Improving Gene Model Accuracy for Genes Involved in Capsule Formation of Fungal Pathogen *Cryptococcus neoformans*

by

Chase Burton

A thesis presented to the Honors College of Middle Tennessee State University in partial fulfillment of the requirements for graduation from the University Honors College

Fall 2020

Improving Gene Model Accuracy for Genes Involved in Capsule Formation of Fungal Pathogen *Cryptococcus neoformans*

by Chase Burton

APPROVED:

_____

Dr. Rebecca Seipelt-Thiemann, Thesis Director
Professor, Biology Department

Dr. Stephen Wright, Second Reader
Professor, Biology Department

_____

Dr. Dennis Mullen, Thesis Committee Chair
Chair, Biology Department

ACKNOWLEDGEMENTS

I want to thank Dr. Rebecca Seipelt-Thiemann for her mentorship and guidance over the past years and for allowing me to be on her research team. I am thankful for the opportunity to be able to conduct the research that is the subject of this thesis. I would also like to thank my family for their encouragement and support each day.

ABSTRACT

Cryptococcus neoformans is a fungal pathogen that causes disease in immunocompromised individuals. One factor that contributes to the virulence of *C. neoformans* is its large capsule, preventing the pathogen from being destroyed by the human immune system. At least eight genes are related to capsule formation in *C. neoformans*. This study aimed to improve the accuracy of capsule formation gene models based on experimentally derived transcript evidence. Gene model accuracy improvement was conducted using a collaborative annotation platform by comparing transcriptome data to the original computational gene models to generate a super-transcript of each gene. Each gene model required reannotations and the models were annotated in GenSAS and are available for future research, such as *in silico* reconstruction of RNA and protein isoforms to provide insight into the protein functions.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

INTRODUCTION

Infectious diseases commonly result from the pathogenic characteristics of microorganisms. Pathogenic microorganisms evolve specific qualities, allowing them to survive, reproduce, and cause disease in host organisms, known as virulence (Alspaugh 2015). *Cryptococcus neoformans* is a fungal pathogen that is virulent in humans (Alspaugh 2015; Nielsen et al. 2005) and can be acquired from the environment (Karkowska-Kuleta et al. 2009), primarily through inhalation of soil contaminated with bird feces (Mada and Alam 2019). There are three variants of *C. neoformans*: *grubii*, *gatti*, and *neoformans* (Nielson et al. 2005). Variant *grubii* is the most pathogenic of the three variants (Nielson et al. 2005). This pathogen is the cause of cryptococcal meningitis, which is an infection of the central nervous system and causes meningoencephalitis, resulting in death if left untreated (Nielson et al. 2005; Odom 1997; Chang and Chung 1994). Although *C. neoformans* is life-threatening, it is predominantly found to affect immunocompromised individuals, such as those living with AIDS (Alspaugh 2015; Nielson et al. 2005; Odom 1997; Chang and Chung 1994).

The virulence of *C. neoformans* is due to several phenotypic factors, including capsule formation, melanin production (Nielson et al. 2005; Karkowska-Kuleta et al. 2009), and phospholipase activity (Alspaugh 2015). The large capsule is composed of polysaccharides and protects *C. neoformans* from its environment (Karkowska-Kuleta et al. 2009); such as changes in pH, carbon dioxide concentration, or iron levels (Alspaugh 2015). It also protects the fungus from lysis when phagocytized by macrophages as part of the normal human immune response (del Poeta 2004). *C. neoformans* replicates inside

the phagocytic compartment and can either be extruded or result in the destruction of the macrophage and disbursement of the pathogen in the lungs (del Poeta 2004).

Several genes contribute to *C. neoformans* capsule formation (see Table 1), most notably *CAP64* and *CAP59*, which, when removed from its genome, result in the absence of capsule formation within the organism (Karkowska-Kuleta et al. 2009; Chang and Chung 1994; Chang et al. 1996; O'Meara and Alspaugh 2012; Buchanan and Murphy 1998). *CAP10*, *CAP59*, *CAP60*, and *CAP64* are important in the synthesis of glucuronoxylomannans (GXMs), which compose the polysaccharides within the capsule (Janbon et al. 2001). *CAP59* and *CMT1* encode for α-1,3-mannosyltransferase, which performs mannosylation, contributing to the capsule's carbohydrate backbone (O'Meara and Alspaugh 2012; Zaragoza et al. 2009; Grijpstra et al. 2009). *CAP10* is involved in the addition of β-1,3-linked xylose to the capsule and regulates the overall change in the capsule's structure (O'Meara and Alspaugh 2012; Grijpstra et al. 2009). *CAS1* and *CAS3* code for O-acetyltransferase, which is necessary for the formation of O-acetyl residues on GXM structures within the capsule (Janbon et al. 2001; Moyrand et al. 2004). *PKA1* produces enzymatic subunits essential in the cAMP/PKA pathway within *C. neoformans*, and the deletion of the gene results in capsule absence and loss of virulence (Bose et al. 2003).

Despite growing knowledge about the biology of this pathogenic fungus, few genomic-level resources, other than a sequenced genome and computationally predicted gene set, are currently available. The genome of *C. neoformans* mostly remains unresearched, allowing for unrefined gene structures and models to display inaccuracies

when accessed. These models are predicted through the Program to Assemble Spliced Alignments (PASA) algorithm based on genome and RNA transcriptome data uploaded to the Genome Sequence Annotation Server (GenSAS). The PASA algorithm generates gene transcript models that have variable accuracy. By comparing RNA transcriptome data with the PASA-generated gene transcript models, more accurate gene transcript models can be created. Improved gene transcript models allow for improved opportunities to translate the RNA transcripts, determine differences in proteins, and predict functional changes to proteins based on domain functions.

This study aimed to use a newly acquired computational analysis platform, GenSAS, with existing RNA transcriptome data and an improved gene model prediction algorithm, PASA, to improve the original computationally predicted gene models, beginning with the genes associated with the critical virulence factor of capsule formation.

MATERIALS AND METHODS

C. neoformans Genome and Capsule-Forming Genes

The genome sequence for *C. neoformans* var. *grubii* H99 (GCA_000149245) and its genome annotation were obtained from the Ensembl Fungi, a genomic database specifically for fungi (Howe et al. 2019). The genomic coordinates for eight genes associated with virulence through capsule formation within *C. neoformans*, for the focus of this research, were identified using the genomic database FungiDB (Basenko et al. 2018) and used to view the gene structures within the GenSAS structural genome annotation platform (Table 1) (Humann et al. 2019).

Table 1. Summary of Eight Genes Involved in Capsule Formation in *C. neoformans*. Each of the eight genes is listed, along with their respective CNAG identification number, function, and location coordinates for locating the genes within Apollo. The CNAG identification number and location coordinates for the genes represented were sourced from the FungiDB genomic database (Basenko et al. 2018).

Gene	CNAG Identification	Function	Coordinates	Source
CAP64	CNAG_02885	Encodes for a capsular associated protein important in the synthesis of glucuronoxylomannans (GXMs), which compose the polysaccharides within the capsule (Janbon et al. 2001).	Chromosome 3, 530,546..532,814(-)	FungiDB

Table 1. Summary of Eight Genes Involved in Capsule Formation in *C. neoformans* *(Cont'd)*

Gene	CNAG Identification	Function	Coordinates	Source
CAP59	CNAG_00721	Encodes for α-1,3-mannosyltransferase, which performs mannosylation contributing to the carbohydrate backbone (O'Meara and Alspaugh 2012; Zaragoza et al. 2009; Grijpstra et al. 2009). *CAP59* is important in the synthesis of GXMs, which compose the polysaccharides within the capsule (Janbon et al. 2001).	Chromosome 1, 1,878,927..1,881,469(+)	FungiDB
CAP10	CNAG_07554	Encodes for a capsular associated protein involved in the addition of β-1,3-linked xylose to the capsule and regulates the overall change in the capsule's structure (O'Meara and Alspaugh 2012; Grijpstra et al. 2009). *CAP10* is important in the synthesis of GXMs, which compose the polysaccharides within the capsule (Janbon et al. 2001).	Chromosome 3, 1,190,190..1,192,419(-)	FungiDB
CAP60	CNAG_00600	Encodes for a capsular associated protein important in the synthesis of GXMs, which composes the polysaccharides within the capsule (Janbon et al. 2001).	Chromosome 1, 1,544,106..1,546,340(-)	FungiDB
CMT1	CNAG_03158	Encodes for α-1,3-mannosyltransferase, which performs mannosylation contributing to the carbohydrate backbone (O'Meara and Alspaugh 2012; Zaragoza et al. 2009; Grijpstra et al. 2009).	Chromosome 8, 1:211,670..214,236(+)	FungiDB

Table 1. Summary of Eight Genes Involved in Capsule Formation in *C. neoformans (Cont'd)*

Gene	CNAG Identification	Function	Coordinates	Source
CAS1	CNAG_07937	Encodes for O-acetyltransferase, which is required for GXM O-acetylation in the capsule (Janbon et al. 2001; Moyrand et al. 2004).	Chromosome 13, 452,065..455,631(-)	FungiDB
CAS3	CNAG_03644	Encodes for O-acetyltransferase, which is required for GXM O-acetylation in the capsule (Janbon et al. 2001; Moyrand et al. 2004).	Chromosome 2, 1:423,088..425,925(+)	FungiDB
PKA1	CNAG_00396	Encodes for PKA catabolic subunits essential in the PKA/cAMP pathway (Bose et al. 2003).	Chromosome 1, 1,039,957..1,042,941(-)	FungiDB

Transcriptome RNA Sequencing Data

Transcriptome RNA sequencing data from previous research were also added to the GenSAS to provide evidence for the accuracy of the computationally predicted gene structures in the version 1.0 genome annotation. These data were uploaded to the GenSAS as ".bam" and ".bai" files from Cyverse using Cyberduck and BAM importer (Humann et al. 2019). The transcriptome RNA sequencing data are the results of three previously conducted experiments, each composed of three biological replicates, including seven clinical isolates and one reference strain (E. McClelland, personal communication, March 18, 2019), H99S reference strain treated with estrogen, testosterone, or ethanol (J. Tucker, personal communication, March 18, 2019), and the

H99S reference strain treated with scytovirin and/or testosterone (R. McFeeters, personal communication, March 18, 2019) (Table 2).

Table 2. Transcriptome RNA Sequencing Evidence Data. Presented are the names of the strains used and their respective treatments, descriptions, and sources.

Strain	Treatment	Description	Source
B15	N/A	Clinical Isolate Patient: Botswana Male	E. McClelland, personal communication, March 18, 2019
B18	N/A	Clinical Isolate Patient: Botswana Female	E. McClelland, personal communication, March 18, 2019
B30	N/A	Clinical Isolate Patient: Botswana Female	E. McClelland, personal communication, March 18, 2019
B40	N/A	Clinical Isolate Patient: Botswana Male	E. McClelland, personal communication, March 18, 2019
B43	N/A	Clinical Isolate Patient: Botswana Female	E. McClelland, personal communication, March 18, 2019
B45	N/A	Clinical Isolate Patient: Botswana Male	E. McClelland, personal communication, March 18, 2019
B58	N/A	Clinical Isolate Patient: Botswana Male	E. McClelland, personal communication, March 18, 2019

Table 2. Transcriptome RNA Sequencing Evidence Data

Strain	Treatment	Description	Source
H99S	N/A	Clinical Isolate Reference Strain (H99S_Clinical)	E. McClelland, personal communication, March 18, 2019
H99S_RM	N/A	Control Treatment Time: 4 h	R. McFeeters, personal communication, March 18, 2019
H99S Scytovirin	100 µg scytovirin	Treatment Time: 4 h (H99S_Sc100_RM)	R. McFeeters, personal communication, March 18, 2019
H99S Scytovirin and Testosterone	100 µg scytovirin and testosterone	Treatment Time: 4 h (H99S_Sc100_T_RM)	R. McFeeters, personal communication, March 18, 2019
H99S Testosterone	Testosterone	(H99S_T_JT)	J. Tucker, personal communication, March 18, 2019
H99S Estrogen	Estrogen	(H99S_E_JT)	J. Tucker, personal communication, March 18, 2019
H99S Ethanol	Ethanol	(H99S_v_JT)	J. Tucker, personal communication, March 18, 2019

GenSAS Genome Annotation Platform and Apollo

The Genome Sequence Annotation Server v6.0 (GenSAS v6.0) is the genome annotation platform used in accessing Apollo, the workspace to view and annotate gene transcript models (Humann et al. 2019). Using the RNA transcriptome sequencing data, along with the new Program to Assemble Spliced Alignments (PASA) algorithm, PASA Refinement_Cn01, used to improve computationally predicted genes by combining expressed sequencing tags and complementary DNA (cDNA) at points of alignment

within the genome (Haas et al. 2003), annotations were made within the Apollo user-created workspace within GenSAS. By uploading genome sequences, users can visualize and edit gene models and structures by placing models of interest within the Apollo workspace and compare RNA transcriptome data easily within the GenSAS user interface (Figure 1) (Humann et al. 2019). This workspace allows for collaboration among users in gene model reannotation and publication once manual curation is completed (Humann et al. 2019).

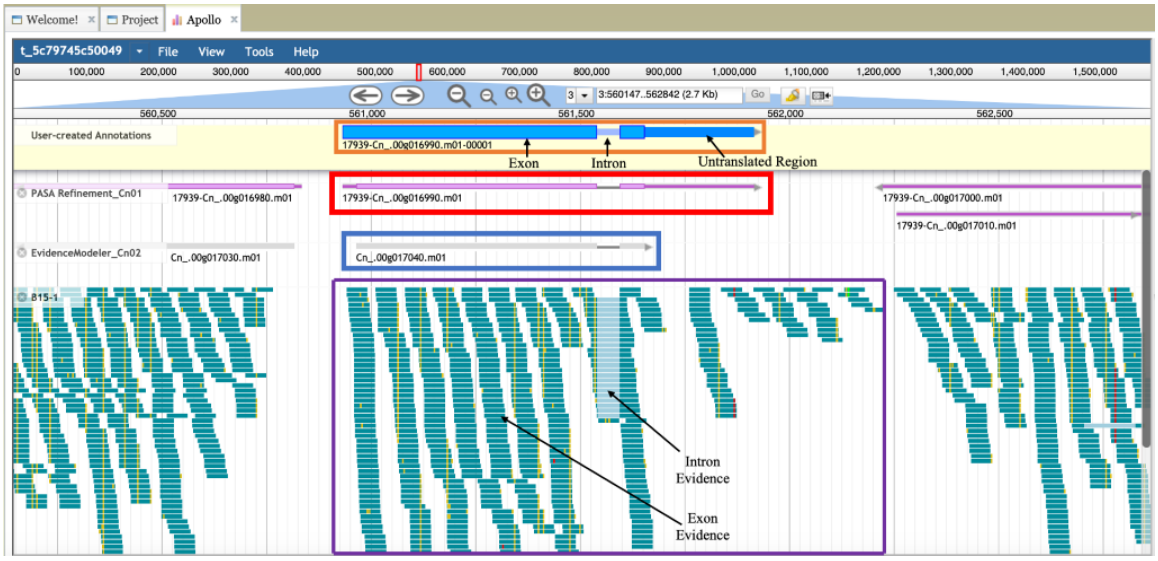

Figure 1. The Apollo Genome Annotation Workspace as Viewed by GenSAS. The orange rectangle contains the PASA Refinement_Cn01 gene transcript model that is present in the red rectangle but added to the Apollo user-created annotation workspace. The exons, intron, and untranslated regions are identified on the gene transcript model within the orange rectangle. The gene transcript model within the blue rectangle is the original *C. neoformans* var. *grubii* H99 genome annotation uploaded into GenSAS, called EvidenceModeler_Cn02. Within the purple rectangle are the RNA transcriptome data that contain exon and intron evidence. The RNA transcriptome data presented is B18-1. As labeled, the intron evidence is represented in GenSAS by light green, while the exon evidence is represented by dark green.

