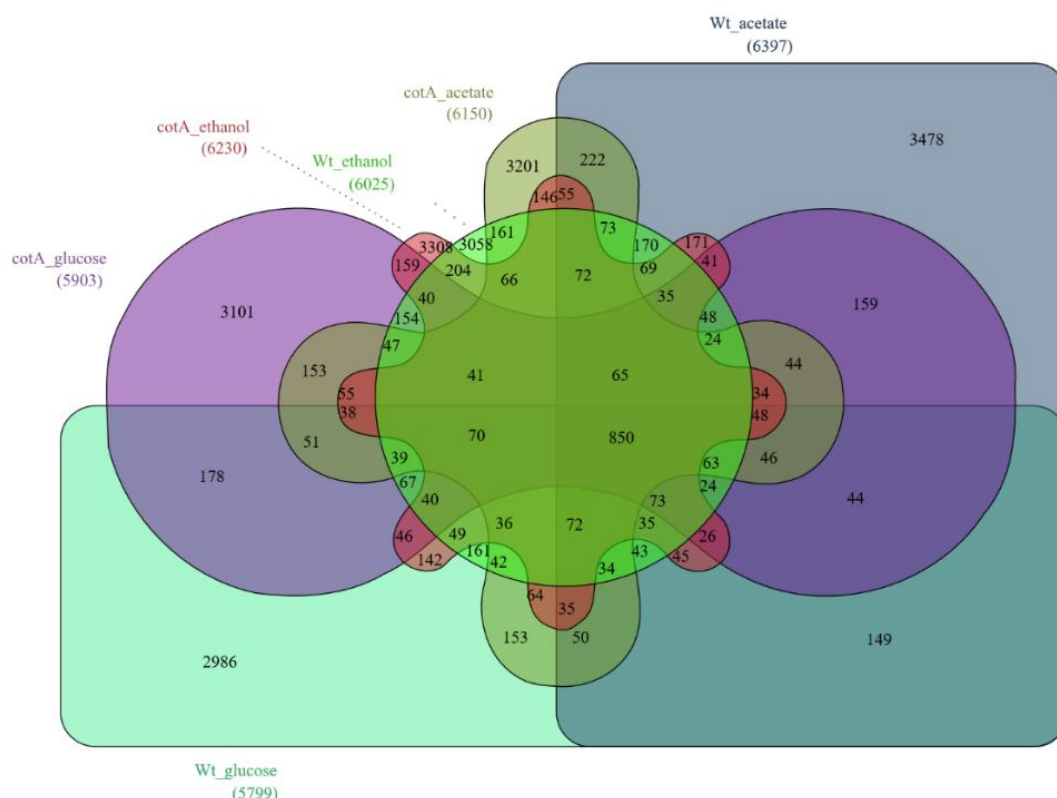


Figure 1 Comparison of alternative splicing events in the wild type and *cotA* mutant strain growing in glucose, acetate or ethanol

Table 2 Types of alternative splicing events in the wild type and *cotA* mutant strain and in the whole project

	ES (%)	AltD (%)	AltA (%)	IR (%)	Others (%)	Total
Wild type	247 (1.6)	828 (5.4)	1961 (12.7)	6549 (42.5)	5810 (37.7)	15395
<i>cotA</i> mutant	239 (1.5)	829 (5.3)	1911 (12.3)	6350 (40.7)	6254 (40.1)	15583
Common events	166 (3.4)	312 (6.4)	891 (18.4)	2419 (49.9)	1058 (21.8)	4846
Wild type specific	81 (0.8)	516 (4.9)	1070 (10.1)	4130 (39.2)	4752 (45.0)	10549
<i>cotA</i> mutant specific	73 (0.7)	517 (4.8)	1020 (9.5)	3931 (36.6)	5196 (48.4)	10737
Wild type and <i>cotA</i> mutant merged	341 (1.2)	1404 (5.0)	3024 (10.8)	10312 (36.9)	12898 (46.1)	27979