Reannotated Gene Transcript Models and Super-Transcripts

After each of the eight genes and their respective gene transcript models was reannotated for improved accuracy to further support the transcriptome RNA sequencing evidence tracks, PASA gene transcript models, and EvidenceModeler_Cn02, then the transcripts are compiled into super-transcripts. These super-transcripts contain a compilation of all reannotated gene transcript models made for a single gene in one gene transcript model. This allows for observing all possible mature transcripts produced from the gene of interest in a single gene transcript model.

RESULTS

The goal of this project was to use gene expression evidence to correct gene structures (annotations) for each of the eight genes involved in the virulence-causing capsule formation in *Cryptococcus neoformans* within the Apollo workspace in the Genome Sequence Activation Server (GenSAS) (Table 1).

Regions of Interest in GenSAS and Apollo

The GenSAS interface with Apollo workspace is a collaborative web-based platform for analyzing and manipulating gene transcript models (Figure 1) (Humann et al. 2019). In addition to observing the transcriptome RNA sequencing evidence tracks, annotations were created for each gene transcript models present within the Apollo workspace by comparing inconsistencies in gene structure of the PASA gene transcript model (Figure 2c) added to the Apollo workspace (Figure 2b), the gene transcript model proposed by FungiDB (Figure 2a), and the original genomic data for *C. neoformans* var. *grubii* H99 stored within GenSAS (Figure 2d). New gene transcript models were created using the PASA gene transcript models as a starting template with annotations added to the models based on supporting evidence (Figure 2e). Reasons for reannotation of PASA gene transcript models may include the skipping of one or more exons, the retention of one or more introns, the lengthening or shortening of the size of one or more exons, and the lengthening or shortening of the size of one or both untranslated regions represented in the gene transcript models located on the 5’ and 3’ ends (Figure 3). An example comparison of the original and reannotated models for one gene shows intron retention (Figure 4).

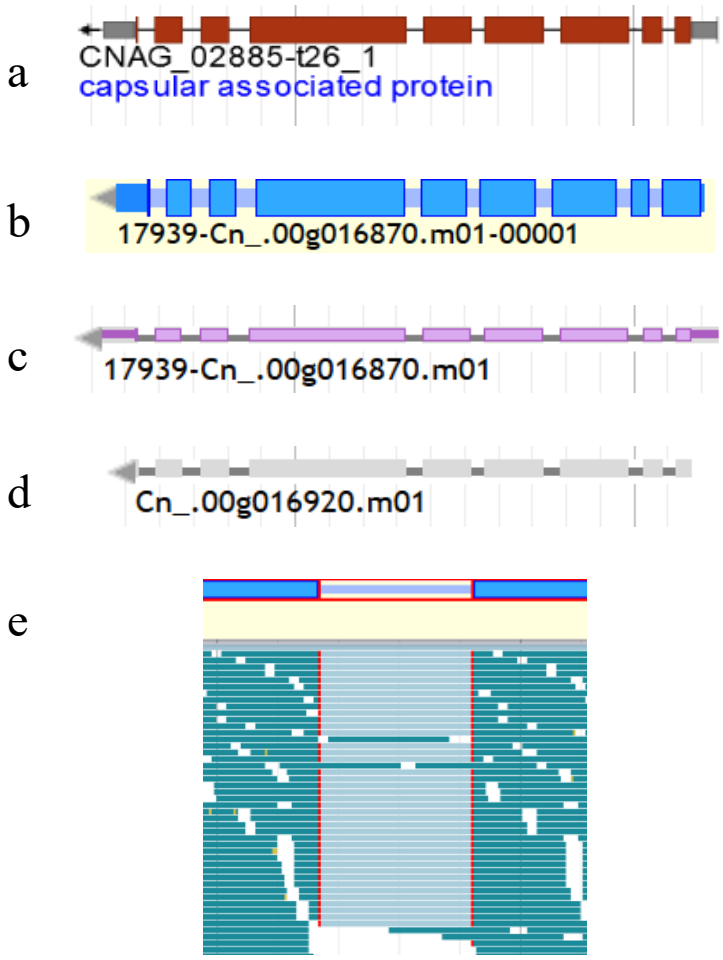

Figure 2. Example Gene Structures as Viewed by Multiple Programs within FungiDB and GenSAS. (a) The structure of a gene as proposed by the FungiDB genomic database. (b) The model of the structure of the gene proposed by the PASA algorithm inserted into the Apollo workspace within GenSAS. (c) The gene structure proposed from the PASA computationally generated algorithm. (d) The gene structure proposed by the original *C. neoformans* var. *grubii* H99 genomic data populated into GenSAS. (e) Intron evidence (light green) and exon evidence (dark green) provided by the previously uploaded RNA transcriptome data, used to determine necessary manipulations to the genes (blue boxes and line) within the Apollo workspace (yellow).

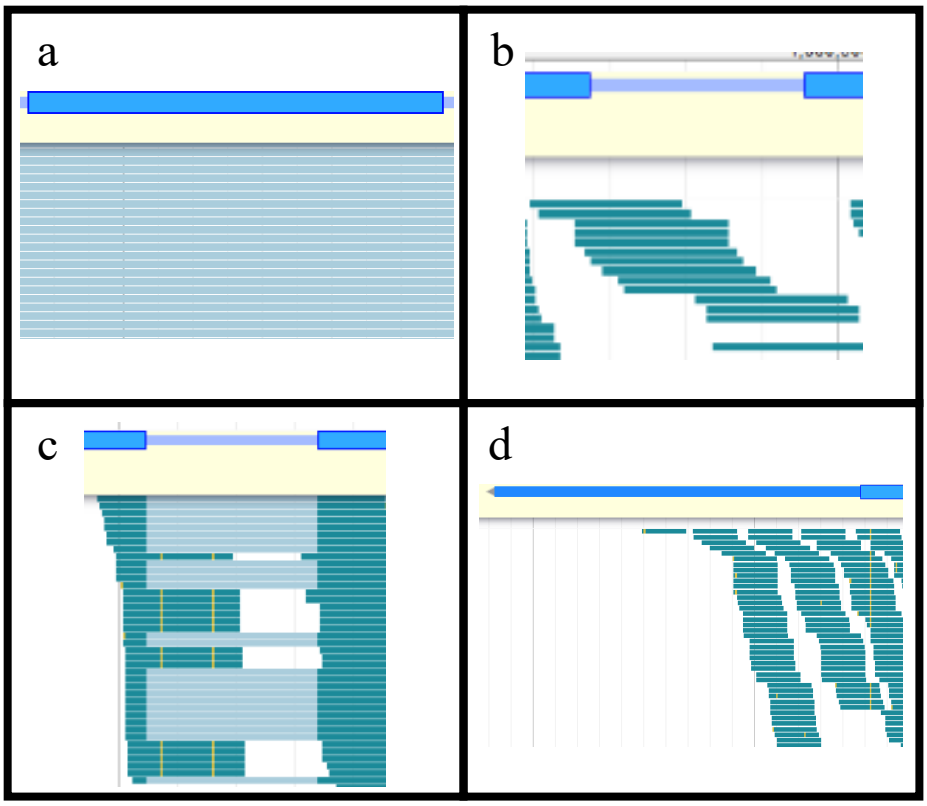

Figure 3. Reasons to Reannotate Gene Structure in GenSAS. (a) Exon skipping is shown, where the exon within the gene transcript model (blue box) is not supported by the RNA transcriptome data due to the presence of only or mostly intron evidence (light green). (b) Intron retention is presented, supported by the exon evidence in the RNA transcriptome data (dark green) beneath the intron (blue line) of the gene structure model. (c) Exon size alteration, in this case, exon lengthening, is supported by substantial exon evidence (dark green) within the space of the proposed intron suggested by the gene transcript model (blue line). The untranslated region of the gene transcript model (thin blue box) is shown, with exon (dark green) RNA transcriptome data suggesting the shortening of the untranslated region, consequently lengthening the exon in the gene transcript model (blue box).

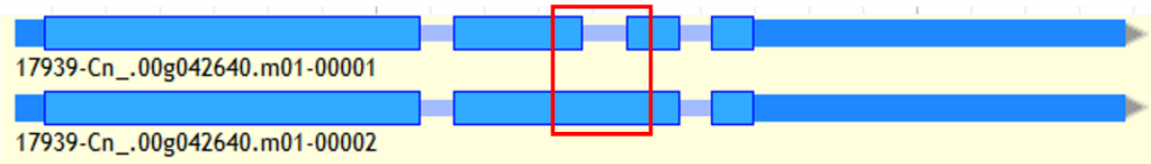

Figure 4. Example Gene Annotation. The red rectangle presents the annotation made between the two gene transcript models. The first gene transcript model is the model proposed by PASA that was placed in the Apollo workspace. The second gene transcript model is the same as the first model but with the retention of intron 2, suggested by the transcriptome RNA sequencing data.

General Overview of Processes

When comparing the gene transcript models from FungiDB, the PASA algorithm, and the original *C. neoformans* var. *grubii* H99S genome (Figure 2 a, c, and d) for each gene, the untranslated regions were present in the FungiDB and PASA gene transcript models, but not in the gene transcript model presented by the original. The RNA transcriptome data (Table 2) were used to identify areas to reannotate and improve the accuracy of gene transcript models proposed by the refined PASA algorithm. RNA transcriptome data suggested reannotations for each capsule-forming virulence-causing gene in *C. neoformans* var *grubii* H99S (Table 3). Based on the RNA transcriptome data, gene transcript models were created for each gene. Using the gene transcript models created based on reannotation suggestions from each data set composed of the RNA transcriptome data, super-transcript gene models were created to compile all the gene transcript models into one model for each gene.

Table 3. Summary of Gene Reannotations. This table shows the genes, inconsistencies, exon coordinates of the reannotations, and RNA transcriptome data tracks that suggest the reannotations.

Gene	Inconsistency	Exon Coordinates	Data Track
CAP64	Retention of intron 3	Exon 3b: 3:531948-532484 (-) strand	B58-3, H99S_A-1, H99S_Sc100_RM2
CAP64	Retention of introns 3 and 7	Exon 3b: 3:531948-532484 (-) strand Exon 7b: 3:530734-531011 (-) strand	B43-2, B58-1
CAP59	Retention of introns 2 and 4	Exon 2b: 1:1879303-1880204 (+) strand Exon 4b: 1:1880286-1880617 (+) strand	B15-1, B18-2
CAP59	Retention of intron 4	Exon 4b: 1:1880286-1880617 (+) strand	B30-2, B30-3, B43-2, B43-3, H99S_A-1
CAP59	Retention of introns 4 and 5	Exon 4c: 1:1880286-1880961 (+) strand	B18-1
CAP59	Retention of introns 1 and 4	Exon 1b: 1:1878927-1879320 (+) strand Exon 4b: 1:1880286-1880617 (+) strand	B58-2
CAP59	Retention of intron 5	Exon 5b: 1:1880489-1880961 (+) strand	H99S_A-2, H99S_Sc100_T_RM2
CMT1	Retention of intron 2	Exon 2b: 8:212681-213204 (+) strand	B58-3, H99S_Sc100_T_RM2
CAP10	Increase in length of exon 4	Exon 4b: 3:1190190-1190360 (-) strand	B15-1, B15-2, B15-3, B18-1, B18-2, B30-2, B45-3, B58-2, H99S_A-2
CAP60	Retention of intron 1	Exon 1b: 1:1544414-1546340 (-) strand	B43-1, B43-2, B43-3
CAP60	Retention of introns 1 and 2	Exon 3c: 1:1544106-1546340 (-) strand	B30-2, B30-3, B58-1
CAS1	Retention of intron 6	Exon 6b: 13:453616-453782 (-) strand	B40-1
CAS1	Retention of intron 8	Exon 8b: 13:452065-453569 (-) strand	B30-3, B43-1, B58-1

Table 3. Summary of Gene Reannotations (Cont'd)

Gene	Inconsistency	Exon Coordinates	Data Track
CAS1	Retention of introns 7 and 8	Exon 7b: 13:452065-453633 (-) strand	B43-3
CAS3	Retention of intron 5	Exon 5b: 2:424361-424540 (+) strand	B43-1, B43-3, B45-1
CAS3	Retention of introns 5 and 11	Exon 5b: 2:424361-424540 (+) strand Exon 11b: 2:425432-425925 (+) strand	B58-1
CAS3	Retention of introns 1, 5, 8, 9, and 11; increase in the length of exon 12	Exon 1b: 2:423088-423934 (+) strand Exon 5b: 2:424361-424540 (+) strand Exon 8b: 2:424790-425376 (+) strand Exon 11c: 2:425432-425925 (+) strand	B18-1
PKA1	Retention of intron 3	Exon 3b: 1:1041613-1042105 (-) strand	B30-3
PKA1	Retention of introns 7 and 9	Exon 7b: 1:1040800-1040994 (-) strand Exon 9b: 1:1039957-1040749 (-) strand	B30-2, B40-1
PKA1	Retention of intron 9	Exon 9b: 1:1039957-1040749 (-) strand	B40-2, B58-1
PKA1	Increase in length of exon 10	Exon 10b: 1:1039957- 1042105 (-) strand	B15-2, B18-1, B30-2, B30-3, B58-2

CAP64

When viewing the FungiDB, PASA, and original genome transcript models for *CAP64*, each gene model was similar, and no differences were found (Figure 5). *CAP64* required two sets of reannotations for the PASA gene transcript models. The RNA transcriptome evidence presented inconsistencies in which set one presented retention of intron 3 (Figure 6) and set two presented retention of introns 3 and 7 occurring simultaneously (Figure 7). For the first set of *CAP64* reannotations, the PASA gene transcript model was modified to display the retention of intron 3 (Figure 8). The PASA

gene transcript model for the second set of *CAP64* reannotations was modified to show the simultaneous retention of introns 3 and 7 (Figure 9). The super-transcript gene model for *CAP64* compiles each of the reannotations viewed in Figures 8 and 9 (Figure 10).

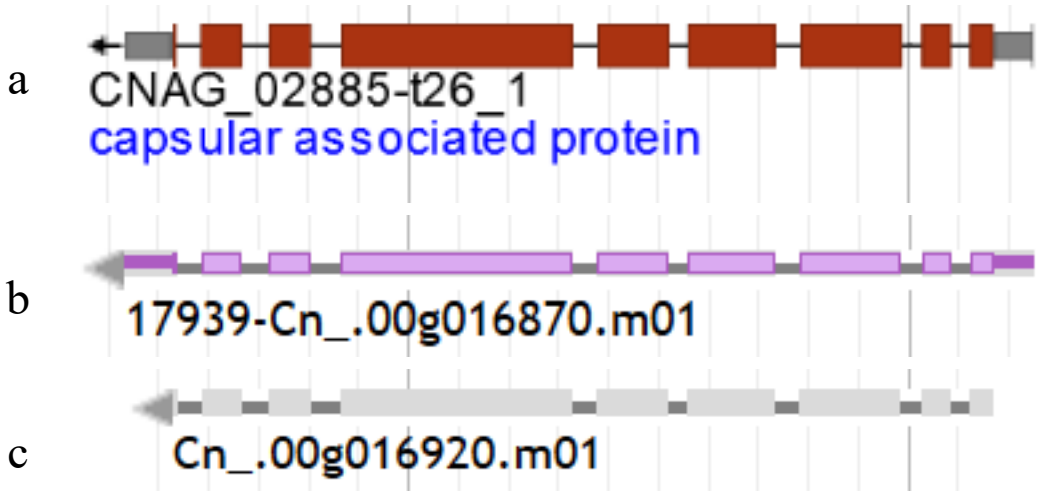

Figure 5. Reference Gene Transcript Models for *CAP64*. (a) The *CAP64* gene transcript model presented from FungiDB. (b) The gene transcript model generated by the PASA algorithm. (c) The original *C. neoformans* genome gene transcript model.