3.3 Differential transcripts expression in different carbon sources in the wild type and *cotA* mutant

As our analysis showed AS events were generated from 3,610 genomic loci, comparison of transcripts expression among different treatments including both carbon sources and strains would provide more information of gene regulations in the organism. First, we compared transcripts expressions between the mutant strain and the wild-type strain for each carbon source; then, we compared transcripts expressions among carbon sources within each strain. There were no differentially expressed transcripts found in glucose treatment between two strains. However, comparing the *cotA* mutant with the wild type, there were 15 and 11 transcripts up-regulated, 26 and 19 transcripts down-regulated, in acetate and ethanol treatment, respectively, with only two transcripts (TCONS_00006223, TCONS_00020271) down-regulated commonly found in acetate and ethanol treatment (Table 3; Table 4).

Table 3 Differentially expressed transcripts in *cotA* mutant compared with the wild-type strain in acetate treatment

Transcript ID	Wild type	<i>cotA</i> mutant	log2(fold_change)	Function
TCONS_00020655	0	2.7	+	
TCONS_00020660	0	13.6	+	
TCONS_00061821	0	1.1	+	
TCONS_00033708	9.7	113.6	3.55	
TCONS_00023856	28.9	101.8	1.82	
TCONS_00010041	10.9	37.9	1.79	
TCONS_00061169	30.4	93.9	1.63	
TCONS_00040084	14.5	37.1	1.35	
TCONS_00026253	22.8	55.1	1.27	
TCONS_00038236	3.9	9	1.2	
TCONS_00040987	555.3	1019.3	0.88	
TCONS_00058105	19.7	36.1	0.87	Cyclic dipeptide prenyltransferase
TCONS_00037816	173.5	303.3	0.81	High affinity iron permease ftrA
TCONS_00056047	43.7	74.1	0.76	
TCONS_00054060	61.8	96.5	0.64	
TCONS_00047940	397.6	213.7	-0.9	Solute carrier RCH1
TCONS_00028166	243.2	129.5	-0.91	
TCONS_00058948	115.8	60.7	-0.93	Protein LURP-one-related 15
TCONS_00006223	2131.1	1064	-1	
TCONS_00023783	18.4	8.8	-1.07	
TCONS_00023503	26.9	12.2	-1.13	
TCONS_00059406	44.6	19.2	-1.22	
TCONS_00004586	21.7	8.9	-1.3	
TCONS_00060313	105.1	37.3	-1.5	
TCONS_00017084	136.5	40.3	-1.76	Putative formamidase C869.04
TCONS_00044194	88.4	23.2	-1.93	Putative agmatinase 1
TCONS_00037163	6.4	1.6	-1.99	
TCONS_00058098	23.3	5.8	-2.01	Cytochrome P450 monooxygenase psoD
TCONS_00059309	98.9	23.1	-2.1	
TCONS_00020271	254.8	58.9	-2.11	
TCONS_00061156	53.7	12.2	-2.14	
TCONS_00058095	8.1	1.8	-2.19	PKS-NRPS hybrid synthetase psoA
TCONS_00059307	70	15.2	-2.2	
TCONS_00058100	31.1	6.3	-2.3	Cytochrome P450 monooxygenase psoD
TCONS_00059310	463.2	92.8	-2.32	
TCONS_00061165	150	29.9	-2.33	
TCONS_00059313	101.4	18.9	-2.42	
TCONS_00044195	58.4	10.1	-2.53	Putative agmatinase 1
TCONS_00016922	0.7	0	-	
TCONS_00037241	1.2	0	-	Efflux pump himE
TCONS_00051044	1	0	-	

Note: P-value: 0.00005, Q-value: 0.0438967. +: up-regulation; -: down-regulation

In contrast to the small numbers of differentially expressed transcripts (DETs) in each carbon source between two strains, larger numbers of DETs between carbon sources were found in both strains. Since there was no DET detected in glucose treatments between two strains, glucose treatment was used for identification of DETs in acetate and ethanol treatment. We identified 485 up-regulated and 516 down-regulated DETs in acetate treatment, 386 up-regulated and 424 down-regulated DETs in ethanol treatment in the wild type; and 608 up-regulated and 714 down-regulated DETs in acetate treatment, 449 up-regulated and 475 down-regulated DETs in ethanol treatment in *cotA* mutant (Figure 2). The identifiers, expression values and functional information of these DETs can be found in supplementary files. Both the wild type and *cotA* mutant generated more DETs in responses to different carbon sources than the differences between the two strains for the same carbon source.