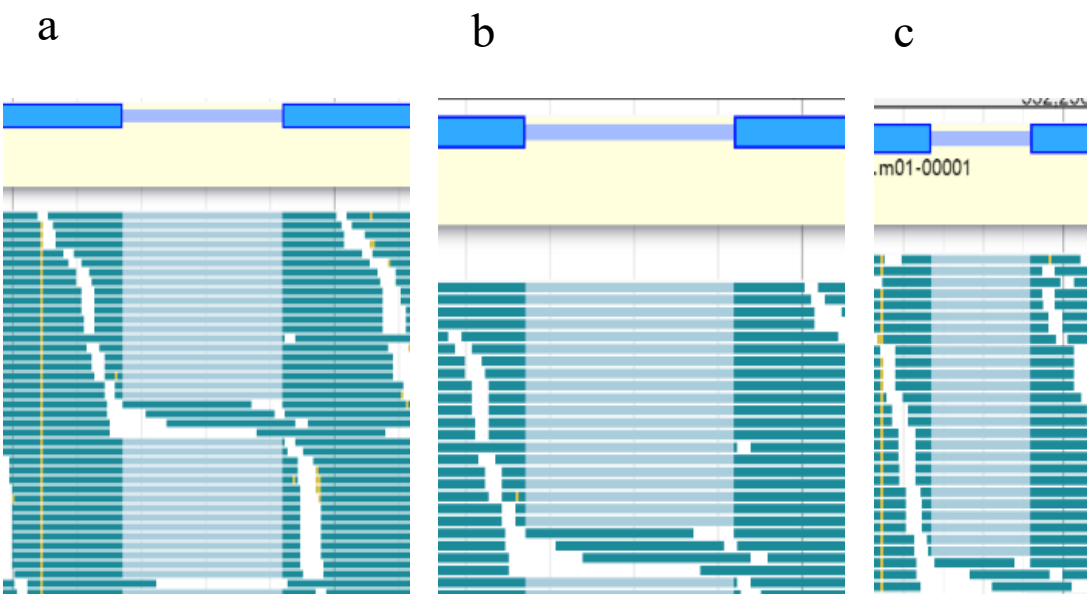

Figure 6. *CAP64* Evidence of Intron 3 Retention. The exon evidence (dark green) suggests the retention of intron 3 in the CAP64 gene. RNA transcriptome data that support this reannotation are (a) B58-3, (b) H99S_A-1, and (c) H99S_Sc100_RM2.

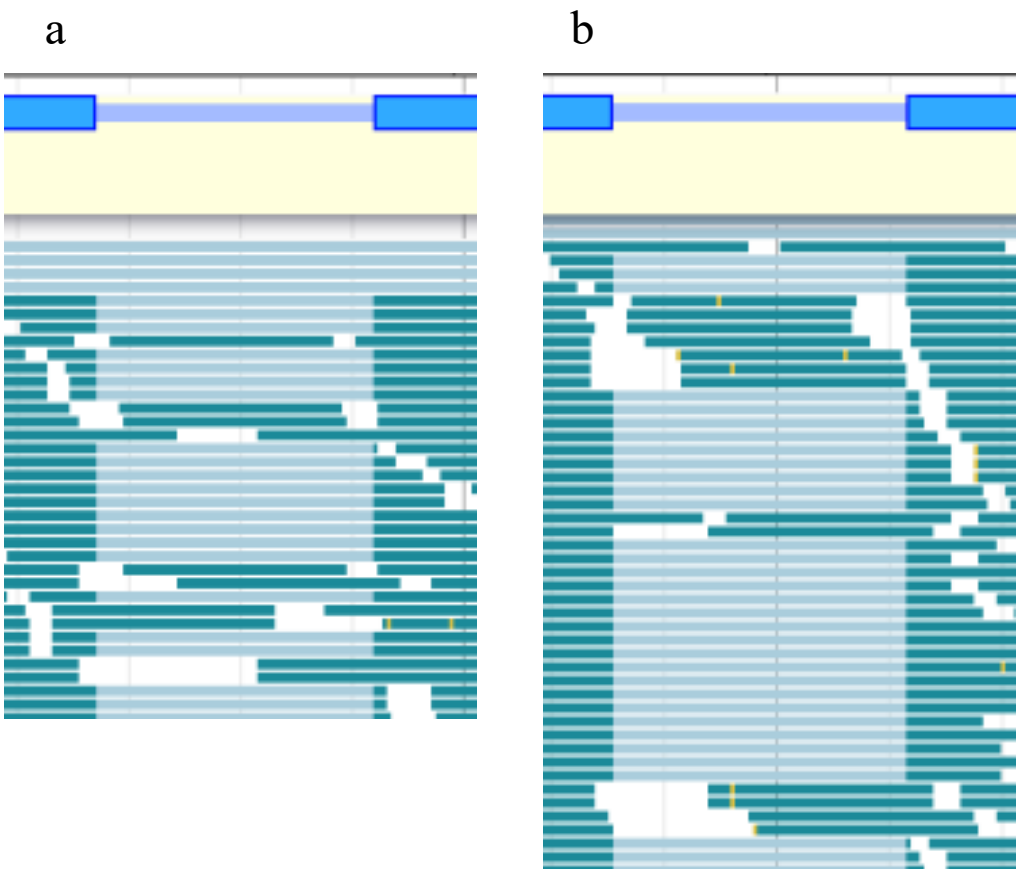

Figure 7. *CAP64* Evidence of Introns 3 and 7 Simultaneous Retention. The exon evidence (dark green) suggests the retention of (a) intron 3 and (b) intron 7 occurring in the same evidence tracks simultaneously. The RNA transcriptome data used to suggest these reannotations are B58-1 and B43-2 (not shown).

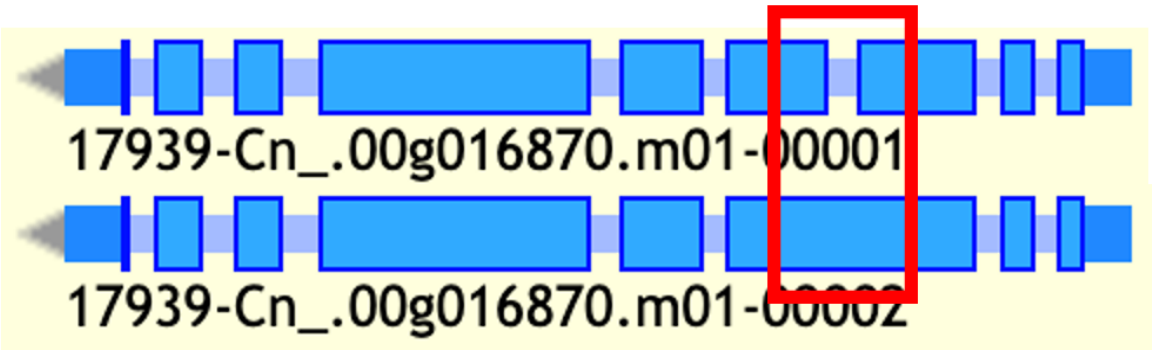

Figure 8. *CAP64* Gene Transcript Model Reannotation of Intron 3 Retention. The red rectangle indicates that intron 3 in the original PASA gene transcript model (top) is retained in the modified gene transcript model (bottom), combining exons 3a and 4 into exon 3b.

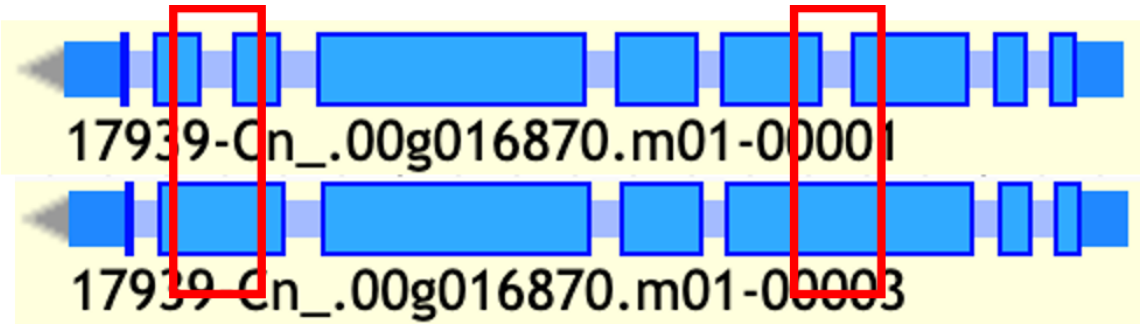

Figure 9. *CAP64* Gene Transcript Model Reannotation of Introns 3 and 7 Simultaneous Retention. The red rectangle on the right indicates that intron 3 present in the original PASA gene transcript model (top) is retained in the modified gene transcript model (bottom), combining exons 3a and 4 to form exon 3b. The rectangle on the left indicates intron 7 present in the original PASA gene transcript model (top) is retained in the modified gene transcript model (bottom), combining exons 7a and 8 to create exon 7b. Both of these retentions occurred in the same evidence tracks simultaneously.

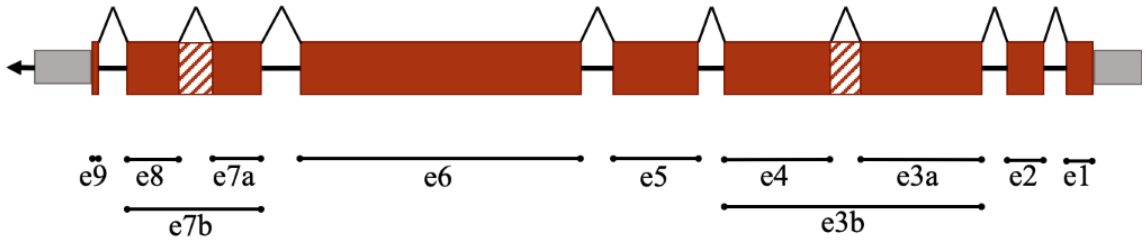

Figure 10. *CAP64* Super-Transcript Gene Model. This gene transcript model represents all possible paths of translation as supported by the RNA transcriptome data. The retention of introns 3 and 7 is displayed by the boxes with red and white diagonal lines, forming exons 3b and 7b.

CAP59

The comparison of the FungiDB, PASA, and original gene transcript models for *CAP59* presents an inconsistency between the FungiDB gene transcript model (Figure 11a) and the PASA gene transcript model (Figure 11b) in which the 3' untranslated region contains an intron in the FungiDB gene transcript model that is not present in the PASA gene transcript model. Five sets of reannotations were created for *CAP59* based on the supporting evidence. Set one presented evidence for the retention of introns 2 and 4 occurring simultaneously (Figure 12), set two suggested evidence for the retention of intron 4 (Figure 13), set three presented evidence for the retention of introns 4 and 5 occurring simultaneously (Figure 14), set four suggested evidence for the simultaneous retention of introns 1 and 4 (Figure 15), and set five showed evidence for the retention of intron 5 (Figure 16). Gene transcript models were modified based on the PASA gene transcript model. The gene transcript model proposed for set one was modified to present the retention of introns 2 and 4 occurring simultaneously (Figure 17). For set two, intron 4 retention is displayed in the modified gene transcript model (Figure 18). Set three of *CAP59* was used to modify a gene transcript model into displaying the reannotation of introns 4 and 5 occurring simultaneously (Figure 19). A gene transcript model of the simultaneous retention of introns 1 and 4 was created using the evidence from set four of *CAP59* (Figure 20). Set five was used to modify a PASA gene transcript model to display the retention of intron 5 (Figure 21). Figures 17, 18, 19, 20, and 21 present the gene transcript models used to create the super-transcript gene model for *CAP59* (Figure 22).

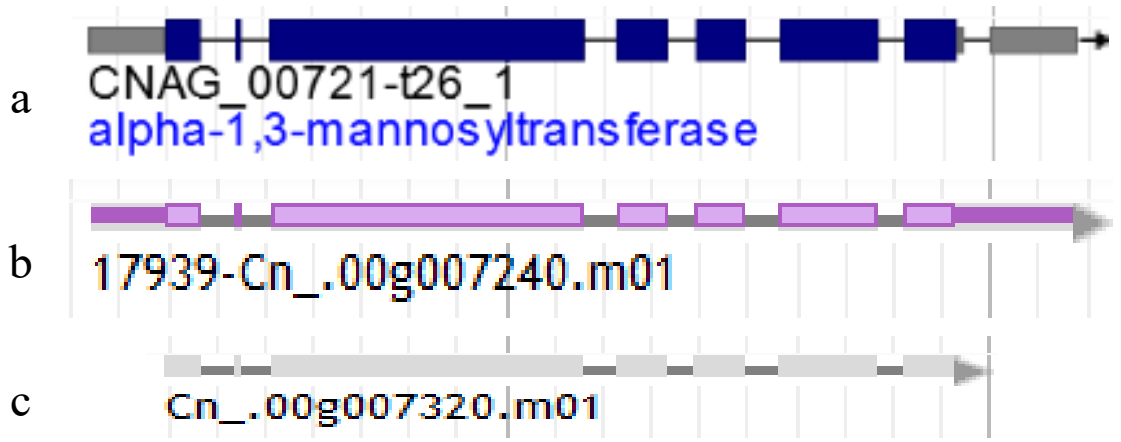

Figure 11. Reference Gene Transcript Models for *CAP59*. (a) The gene transcript model proposed by FungiDB. Note the intron within the 3' untranslated region. (b) The PASA algorithm-proposed gene transcript model for *CAP59*. Note the absence of the intron within the 3' untranslated region. (c) The gene transcript model proposed by the original genome data.

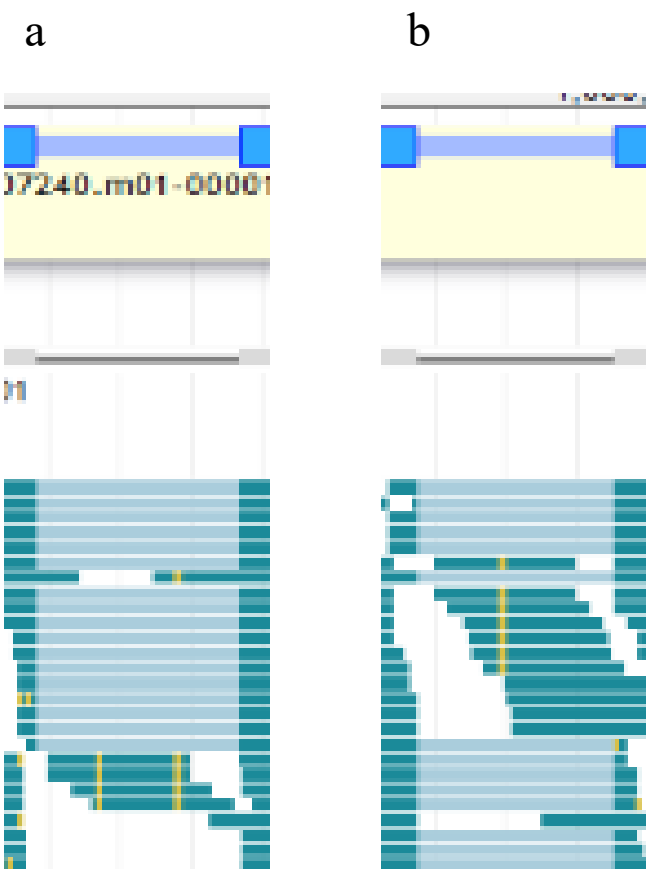

Figure 12. *CAP59* Evidence of Introns 2 and 4 Simultaneous Retention. The exon evidence (dark green) suggests the retention of (a) intron 2 and (b) intron 4 occurring in the same evidence tracks simultaneously. The RNA transcriptome data that suggest these retentions are B15-1 and B18-2 (not shown).

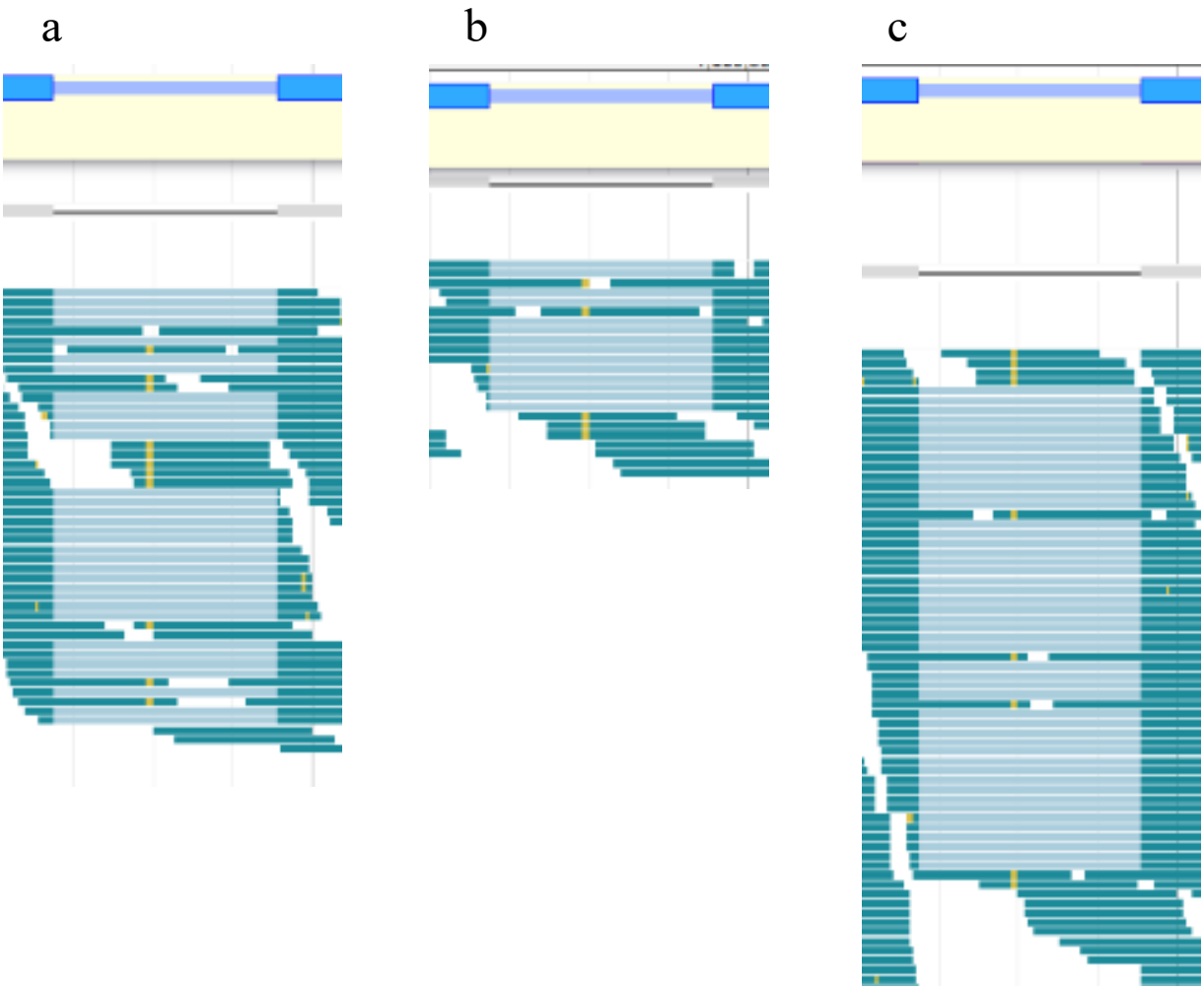

Figure 13. *CAP59* Evidence of Intron 4 Retention. The exon evidence (dark green) suggests the retention of intron 4. (a) B30-2, (b) B30-3, (c) B43-2, B43-3 (not shown), and H99S_A-1(not shown) are RNA transcriptome data that suggest the retention.

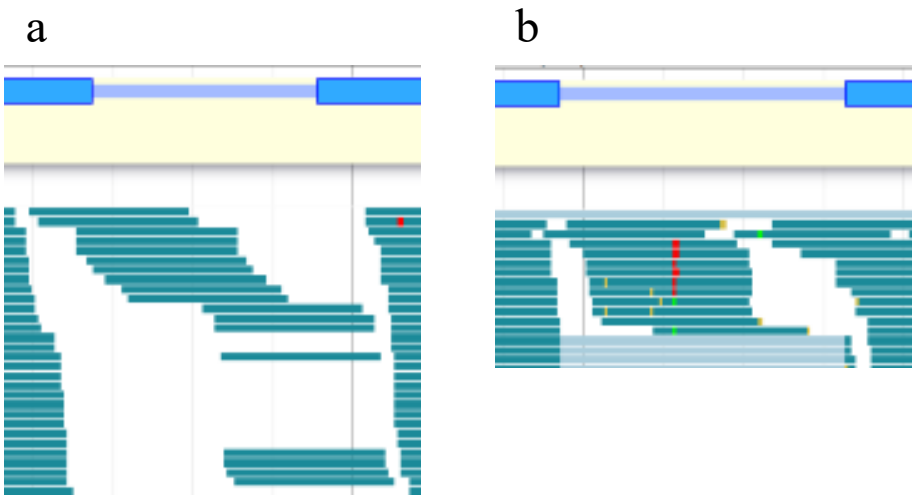

Figure 14. *CAP59* Evidence of Introns 4 and 5 Simultaneous Retention. The exon evidence (dark green) suggests the retention of (a) intron 4 and (b) intron 5 occurring in the same evidence track simultaneously. B18-1 is the RNA transcriptome data that suggest these retentions.

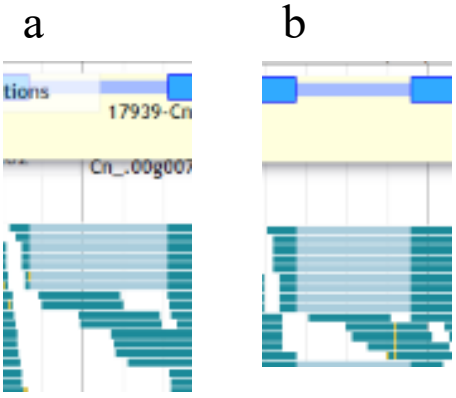

Figure 15. *CAP59* Evidence of Introns 1 and 4 Simultaneous Retention. The exon evidence (dark green) suggests the retention of (a) intron 1 and (b) intron 4 occurring in the same evidence track simultaneously. B58-2 is the RNA transcriptome data that suggest these retentions.

Figure 16. *CAP59* Evidence of Intron 5 Retention. The exon evidence (dark green) suggests the retention of intron 5. The RNA transcriptome data that suggest this retention are (a) H99S_A-2 and (b) H99S_Sc100_T_R_M2.

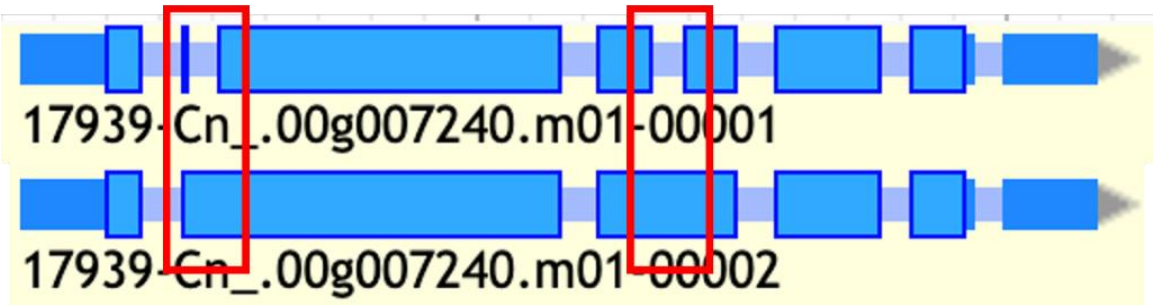

Figure 17. *CAP59* Gene Transcript Model Reannotation of Introns 2 and 4 Simultaneous Retention. The red rectangle on the left presents intron 2 being present in the original PASA transcript gene transcript model (top) but retained in the modified gene transcript model (bottom), combining exons 2a and 3 to form exon 2b. The red rectangle to the right indicates intron 4, which is present in the original PASA gene transcript model (top) being retained in the modified gene transcript model (bottom), combining exons 4a and 5a into exon 4b. These retentions occurred in the same evidence tracks simultaneously.

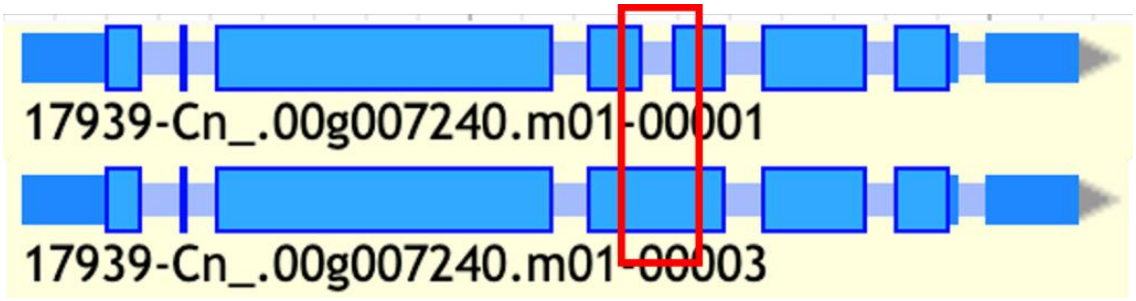

Figure 18. *CAP59* Gene Transcript Model Reannotation of Intron 4 Retention. The red rectangle indicates intron 4 is present in the original PASA gene transcript model (top) but is retained in the modified gene transcript model (bottom), combining exons 4a and 5a to form exon 4b.

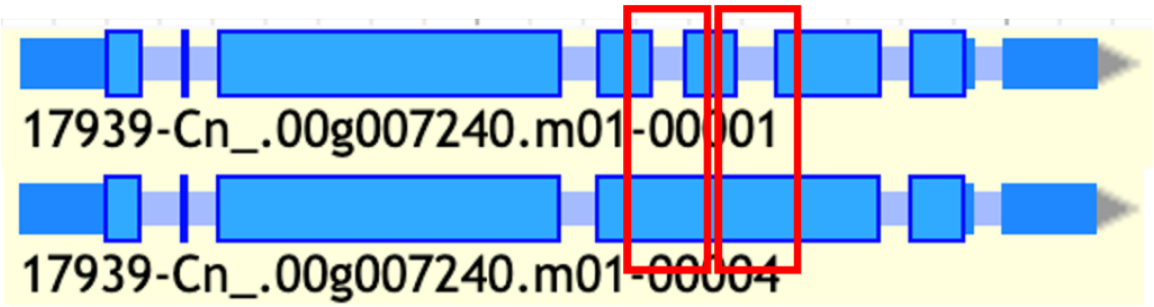

Figure 19. *CAP59* Gene Transcript Model Reannotation of Introns 4 and 5 Simultaneous Retention. The red rectangle on the left indicates that intron 4 present in the original PASA gene transcript model (top) is retained in the modified gene transcript model (bottom). The rectangle on the right indicates that intron 5 present in the original PASA gene transcript model (top) is retained in the modified gene transcript model (bottom). Together, these two annotations combine exons 4a, 5a, and 6 to form exon 4c. These retentions occurred in the same evidence track simultaneously.

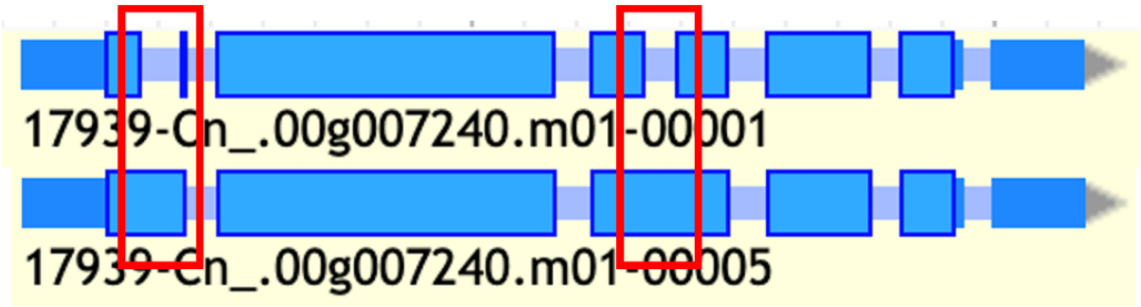

Figure 20. *CAP59* Gene Transcript Model Reannotation of Introns 1 and 4 Simultaneous Retention. The red rectangle on the left indicates that intron 1 present in the original PASA gene transcript model (top) is retained in the modified gene transcript model (bottom), combining exons 1a and 2 to form exon 1b. The red rectangle on the right indicates that intron 4 present in the original PASA gene transcript model (top) is retained in the modified gene transcript model (bottom), combining exons 4a and 5a to form exon 4b. These retentions occurred in the same evidence track simultaneously.

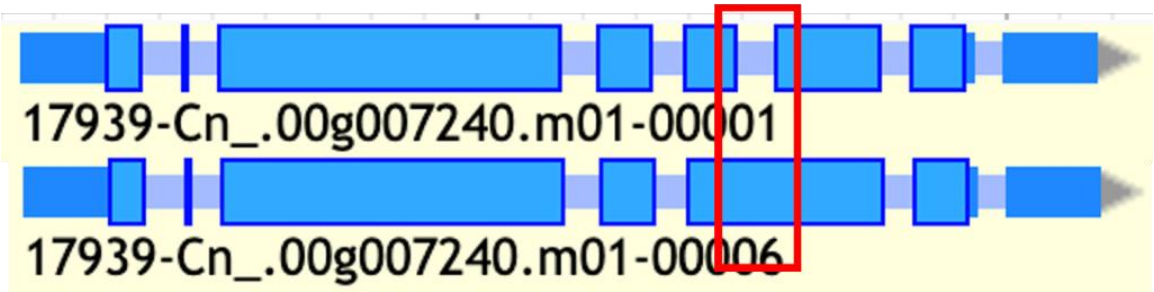

Figure 21. *CAP59* Gene Transcript Model Reannotation of Intron 5 Retention. The red rectangle indicates that intron 5 present in the original PASA gene transcript model (top) is retained in the modified gene transcript model (top), combining exons 5a and 6 to form exon 5b.

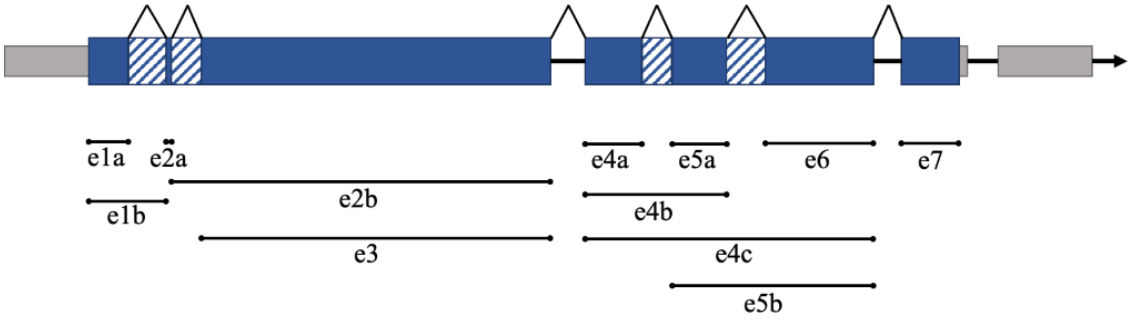

Figure 22. *CAP59* Super-Transcript Gene Model. The RNA transcriptome data that support the reannotations for *CAP59* were used to create the gene transcript model representing all possible paths for translation. The retention of introns 1, 2, 4, and 5 is displayed by the boxes with blue and white diagonal lines, forming exons 1b, 2b, 4b, 4c, and 5b.