Table 4 Differentially expressed transcripts in *cotA* mutant compared with the wild-type strain in ethanol treatment

Transcript ID	Wild type	CotA mutant	log2(fold change)	
TCONS_00022959	7.1	52.6	2.9	Adhesin-like cell surface protein MAD1
TCONS_00041628	165	590.6	1.8	
TCONS_00033707	23.7	65.9	1.5	
TCONS_00003112	2.9	6.5	1.2	
TCONS_00023450	131.2	243.1	0.9	Carboxypeptidase S1 homolog A
TCONS_00001889	336.7	610.7	0.9	
TCONS_00016143	23.7	42.7	0.8	
TCONS_00038422	83	147.6	0.8	
TCONS_00059439	48.1	83.6	0.8	Major facilitator-type transporter ecdD
TCONS_00029283	15.5	26.7	0.8	
TCONS_00059013	377.3	628.9	0.7	Proline dehydrogenase 1
TCONS_00051851	121.2	67.6	-0.8	
TCONS_00005641	48.1	25.7	-0.9	Nitrate transporter
TCONS_00055764	20.8	11.1	-0.9	
TCONS_00003061	249.9	125.1	-1	Zinc finger protein C25B8.19c
TCONS_00012663	22.1	9.8	-1.2	
TCONS_00037784	83.8	34.5	-1.3	Probable thioredoxin reductase ARB_06224
TCONS_00010838	904	365.2	-1.3	
TCONS_00006223	2257.1	911.2	-1.3	Acetate permease A
TCONS_00037785	287	112.8	-1.3	
TCONS_00020271	205.6	63.2	-1.7	Probable thioredoxin reductase ARB_06224
TCONS_00020936	17.4	5.3	-1.7	
TCONS_00042756	10.9	1.6	-2.8	Polyenoic acids biosynthesis gene cluster
TCONS_00058380	127.2	13.3	-3.3	
TCONS_00059588	12	1.2	-3.3	MFS-type transporter criB
TCONS_00059590	172.1	17	-3.3	
TCONS_00026188	11.7	0	-	
TCONS_00060548	8.7	0	-	
TCONS_00000511	2	0	-	
TCONS_00056432	1.9	0	-	

Note: P-value: 0.00005, Q-value: 0.0438967. + or positive value: up-regulation; - or negative value: down-regulation

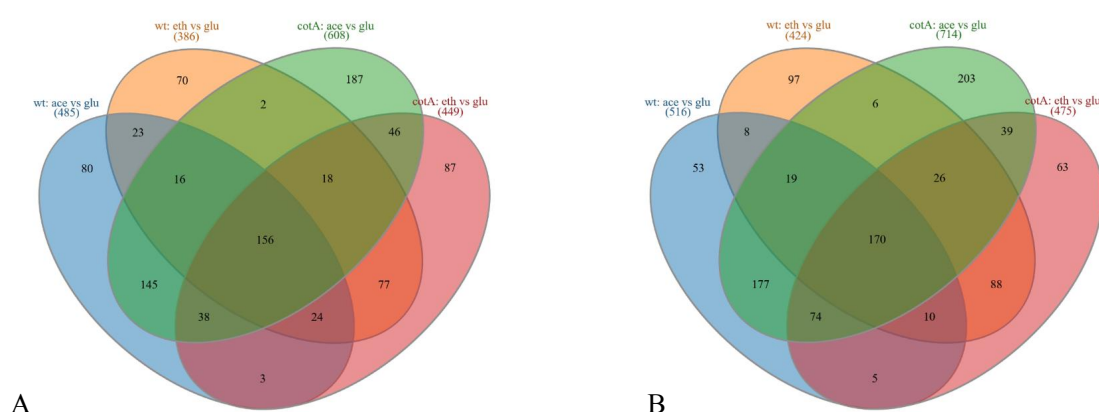


Figure 2 Differentially expressed transcripts in acetate or ethanol treatment compared with glucose treatment in the wild type or *cotA* mutant strain