CMT1

The lack of differences present when comparing the FungiDB, PASA, and original genome transcript models for *CMT1* suggested that each model was similar (Figure 23). One set of reannotations was required, suggesting the retention of intron 2 (Figure 24), based on supporting evidence. A gene transcript model was modified to present the retention of intron 2 within *CMT1* (Figure 25). The super-transcript gene model for *CMT1* was created using the reannotation in the gene transcript model displayed in Figure 25 (Figure 26).

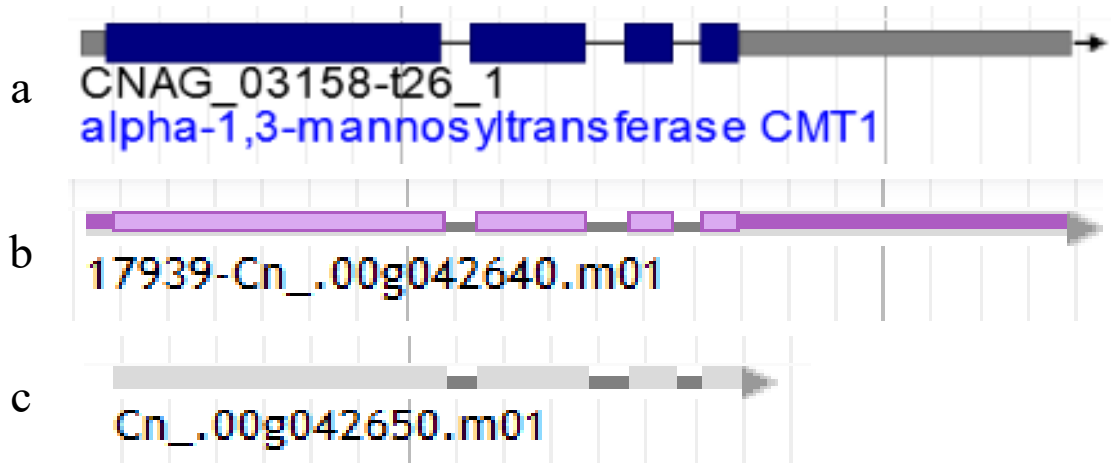

Figure 23. Reference Gene Transcript Models for *CMT1*. (a) The FungiDB model of *CMT1*. (b) The gene transcript model generated by the PASA algorithm. (c) The original genome gene transcript model for *C. neoformans*.

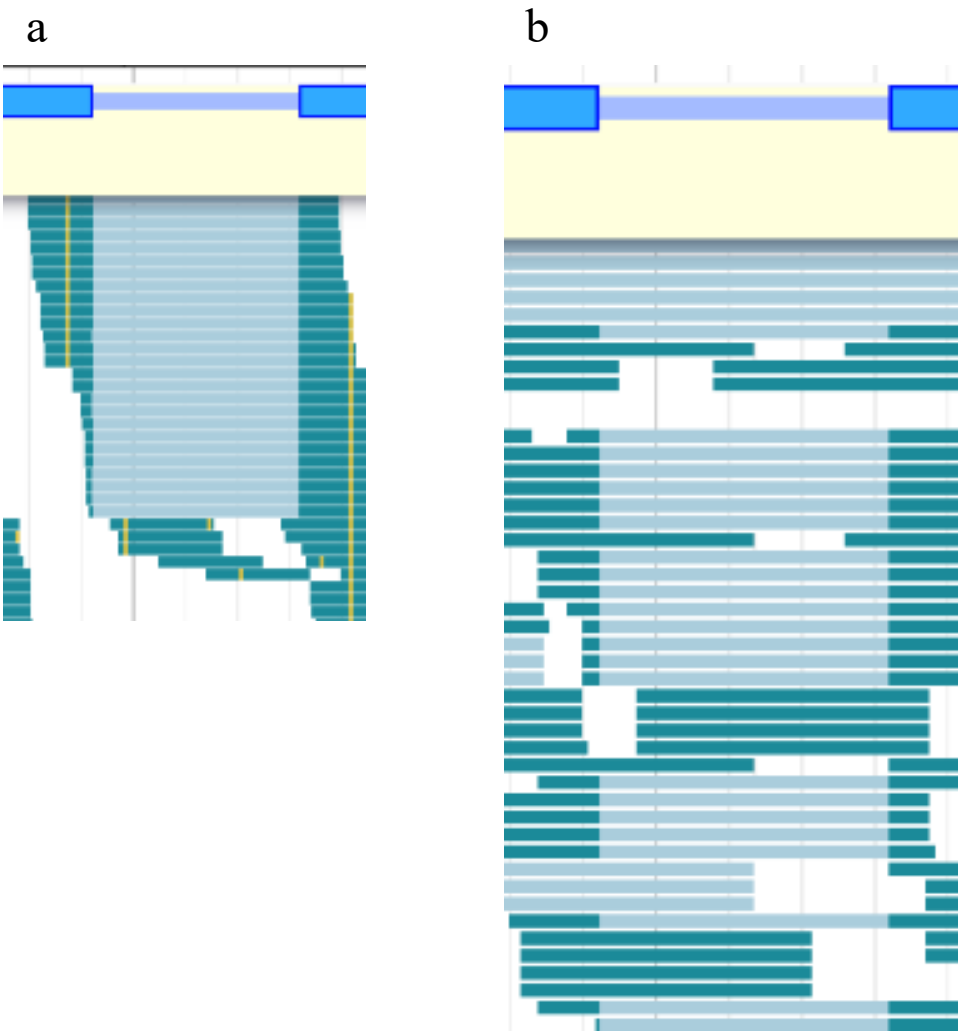

Figure 24. *CMT1* Evidence of Intron 2 Retention. The exon data (dark green) suggest the retention of intron 2. (a) B58-3 and (b) H99S_Sc100_T_R_M2 are the RNA transcriptome data that suggest the retention.

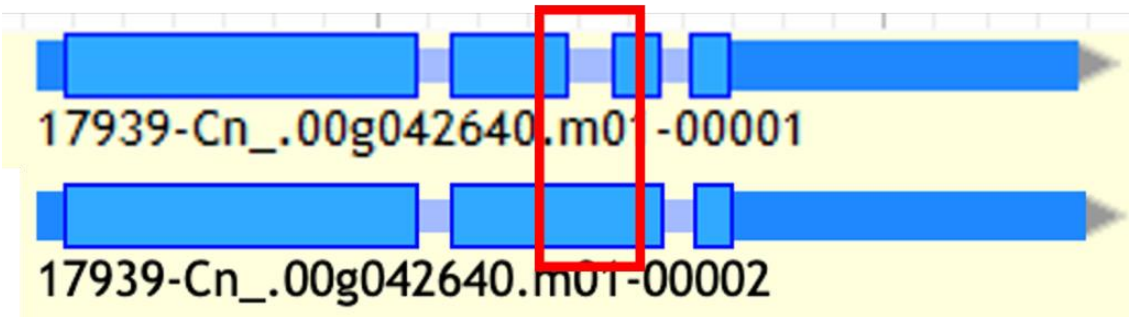

Figure 25. *CMT1* Gene Transcript Model Reannotation of Intron 2 Retention. The red rectangle indicates that intron 2 present in the original PASA gene transcript model (top) is retained in the modified gene transcript model (bottom), combining exons 2a and 3 into exon 2b.

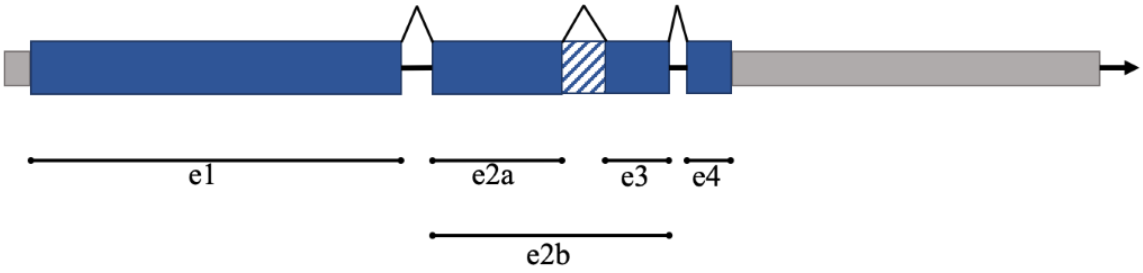

Figure 26. *CMT1* Super-Transcript Gene Model. This gene transcript model represents all possible paths of translation as supported by the RNA transcriptome data. The retention of intron 2 displayed by the boxes with blue and white diagonal lines formed exon 2b.

CAP10

When viewing the FungiDB, PASA, and original genome transcript models for *CAP10*, no differences were found, so the gene models were similar (Figure 27). One set of reannotations was required, which suggested an increase in the length of exon 4 (Figure 28). A gene transcript model was modified to display the increase in the length of exon 4 in *CAP10* (Figure 29). Figure 29 presents the gene transcript model used in the super-transcript gene model for *CAP10* (Figure 30).

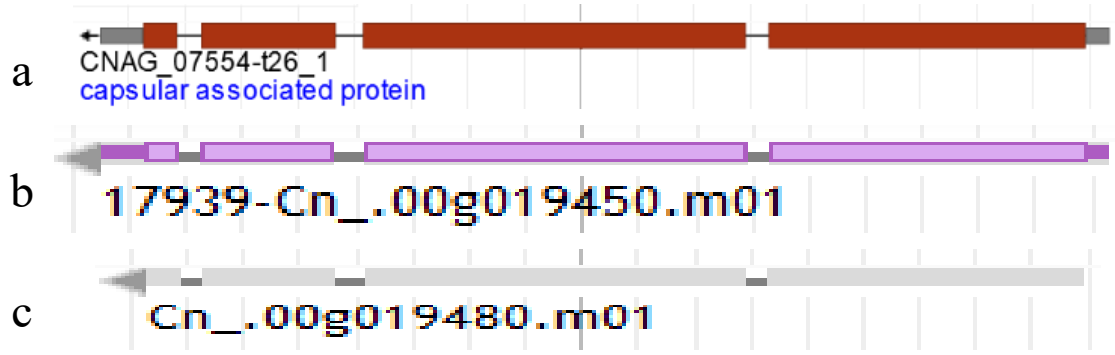

Figure 27. Reference Gene Transcript Models for *CAP10*. (a) The *CAP10* gene transcript model presented in FungiDB. (b) The gene transcript model proposed by the PASA algorithm. (c) The gene transcript model proposed by the original genome.

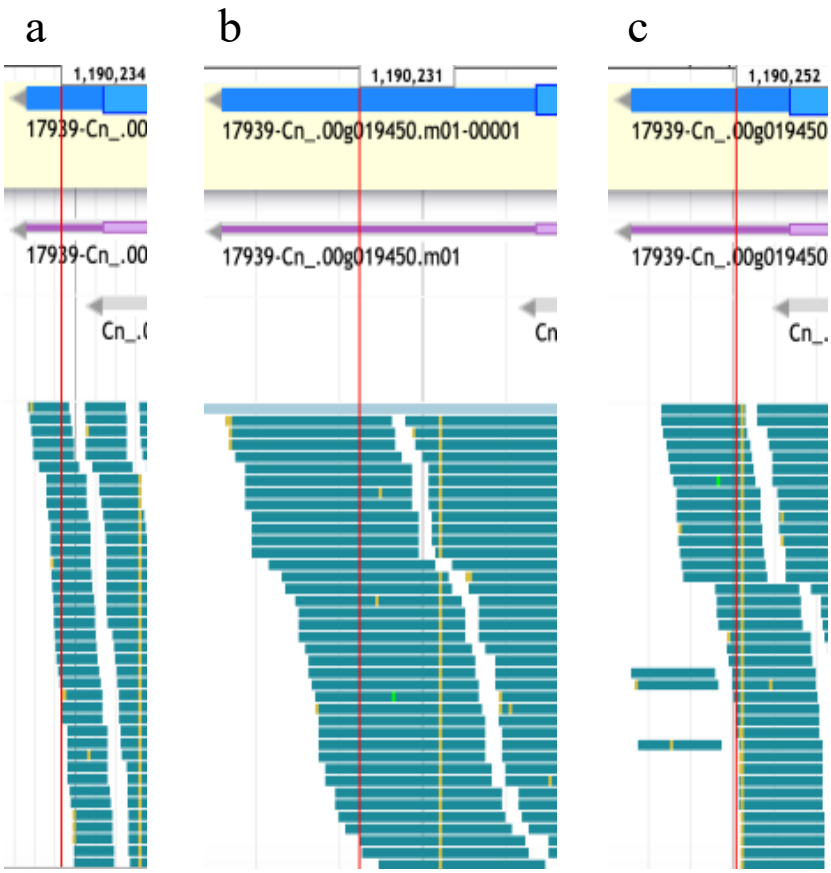

Figure 28. *CAP10* Evidence of Increase in Exon 4 Length. The exon data (dark green) suggest an increase in the length of exon 4 because of the presence of exon data in the 5' untranslated region. (a) B15-1, (b) B15-2, and (c) B15-3 are the RNA transcriptome data that suggest the increase in the length of exon 4. B18-1, B18-2, B30-2, B45-3, B58-2, and H99S_A-2 are RNA transcriptome data that suggest the increase in the length of exon 4 but are not shown.

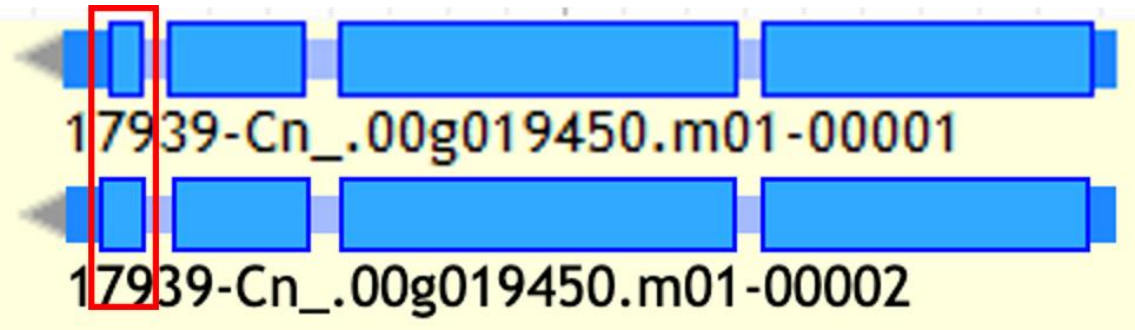

Figure 29. *CAP10* Gene Transcript Model Reannotation of the Increase of Exon 4a Length. The red rectangle presents that exon 4a, which is present in the original PASA gene transcript model (top), is extended in length in the modified gene transcript model (bottom) to form exon 4b. This decreases the length of the 3' untranslated region.

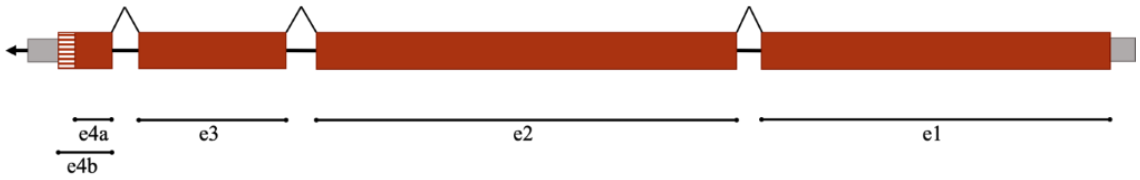

Figure 30. *CAP10* Super-Transcript Gene Model. The RNA transcriptome data that support the reannotation for *CAP10* were used to create the gene transcript model presenting all possible paths for translation. The increase in the length of exon 4a displayed by the boxes with red and white horizontal lines formed exon 4b.