Note: (A) Up-regulated transcripts; (B) Down-regulated transcripts

3.4 Functional analysis of genes, transcripts, and proteins

Cufflinks tool assembled 8,184 genomic loci, i. e., expressed genes, based on the RNA-seq data mapping to the genome. In the reference genome downloaded from the NCBI genome database, a total of 9,857 genes were annotated. By comparing the locations of reference gene models with genomic loci assembled by Cufflinks, all 9,857 gene models were mapped to 6,130 genomic loci from Cufflinks with some Cufflinks genomic loci having

two or more gene models (Table 5). Among these genomic loci identified by Cufflinks, 4,082 loci with each of them was mapped to single gene models, 2,048 loci each were matched with more than one gene model. One extreme case was locus XLOC-002925 on chromosome 3 (accession NC_007196) mapped with 46 gene models. This “super locus” problem in Cufflinks is caused by data merging using cuffcompare utility to combine multiple samples with increased noises of novel assembled fragments known as “transfrags” as well as incomplete or inaccurate annotation of the reference genome. Nevertheless, it is worthy to further detailed investigation or manual corrections when the data are used for other applications. In addition to the loci mapped to gene models, 2,054 new genomic loci with 5,939 transcripts were identified in the work, these loci represented unannotated genes, thus, their biological significance needs to be further investigated. To avoid the “super locus” problem, we identified the reference gene models harboring altered splice sites in AS transcripts identified above (Table 2). Among 9,857 gene models, 2,823 gene models had two more transcripts generated by AS, thus, the AS rate was estimated to be about 28.6% in the annotated genes.

Table 5 Basic features and functional annotation of assembled RNA transcripts and predicted proteins in *A. fumigatus*

Feature	Numbers
Total genomic loci from Cufflinks	8184
Loci having one transcript	2177 (26.6%)
Loci having more than one transcript	6007 (73.6%)
Cufflinks loci mapped to reference gene model loci	6130 (74.9%; match with 9857 genes)
Newly identified genomic loci	2054 (25.1%)
Total assembled unique transcripts by Cufflinks	57230
Average transcript length (bp)	3174
Cufflinks transcripts match with gene model transcripts	25257 (44.1 %)
BLASTX search against Swiss-Prot dataset	21278 (37.2%)
Total predicted ORFs (M20)	54743 (95.7%)
Average ORF length (amino acids)	260
Total ORFs with a Pfam match	15784 (28.8%)

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

A total of 57,230 unique transcripts with an average length of 3,174 bp were assembled by Cufflinks tool. Among them, 25,257 transcripts matched gene model transcripts and 21,278 transcripts had a protein match against UniProt Swiss-Prot database. A total of 54,743 polypeptides, i. e., open reading frames (ORFs), starting with methionine (Met) with a minimum length of 20 amino acids were predicted and 15,784 (28.8%) predicted polypeptides were matched to the Pfam database (Table 5).

To compare the AS rates of genes encoding different Pfams we compared the Pfam numbers of AS gene models with all protein-coding gene models. While 6,225 of 9,629 (64.6%) proteins of gene models had Pfam matches, 2,189 of 2,823 (77.5%) proteins from AS genes had Pfam matches. As *A. fumigatus* plays an important role as a saprophyte in decaying organic matters in nature, based on the CAZymes classification (<http://www.cazy.org/Home.html>) (Drula et al., 2022), we compared the number of genes encoding carbohydrate-active enzymes (CAZymes) in *A. fumigatus* and *A. niger* (Table 6). AS in CAZymes genes in *A. niger* was reported previously (Min et al. 2025). We processed *A. niger* data with an updated Pfam database (version 35). A total of 251 and 267 genes encoding CAZymes were identified in *A. niger* and *A. fumigatus*, respectively; and among them 123 (49.0%) and 86 (32.2%) were alternatively spliced, respectively (Table 6). These CAZymes in *A. fumigatus* and *A. niger* may be valuable for further investigation for industrial applications in future (Gruben et al., 2017; Contesini et al., 2021; Joshi et al., 2024).