CAP60

The lack of differences present when comparing the FungiDB, PASA, and original genome transcript models for *CAP60* suggested that each model was similar (Figure 31). Two sets of reannotations for the PASA gene transcript model were required. Evidence suggested the retention of intron 1 for set one (Figure 32) and the simultaneous retention of introns 1 and 2 for set two (Figure 33). Set one was used to modify a gene transcript model to display the retention of intron 1 (Figure 34). For set two, modifications to a gene transcript model were made to display the retention of introns 1 and 2 occurring simultaneously (Figure 35). The super-transcript gene model for *CAP60* was created by compiling the gene transcript models presented in Figures 34 and 35 (Figure 36).

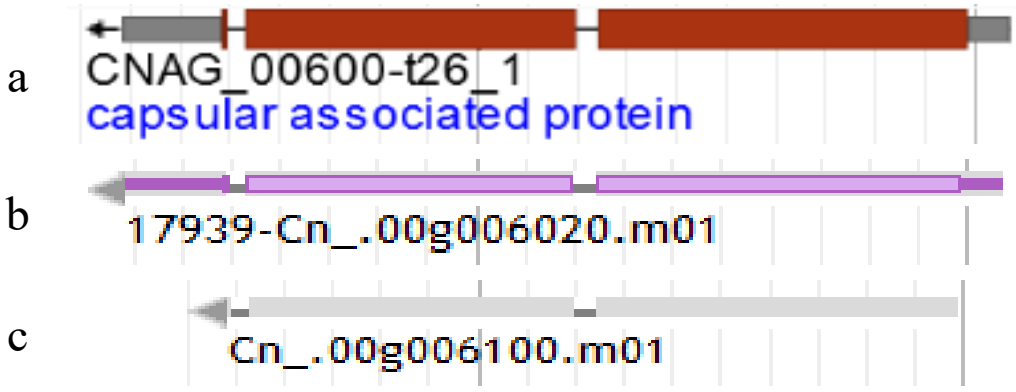

Figure 31. Reference Gene Transcript Models for *CAP60*. (a) The FungiDB gene transcript model for *CAP60*. (b) The PASA gene transcript model. (c) The gene transcript model proposed by the original gene transcript model.

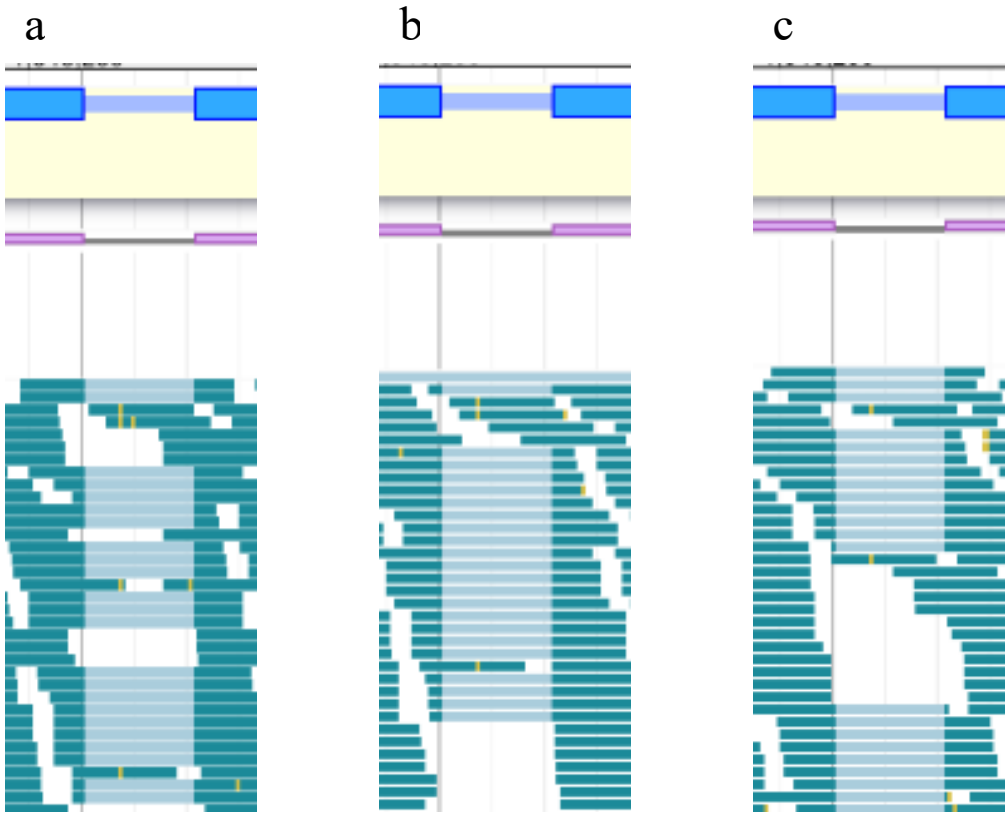

Figure 32. *CAP60* Evidence of Intron 1 Retention. The exon data (dark green) suggest the retention of intron 1. The RNA transcriptome data that suggest the retention are (a) B43-1, (b) B43-2, and (c) B43-3.

Figure 33. *CAP60* Evidence of Simultaneous Retention of Introns 1 and 2. The exon evidence (dark green) suggests the retention of (b) intron 1 and (a) intron 2 occurring in the same evidence tracks simultaneously. B30-2 and B30-3 (not shown) are the RNA transcriptome data that suggest the retentions.

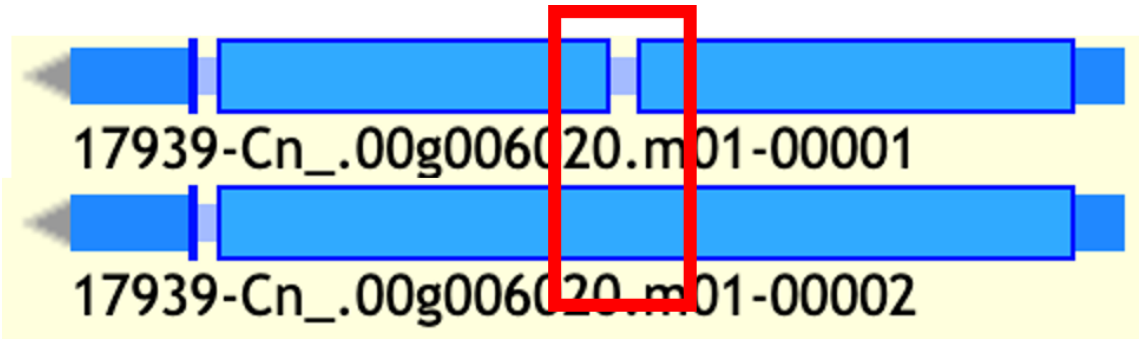

Figure 34. *CAP60* Gene Transcript Model Reannotation of Intron 1 Retention. The red rectangle indicates that intron 1 present in the original PASA gene transcript model (top) is retained in the modified gene transcript model (bottom), combining exons 1a and 2a to form exon 1b.

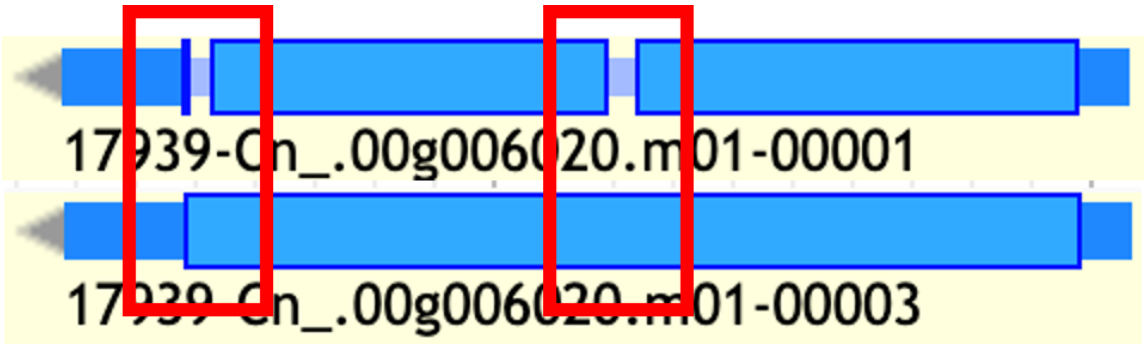

Figure 35. *CAP60* Gene Transcript Model Reannotation of Introns 1 and 2 Simultaneous Retention. The red rectangle to the right indicates that intron 1 present in the original PASA gene transcript model (top) is retained in the modified gene transcript model (bottom). The red rectangle to the left indicates that intron 2 present in the original PASA gene transcript model (top) is retained in the modified gene transcript model (bottom). Both reannotations combined exons 1a, 2, and 3 to form exon 1c. These retentions occurred in the evidence tracks simultaneously.

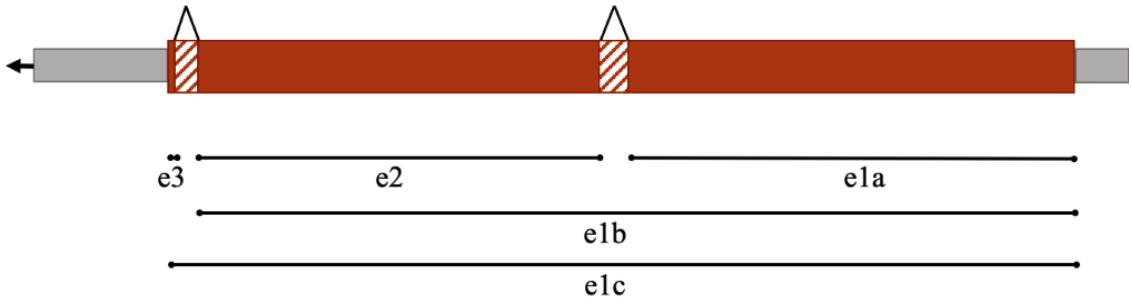

Figure 36. *CAP60* Super-Transcript Gene Model. The gene transcript model represents all possible paths of translation as supported by the RNA transcriptome data. The retention of introns 1 and 3 presented by the boxes with red and white diagonal lines created exons 1b and 1c.

CAS1

The comparison of the FungiDB, PASA, and original gene transcript models for *CAS1* suggests exons 6 and 7 are present in the FungiDB and PASA gene transcript models (Figure 37a and 37b) but are not present in the original genome gene transcript model (Figure 37c). Three sets of reannotations were required in which set one presented evidence for the retention of intron 6 (Figure 38), set two suggested evidence for the retention of intron 8 (Figure 39), and set three presented evidence for the simultaneous retention of introns 7 and 8 (Figure 40). Gene transcript models for the reannotations proposed by each of the three sets were created. A gene transcript model was made to represent the intron 6 retention from set one (Figure 41), intron 8 retention from set two (Figure 42), and introns 7 and 8 retention occurring simultaneously from set three (Figure 43). The super-transcript for *CAS1* was made by compiling the reannotations presented in the gene transcript models in Figures 41, 42, and 43 (Figure 44).

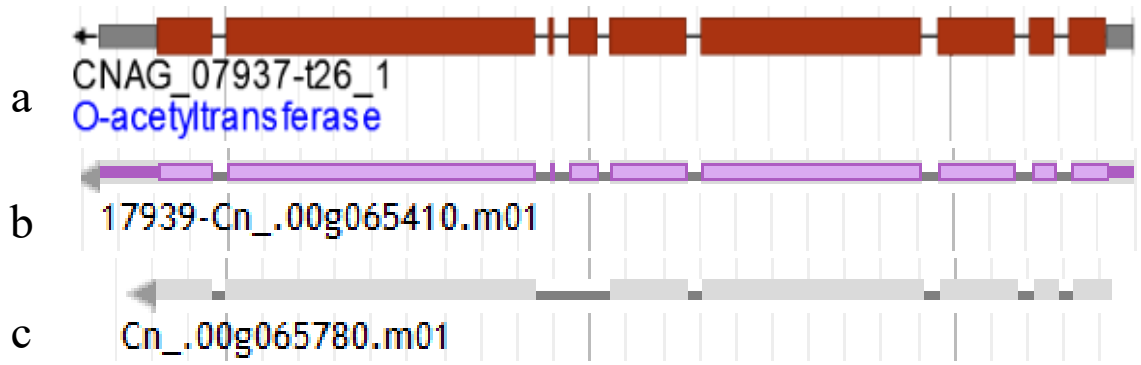

Figure 37. Reference Gene Transcript Models for *CAS1*. (a) The FungiDB-proposed gene transcript model for *CAS1*. (b) The gene transcript model proposed by the PASA algorithm. (c) The gene transcript model proposed by the original genome. Note the absence of exons 6 and 7.

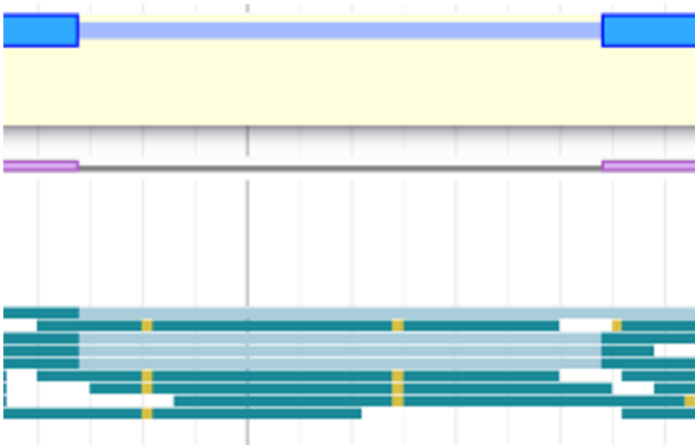

Figure 38. *CAS1* Evidence of Intron 6 Retention. The exon evidence (dark green) suggests the retention of intron 6. B40-1 is the RNA transcriptome data that suggested retention.

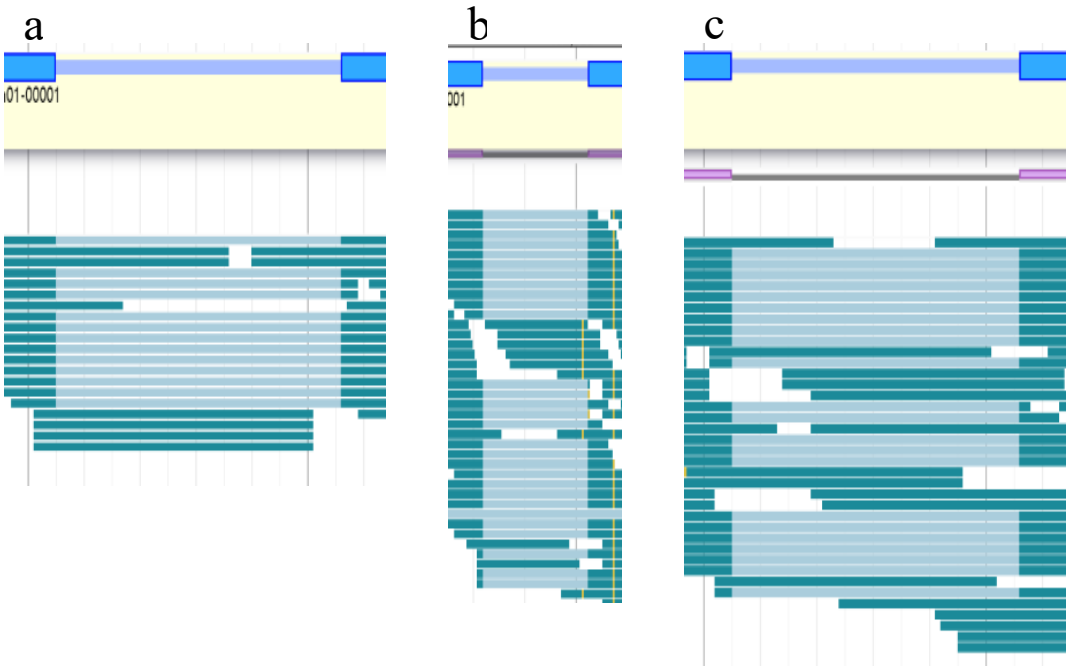

Figure 39. *CAS1* Evidence of Intron 8 Retention. The exon evidence (dark green) suggests the retention of intron 8. The RNA transcriptome data that suggest the retention are (a) B30-3, (b) B43-1, and (c) B58-1.