Table 6 List of genes encoding carbohydrate active enzymes in *A. fumigatus* and *A. niger*

Pfam ID	Domain	Domain description	<i>A. niger</i>		<i>A. fumigatus</i>	
			Total	AS	Total	AS
pfam04616	Glyco_hydro_43	Glycosyl hydrolases family 43	10	4	16	6
pfam00704	Glyco_hydro_18	Glycosyl hydrolases family 18	9	6	16	9
pfam01915	Glyco_hydro_3_C	Glycosyl hydrolase family 3 C-terminal	13	8	14	4
pfam00128	Alpha-amylase	Alpha amylase	14	7	12	4
pfam00295	Glyco_hydro_28	Glycosyl hydrolases family 28	16	5	11	3
pfam03663	Glyco_hydro_76	Glycosyl hydrolase family 76	11	6	8	2
pfam03659	Glyco_hydro_71	Glycosyl hydrolase family 71	7	2	8	2
pfam00150	Cellulase	Cellulase (glycosyl hydrolase family 5)	4	4	8	4
pfam01055	Glyco_hydro_31	Glycosyl hydrolases family 31	7	3	7	4
pfam03198	Glyco_hydro_72	Glucanotransferase	6	4	7	3
pfam00982	Glyco_transf_20	Glycosyltransferase family 20	6	4	7	4
pfam03443	Glyco_hydro_61	Glycosyl hydrolase family 61	5	0	7	0
pfam01522	Polysacc_deac_1	Polysaccharide deacetylase	7	2	6	1
pfam00722	Glyco_hydro_16	Glycosyl hydrolases family 16	5	1	6	2
pfam04488	Gly_transf_sug	Glycosyltransferase sugar-binding	9	3	5	0
pfam01532	Glyco_hydro_47	Glycosyl hydrolase family 47	5	3	5	2
pfam13641	Glyco_transf_2_3	Glycosyltransferase like family 2	4	2	5	1
pfam00232	Glyco_hydro_1	Glycosyl hydrolase family 1	3	1	5	1
pfam00723	Glyco_hydro_15	Glycosyl hydrolases family 15	2	1	5	2
pfam07971	Glyco_hydro_92	Glycosyl hydrolase family 92	5	4	4	2
pfam00933	Glyco_hydro_3	Glycosyl hydrolase family 3 N terminal	4	2	4	2
pfam01301	Glyco_hydro_35	Glycosyl hydrolases family 35	4	3	4	2
pfam00251	Glyco_hydro_32N	Glycosyl hydrolases family 32 N-terminal	3	0	4	0
pfam03901	Glyco_transf_22	Alg9-like mannosyltransferase family	3	2	4	3
pfam05686	Glyco_transf_90	Glycosyl transferase family 90	3	0	4	1
pfam00840	Glyco_hydro_7	Glycosyl hydrolase family 7	2	2	4	0
pfam07335	Glyco_hydro_75	Fungal chitinase of glycosyl hydrolase	2	1	4	1
pfam00331	Glyco_hydro_10	Glycosyl hydrolase family 10	1	1	4	0
pfam03033	Glyco_transf_28	Glycosyltransferase family 28 N-terminal	5	5	3	2
pfam00457	Glyco_hydro_11	Glycosyl hydrolases family 11	3	0	3	0
pfam01670	Glyco_hydro_12	Glycosyl hydrolase family 12	3	2	3	0
pfam01793	Glyco_transf_15	Glycolipid 2-alpha-mannosyltransferase	3	2	3	3
pfam07470	Glyco_hydro_88	Glycosyl Hydrolase Family 88	2	1	3	0
pfam13692	Glyco_trans_1_4	Glycosyl transferases group 1	2	2	3	1
pfam00734	CBM_1	Fungal cellulose binding domain	1	1	3	2
Total	-	-	251	123	267	86