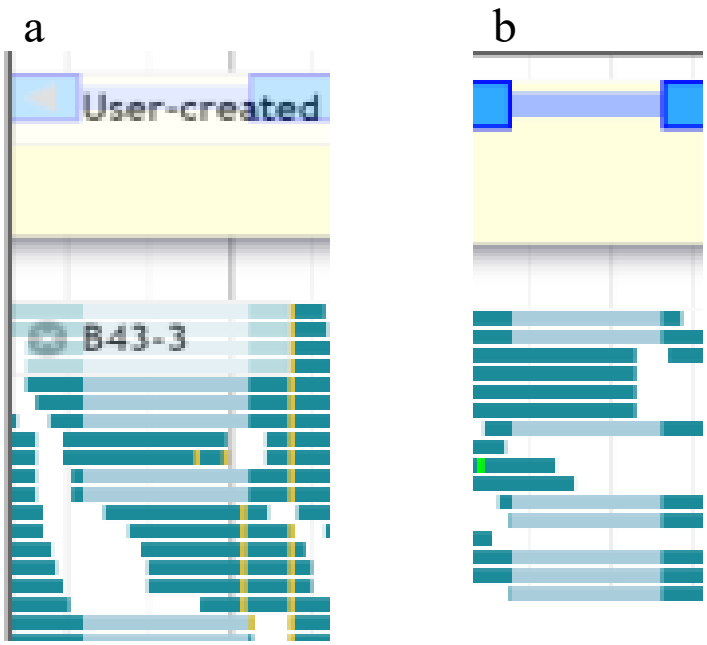

Figure 40. *CAS1* Evidence of Introns 7 and 8 Simultaneous Retention. The exon evidence (dark green) suggests the retention of (b) intron 7 and (b) intron 8 occurring in the same evidence tracks simultaneously. B43-3 is the RNA transcriptome evidence that suggests retention.

Figure 41. *CAS1* Gene Transcript Model Reannotation of Intron 6 Retention. The red rectangle indicates that intron 6 present in the original PASA gene transcript model (top) is retained in the modified gene transcript model (bottom), combining exons 6a and 7a to form exon 6b.

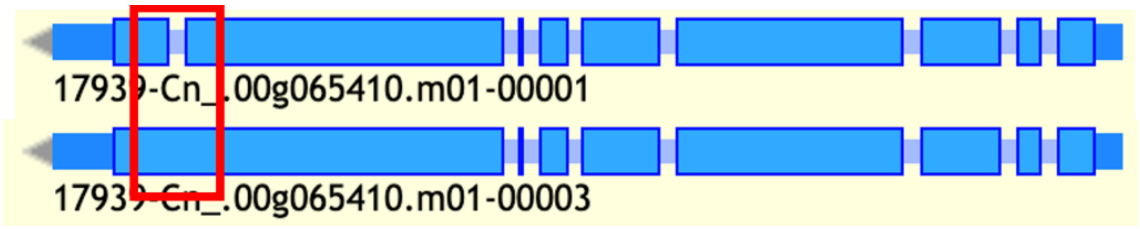

Figure 42. *CAS1* Gene Transcript Model Reannotation of Intron 8 Retention. The red rectangle indicates that intron 8 present in the original PASA gene transcript model (top) is retained in the modified gene transcript model (bottom), combining exons 8a and 9 to form exon 8b.

Figure 43. *CAS1* Gene Transcript Model Reannotation of Introns 7 and 8 Simultaneous Retention. The red rectangle to the right indicates that intron 7 present in the original PASA gene transcript model (top) is retained in the modified gene transcript model (bottom). The red rectangle to the left indicates that intron 8 present in the original PASA gene transcript model (top) is retained in the modified gene transcript model (bottom). The reannotations combined exons 7a, 8a, and 9 to form exon 7c. These retentions occurred in the evidence tracks simultaneously.

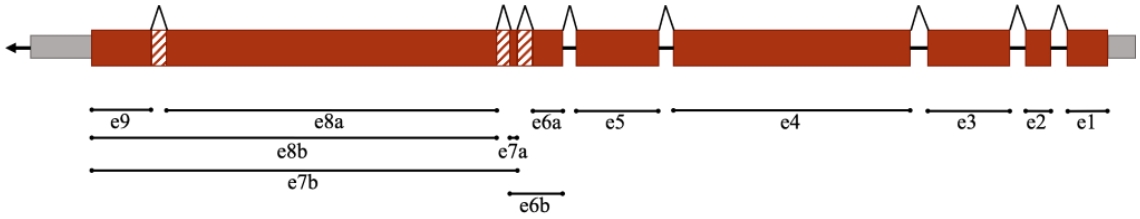

Figure 44. *CAS1* Super-Transcript Gene Model. The RNA transcriptome data that support the reannotations for *CAS1* were used to create the gene transcript model displaying all possible paths for translation. The retention of introns 6, 7, and 8 is represented by the boxes with red and white diagonal lines, forming exons 6b, 7b, and 8b.

CAS3

When viewing the FungiDB, PASA, and original genome transcript models for *CAS3*, each gene model was similar, and no differences were found (Figure 45). Three sets of reannotations were created. Set one suggested evidence for the retention of intron 5 (Figure 46), set two presented the simultaneous retention of introns 5 and 11 (Figure 47), and set three suggested evidence for the retention of introns 1, 5, 8, 9, and 11, with the increase in the length of exon 12 occurring simultaneously (Figure 48). Gene transcript models were made for each of the three sets for *CAS3*, with one model being modified to display the retention of intron 5 from set one (Figure 49) and one model displaying the modification of the simultaneous retention of introns 5 and 11 from set two (Figure 50). Set three RNA transcriptome data was used to create a gene transcript model to present the retention of introns 1, 5, 8, 9, and 11 with the increase in the length of exon 12 occurring simultaneously (Figure 51). The *CAS3* super-transcript gene model was created using the reannotations made in the gene transcript models presented in Figures 49, 50, and 51 (Figure 52).

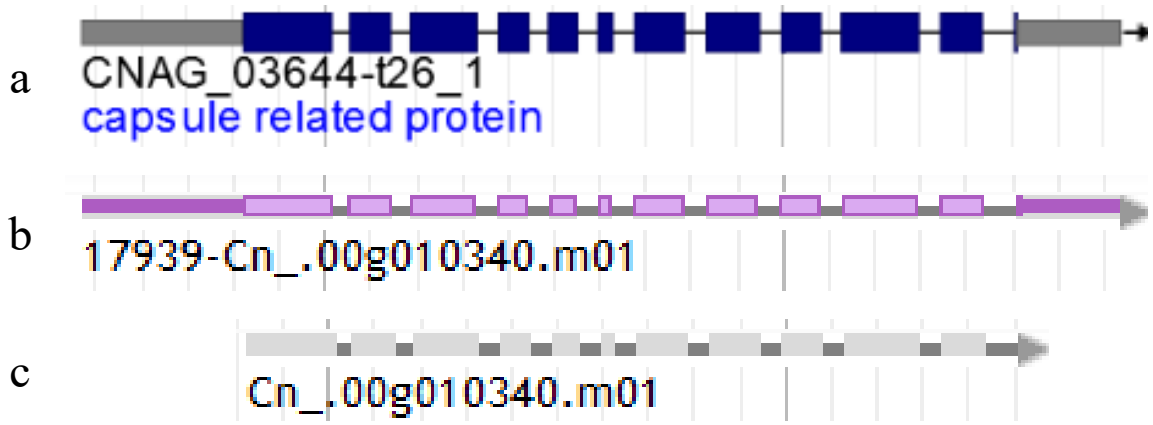

Figure 45. Reference Gene Transcript Models for *CAS3*. (a) The *CAS3* FungiDB gene transcript model. (b) The gene transcript model proposed by the PASA algorithm. (c) The original genome gene transcript model.

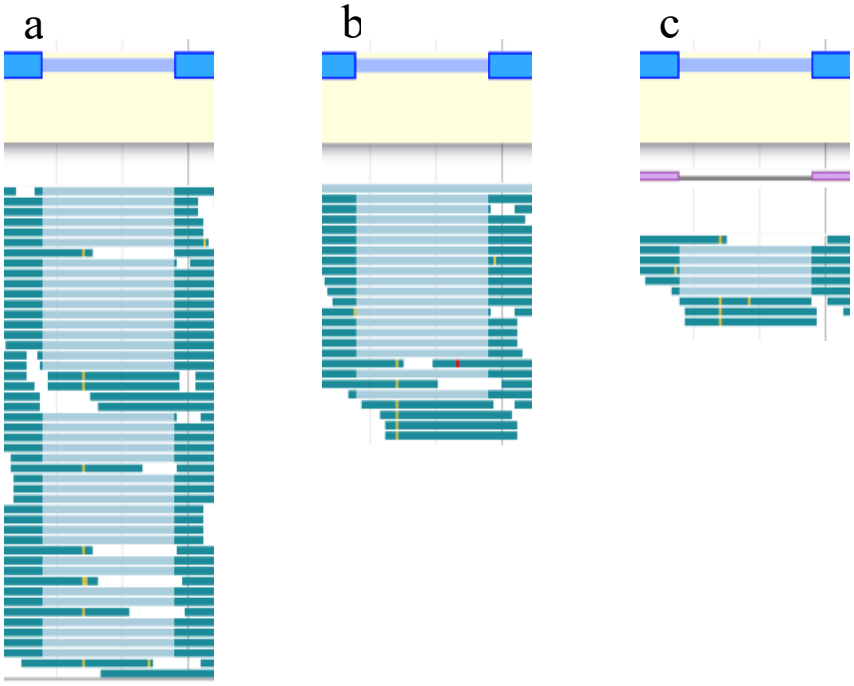

Figure 46. *CAS3* Evidence of Intron 5 Retention. The exon evidence (dark green) suggests the retention of intron 5. The RNA transcriptome data that suggest the retention are (a) B43-1, (b) B43-3, and (c) B45-2.

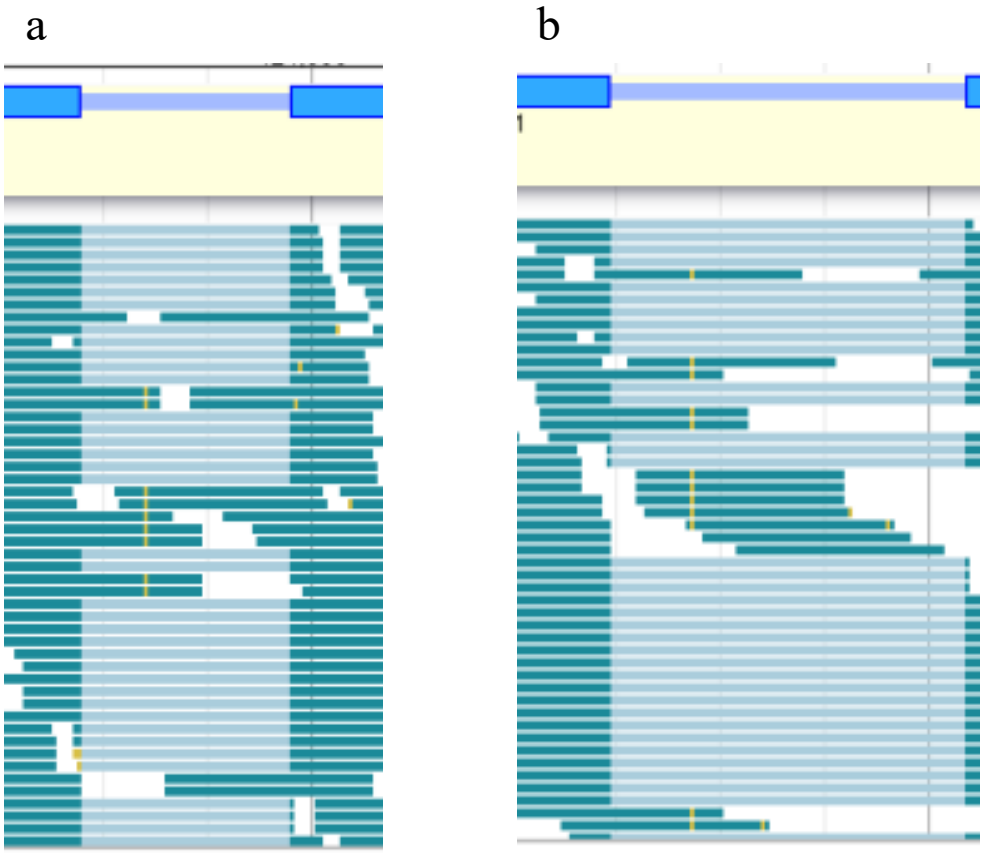

Figure 47. *CAS3* Evidence of Introns 5 and 11 Simultaneous Retention. The exon evidence (dark green) suggests the retention of (a) intron 5 and (b) intron 11 occurring in the same evidence track simultaneously. These retentions are suggested by the RNA transcriptome data B58-1.

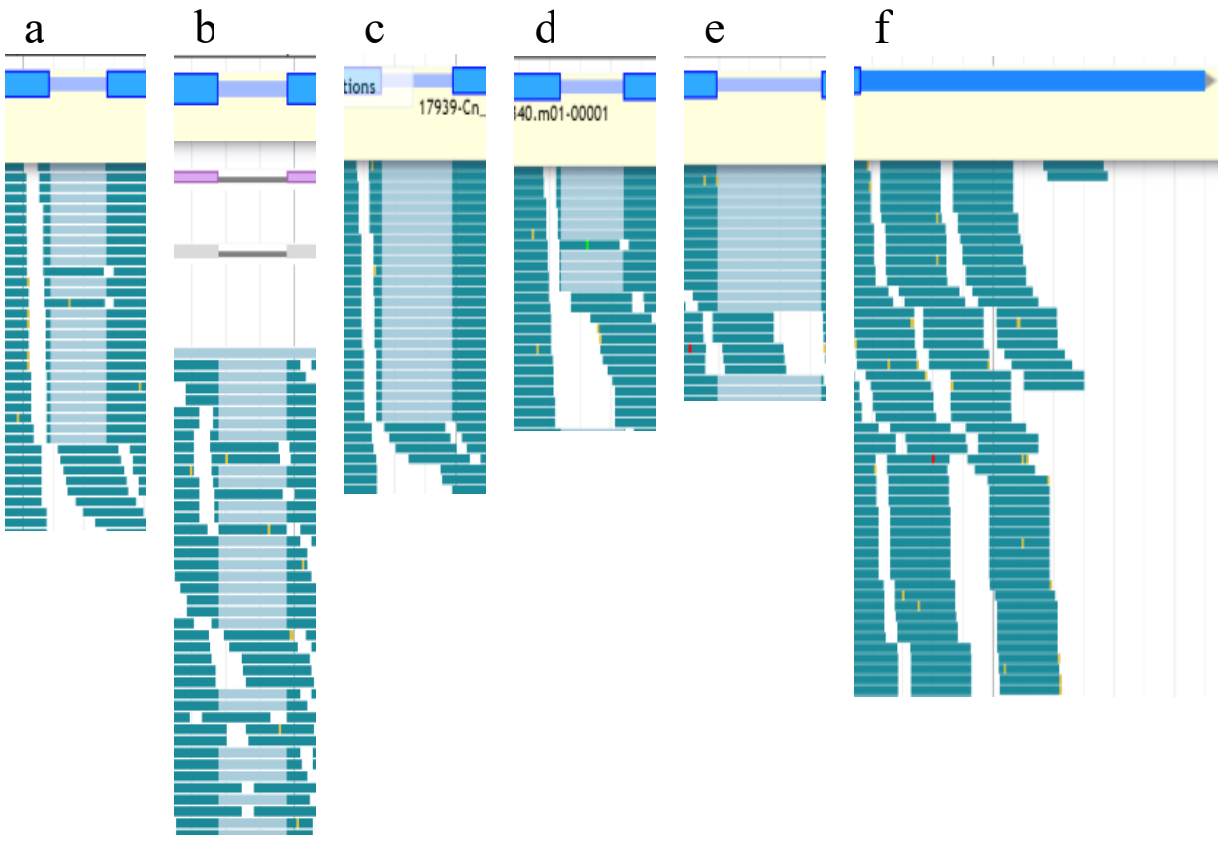

Figure 48. *CAS3* Evidence of Introns 1, 8, 9, and 11 Retention with an Increase in Length of Exon 12 Occurring Simultaneously. The exon evidence (dark green) suggests the retention of (a) intron 1, (b) intron 8, (c) intron 9, and (d) intron 11. (f) The increase in the length of exon 12 is suggested by the exon evidence being present in the 3' untranslated region. B18-1 is the RNA transcriptome data that suggest these retentions and exon lengthening. These reannotations occurred in the same evidence track simultaneously.