Note: only protein families with 3 or more members in *A. fumigatus* are listed in the table. The full list of all protein families can be found in supplementary files

4 Discussion

Fungal species have relatively smaller genomes with less protein coding genes than plant species. However, they can grow and survive in a vastly variable array of environments, such as in soils as saprophytes or in plants and human and animals as pathogens. A number of genome-wide studies on AS in fungi have demonstrated a large proportion of intron containing protein coding genes potentially subjected to AS (Grützmann et al., 2013; Marshall et al., 2013; Xie et al., 2015; Jin et al., 2017; Fang et al., 2020; Ibrahim et al., 2021; Lu et al., 2022; Jeon et al., 2022; Min et al., 2025). Undoubtedly, the increased transcriptome and proteome diversity resulting from AS have provided the capability of fungal species to rapidly adapt and cope with drastic changes of environments including nutrient availability. Our current work represents the first genome-wide analysis of AS in *A. fumigatus*. As in other fungal species by now have been investigated, protein coding genes in *A. fumigatus* are potentially

subjected to AS for growth regulations in coping with changing environments. In current work, the AS rate was estimated to be about 28.6%, even the samples were limited to variations of carbon sources. A much higher AS rate in *A. fumigatus* can be expected when more transcriptome data are integrated in future analysis, as *A. niger* was estimated to have 50% of AS with more RNA-seq data from samples generated from multiple projects were employed for genome mapping (Min et al., 2025).

Identification of carbon source treatment specific AS events in the wild type and *cotA* mutant strain may be helpful for understanding the regulations of genes at transcripts level (Table 1; Table 2). Particularly the list of differentially expressed transcripts identified between the two strains in the same carbon source may be used as potential targets for further investigation of the regulation of hyphal morphogenesis in *A. fumigatus* (Table 3; Table 4). In addition, in *A. fumigatus* genome there are over 200 of genes coding for CAZymes and many of them are subjecting to AS, thus the isoforms of these enzymes need to be verified and experimentally tested for their enzymatic activities for exploring their potential application in biofuel production and bioprocessing (Table 6) (Miao et al., 2015; Joshi et al., 2024). Taken together, the information collected in the work including the assembled transcripts with functional annotation, identified AS events and DETs, and newly identified genomic loci provides a resource for further investigation in understanding the biology of *A. fumigatus*.

Author Contributions

XM and KA designed the experiments. XM provided the methodology and software support. KA carried out RNA-seq data collection and genome mapping. XM and KA analyzed the data and prepared the manuscript. Both authors have read and agreed to the published version of the manuscript.

Acknowledgements

The Ohio Supercomputer Center provided computational resources for part of data processing.

Conflict of Interest Disclosure

The authors affirm that this research was conducted without any commercial or financial relationships that could be construed as a potential conflict of interest.

References

- Adeyemo K. S., 2025, Identification and analysis of alternative splicing of *Aspergillus fumigatus* Thesis for M. S., Youngstown State University, Supervisor: Min X., pp. 25-29.
http://rave.ohiolink.edu/etdc/view?acc_num=ysu1754652601198289.
- Chaudhary S., Khokhar W., Jabre I., Reddy A. S., Byrne L. J., Wilson C. M., and Syed N. H., 2019, Alternative splicing and protein diversity: plants versus animals, *Front. Plant Sci.*, 10: 708.
<https://doi.org/10.3389/fpls.2019.00708>
- Contesini F.J., Frandsen R.J.N. and Damasio A., 2021, Editorial: CAZymes in biorefinery: from genes to application, *Front. Bioeng. Biotechnol.*, 9: 622817.
- Drula E., Garron M.L., Dogan S., Lombard V., Henrissat B., Terrapon N., 2022, The carbohydrate-active enzyme database : functions and literature, *Nucleic Acids Res.*, 50: D571–D577.
- Fang S., Hou X., Qiu K., He R., Feng X., Liang X., 2020, The occurrence and function of alternative splicing in fungi, *Fungal Biol. Rev.*, 34(4): 178-188.
- Foissac S., and Sammeth M., 2007, ASTALAVISTA: dynamic and flexible analysis of alternative splicing events in custom gene datasets, *Nucleic Acids Res.*, 35: W297-299.
<https://doi.org/10.1093/nar/gkm311>
- Gruben B.S., Mäkelä M.R., Kowalczyk J.E., Zhou M., Benoit-Gelber I. and De Vries R.P., 2017, Expression-based clustering of CAZyme-encoding genes of *Aspergillus niger*, *BMC Genomics*, 18(1):900.
- Grützmann K., Szafranski K., Pohl M., Voigt K., Petzold A., Schuster S., 2014, Fungal alternative splicing is associated with multicellular complexity and virulence: a genome-wide multi-species study, *DNA Res.*, 21(1): 27-39.
- Heberle H., Meirelles G.V., da Silva F.R., Telles G.P. and Minghim R., 2015, InteractiVenn: a web-based tool for the analysis of sets through Venn diagrams, *BMC Bioinfo.*, 16:169.
- Ibrahim H.M., Kusch S., Didelon M. and Raffaele S., 2021, Genome - wide alternative splicing profiling in the fungal plant pathogen *Sclerotinia sclerotiorum* during the colonization of diverse host families. *Mol. Plant Pathol.*, 22(1): 31-47.
- Jeon J., Kim K.T., Choi J., Cheong K., Ko J., Choi G., Lee H., Lee G.W., Park S.Y., Kim S. and Kim S.T., 2022, Alternative splicing diversifies the transcriptome and proteome of the rice blast fungus during host infection. *RNA Biol.*, 19(1): 373-386.
- Jin L., Li G., Yu D., Huang W., Cheng C., Liao S., Wu Q. and Zhang Y., 2017, Transcriptome analysis reveals the complexity of alternative splicing regulation in the fungus *Verticillium dahlia*, *BMC Genomics*, 18(1): 130.

- Joshi N., Grewal J., Drewniak L. and Pranaw K., 2024, Bioprospecting CAZymes repertoire of *Aspergillus fumigatus* for eco-friendly value-added transformations of agro-forest biomass, *Biotechnol. Biofuels Bioprod.*, 17(1): 3.
- 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.* 37: 907-915.
<https://doi.org/10.1038/s41587-019-0201-4>
- Kim D., Pertea G., Trapnell C., Pimentel H., Kelley R., and Salzberg S. L., 2013, TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions, *Genome Biol.*, 14: R36.
<https://doi.org/10.1186/gb-2013-14-4-r36>
- Liu X.Y., Fan L., Gao J., Shen X.Y., Hou C.L., 2020, Global identification of alternative splicing in *Shiraia bambusicola* and analysis of its regulation in hypocrellin biosynthesis, *Appl. Microbiol. Biotechnol.*, 104: 211-223.
- Lu P., Chen D., Qi Z., Wang H., Chen Y., Wang Q., Jiang C., Xu J.R. and Liu H., 2022, Landscape and regulation of alternative splicing and alternative polyadenylation in a plant pathogenic fungus, *New Phytologist*, 235(2): 674-689.
- Marshall A.N., Montealegre M.C., Jiménez-López C., Lorenz M.C. and van Hoof A., 2013, Alternative splicing and subfunctionalization generates functional diversity in fungal proteomes, *PLoS Genet.* 9(3): e1003376.
- Martin-Vicente A., Souza A.C.O., Guruceaga X., Thorn H.I., Xie J., Nywening A.V., Ge W. and Fortwendel J.R., 2024, A conserved fungal morphogenetic kinase regulates pathogenic growth in response to carbon source diversity, *Nat. Commun.*, 15(1):8945.
- Miao Y., Liu D., Li G., Li P., Xu Y., Shen Q. and Zhang R., 2015, Genome-wide transcriptomic analysis of a superior biomass-degrading strain of *A. fumigatus* revealed active lignocellulose-degrading genes, *BMC Genomics*, 16(1): 459.
- Min X., Jones C., Logan B., Campean M., Wekhyan B., Dasana C., Patel A., Mudiyansele P., Adeyemo K. and Yu, F., 2025, Genome-wide identification and analysis of alternative splicing in *Aspergillus niger*. *Comput. Mol. Biol.*, 15(1):1-12.
- Min X.J., Butler G., Storms R., Tsang A., 2005, OrfPredictor: predicting protein-coding regions in EST-derived sequences, *Nucleic Acids Res.*, 33: W677-680.
<https://doi.org/10.1093/nar/gki394>
- Muzafar S., Sharma R.D., Chauhan N. and Prasad R., 2021, Intron distribution and emerging role of alternative splicing in fungi, *FEMS Microbiol. Lett.*, 368(19):fnab135.
- Nierman W.C., Pain A., Anderson M.J., Wortman J.R., Kim H.S., Arroyo J., Berriman M., Abe K., Archer D.B., Bermejo C., Bennett J., Bowyer P., Chen D., Collins M., Coulson R., Davies R., Dyer P.S., Farman M., Fedorova N., Fedorova N., Feldblyum T.V., Fischer R., Fosker N., Fraser A., García J.L., García M.J., Goble A., Goldman G.H., Gomi K., Griffith-Jones S., Gwilliam R., Haas B., Haas H., Harris D., Horiuchi H., Huang J., Humphray S., Jiménez J., Keller N., Khouri H., Kitamoto K., Kobayashi T., Konzack S., Kulkarni R., Kumagai T., Lafon A., Latgé J.P., Li W., Lord A., Lu C., Majoros W.H., May G.S., Miller B.L., Mohamoud Y., Molina M., Monod M., Mouyna I., Mulligan S., Murphy L., O'Neil S., Paulsen I., Peñalva M.A., Pertea M., Price C., Pritchard B.L., Quail M.A., Rabinowitsch E., Rawlins N., Rajandream M.A., Reichard U., Renauld H., Robson G.D., Rodríguez de Córdoba S., Rodríguez-Peña J.M., Ronning C.M., Rutter S., Salzberg S.L., Sanchez M., Sánchez-Ferrero J.C., Saunders D., Seeger K., Squares R., Squares S., Takeuchi M., Tekaia F., Turner G., Vazquez de Aldana C.R., Weidman J., White O., Woodward J., Yu J.H., Fraser C., Galagan J.E., Asai K., Machida M., Hall N., Barrell B., and Denning D.W., 2005, Genomic sequence of the pathogenic and allergenic filamentous fungus *Aspergillus fumigatus*, *Nature*, 438(7071):1151-1156.
- Souza A.C.O., Martin-Vicente A., Nywening, A.V., Ge W., Lowes D.J., Peters B.M. and Fortwendel J.R., 2021, Loss of Septation Initiation Network (SIN) kinases blocks tissue invasion and unlocks echinocandin cidal activity against *Aspergillus fumigatus*, *PLoS Pathog.*, 17(8):e1009806.
<https://doi.org/10.1371/journal.ppat.1009806>
- Trapnell C., Williams B.A., Pertea G., Mortazavi A., Kwan G., Van Baren M. J., Salzberg S. L., Wold B. J., and Pachter L., 2010, Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation, *Nat. Biotechnol.*, 28: 511-515.
<https://doi.org/10.1038/nbt.1621>
- Xie B.B., Li D., Shi W.L., Qin Q.L., Wang X.W., Rong J.C., Sun C.Y., Huang F., Zhang X.Y., Dong X.W. and Chen X.L., 2015, Deep RNA sequencing reveals a high frequency of alternative splicing events in the fungus *Trichoderma longibrachiatum*, *BMC Genomics*, 16(1): 54.

Disclaimer/Publisher's Note

The statements, opinions, and data contained in all publications are solely those of the individual authors and contributors and do not represent the views of the publishing house and/or its editors. The publisher and/or its editors disclaim all responsibility for any harm or damage to persons or property that may result from the application of ideas, methods, instructions, or products discussed in the content. Publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.