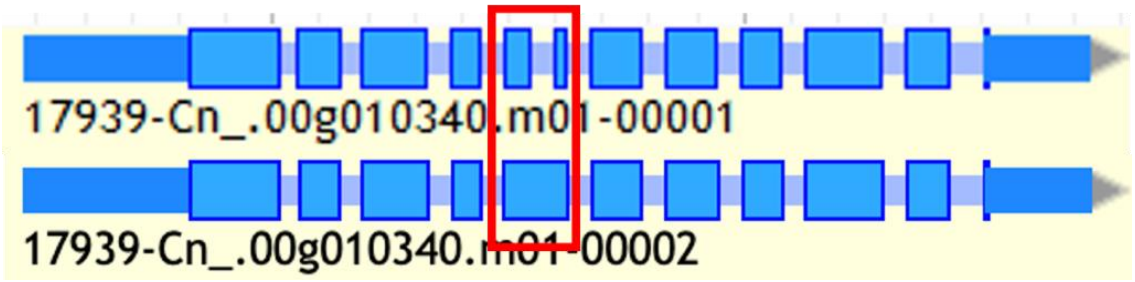

Figure 49. *CAS3* Gene Transcript Model Reannotation of Intron 5 Retention. The red rectangle indicates that intron 5 present in the original PASA gene transcript model (top) is retained in the modified gene transcript model (bottom), combining exons 5a and 6 to form exon 5b.

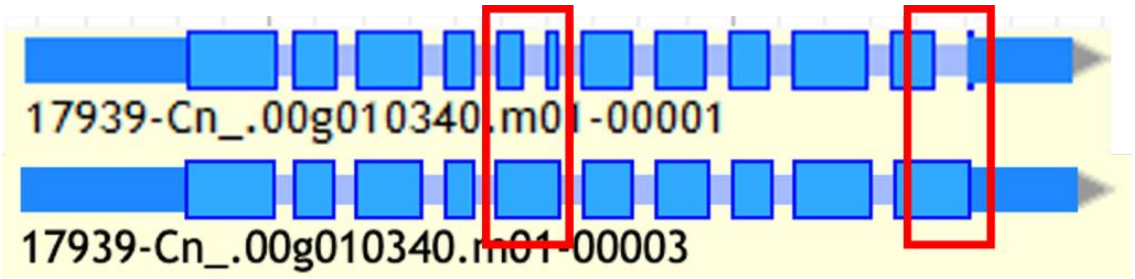

Figure 50. *CAS3* Gene Transcript Model Reannotation of Introns 5 and 11 Simultaneous Retention. The red rectangle to the left indicates that intron 5 present in the original PASA gene transcript model (top) is retained in the modified gene transcript model (bottom), combining exons 5a and 6 to create exon 5b. The red rectangle to the right indicates that intron 11 present in the original PASA gene transcript model (top) is retained in the modified gene transcript model (bottom), combining exons 11a and 12 to form exon 11b. These retentions occurred in the same evidence tracks simultaneously.

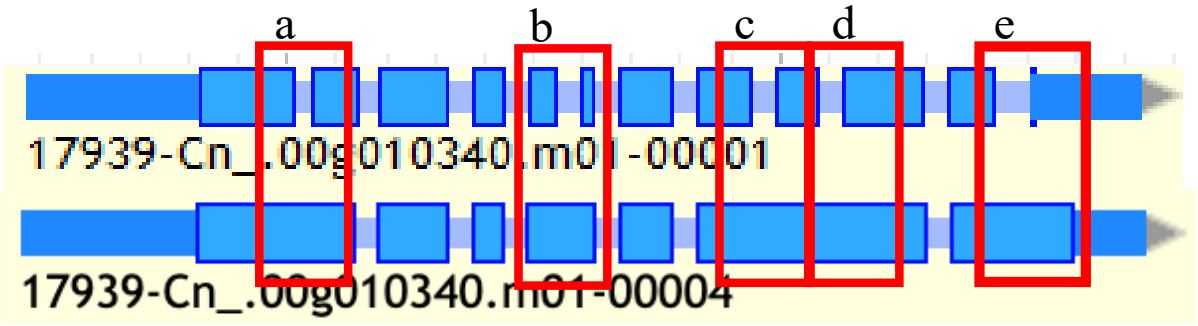

Figure 51. *CAS3* Gene Transcript Model Reannotations of Introns 1, 5, 8, 9, and 11 Retention and the Increase in Exon 12a Length Occurring Simultaneously. (a) The red rectangle indicates that intron 1 present in the original PASA gene transcript model (top) is retained in the modified gene transcript model (bottom), combining exons 1a and 2 to create exon 1b. (b) The red rectangle indicates that intron 5 present in the original PASA gene transcript model (top) is retained in the modified gene transcript model (bottom), combining exons 5a and 6 to create exon 5b. (c) The red rectangle indicates that intron 8 present in the original PASA gene transcript model (top) is retained in the modified gene transcript model (bottom). (d) The red rectangle indicates that intron 9 present in the original PASA gene transcript model (top) is retained in the modified gene transcript model (bottom). The reannotations from (c) and (d) combine exons 8a, 9, and 10 to form exon 8b. (e) The red rectangle indicates that intron 11 present in the original PASA gene transcript model (top) is retained in the modified gene transcript model (bottom). In addition to the retention of intron 11, the length of exon 12 was increased, decreasing the size of the 5' untranslated region (e). For both reannotations in (e), the exons 11a and 12 were combined to form 11c. These reannotations occurred in the same evidence track simultaneously.

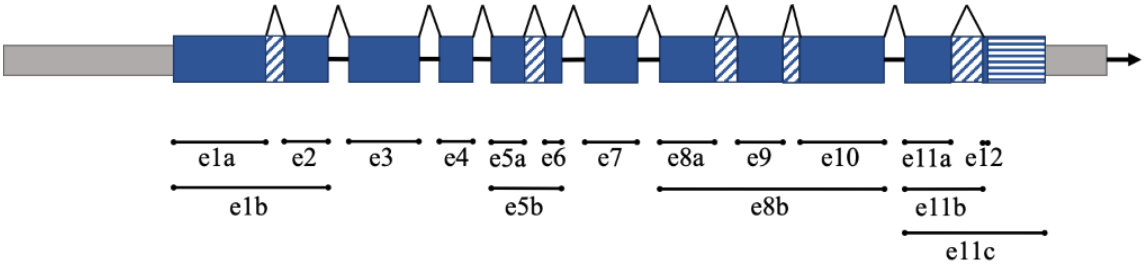

Figure 52. *CAS3* Super-Transcript Gene Model. The gene transcript model represents all possible paths of translation as supported by the RNA transcriptome data. The retention of introns 1, 5, 8, 9, and 11 is represented by the boxes with blue and white diagonal lines, forming exons 1b, 5b, 8b, and 11b. The increase in the length of exon 12 is represented by the box with blue and white horizontal lines, forming exon 11c since this reannotation occurred with the retention of intron 11.

PKA1

The FungiDB, PASA, and original genome transcript models for *PKA1* show that there is an intron within the 5' untranslated region in the FungiDB gene transcript model (Figure 53a) that is not present in the PASA gene transcript model (Figure 53b). Four sets of reannotations were made. Set one suggested the retention of intron 3 (Figure 54), set two presented evidence for the simultaneous retention of introns 7 and 9 (Figure 55), set three presented evidence for the retention of intron 9 (Figure 56), and set four suggested evidence for the increase in the length of exon 10 (Figure 57). Four gene transcript models were created for each of the four sets of reannotations suggested by the RNA transcriptome data with one model being created for the retention of intron 3 from set one (Figure 58), the simultaneous retention of introns 7 and 9 from set two (Figure 59), the retention of intron 9 from set three (Figure 60), and the increase in the length of exon 10 from set four (Figure 61). The *PKA1* super-transcript is compiled of the reannotations present in the gene transcript models presented in Figures 58, 59, 60, and 61 (Figure 62).

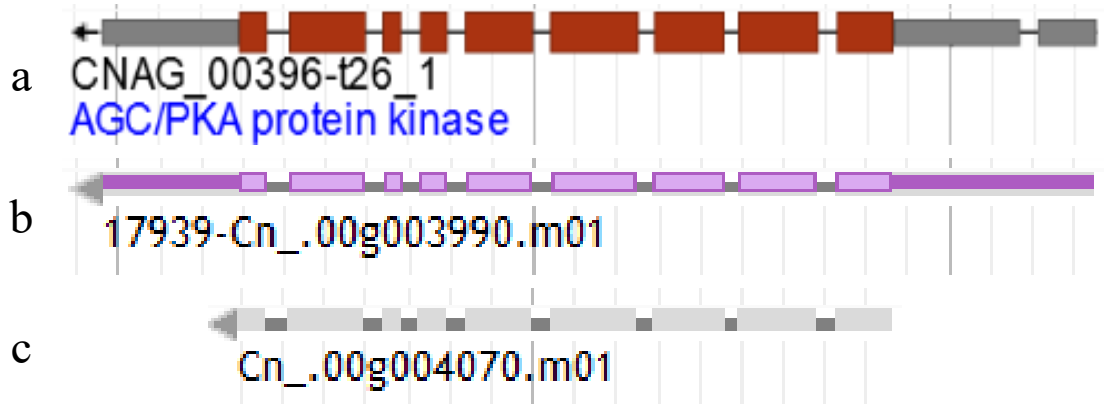

Figure 53. Reference Gene Transcript Models for *PKA1*. (a) The gene transcript model proposed by FungiDB. Note the presence of the intron in the 5' untranslated region. (b) The PASA algorithm-proposed gene transcript model for *PKA1*. Note the absence of the intron in the 5' untranslated region. (c) The gene transcript model proposed by the original genome.

Figure 54. *PKA1* Evidence of Intron 3 Retention. The exon evidence (dark green) suggests the retention of intron 3. B30-3 is the RNA transcriptome data that suggested retention.

Figure 55. *PKA1* Evidence of Introns 7 and 9 Simultaneous Retention. The exon evidence (dark green) suggests the retention of (b) intron 7 and (a) intron 9 occurring in the same evidence tracks simultaneously. B30-2 and B40-1 (not shown) are the RNA transcriptome data that suggest the retentions.

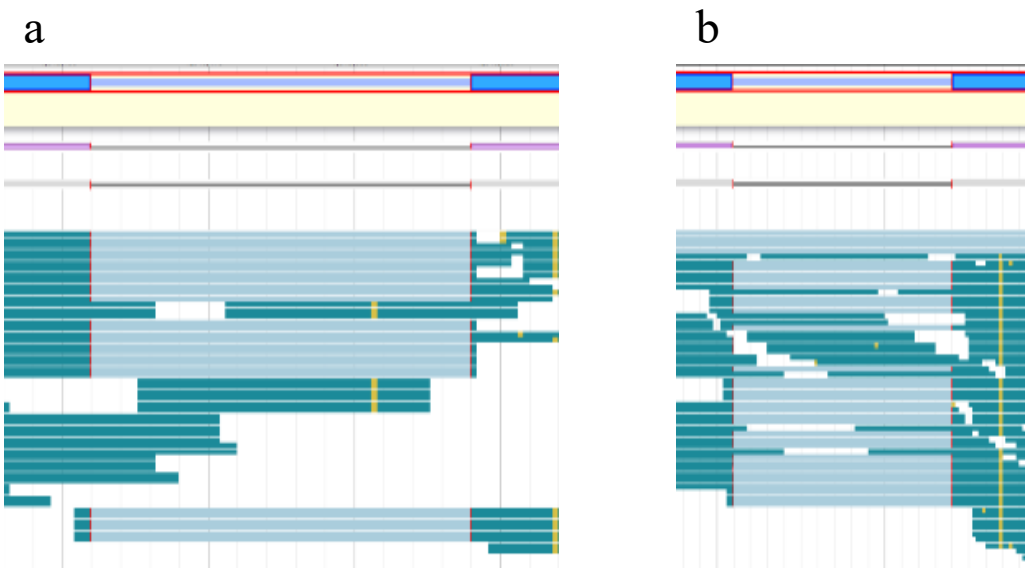

Figure 56. *PKA1* Evidence of Intron 9 Retention. The exon evidence (dark green) suggests the retention of intron 9. (a) B40-2 and (b) B58-1 are the RNA transcriptome data that suggest the retention.

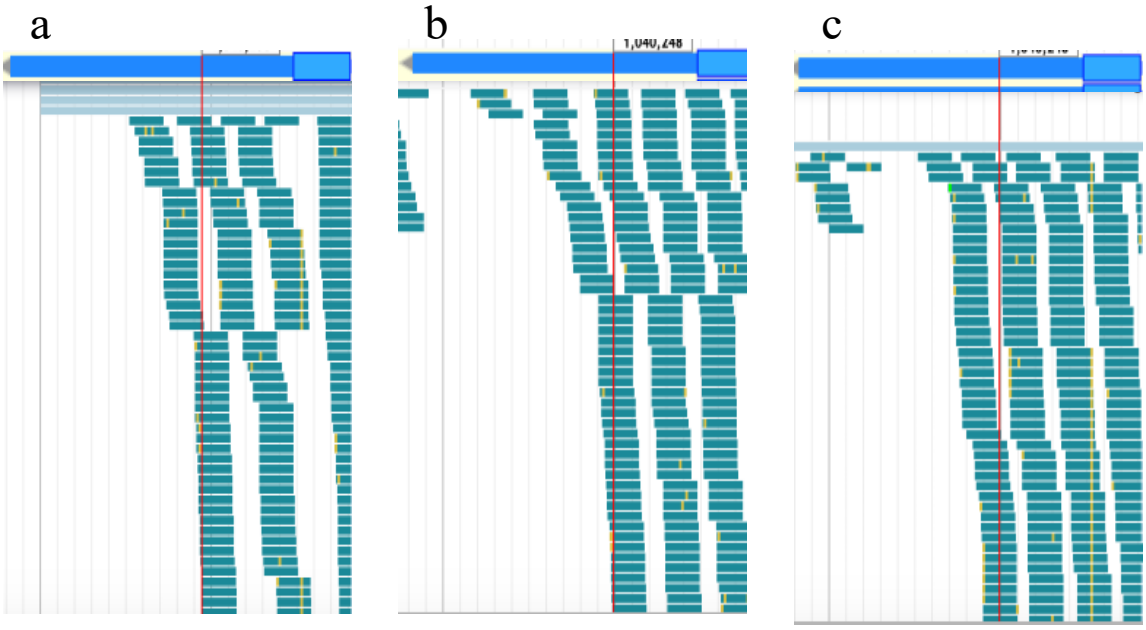

Figure 57. *PKA1* Evidence of Increase in Exon 10 Length. The exon evidence (dark green) within the 5' untranslated region suggests an increase in the length of exon 4. The RNA transcriptome data that suggest this reannotation are (a) B15-2, (b) B18-1, and (c) B30-2. B30-3 and B58-2 are also RNA transcriptome data that suggest the reannotation but are not shown.

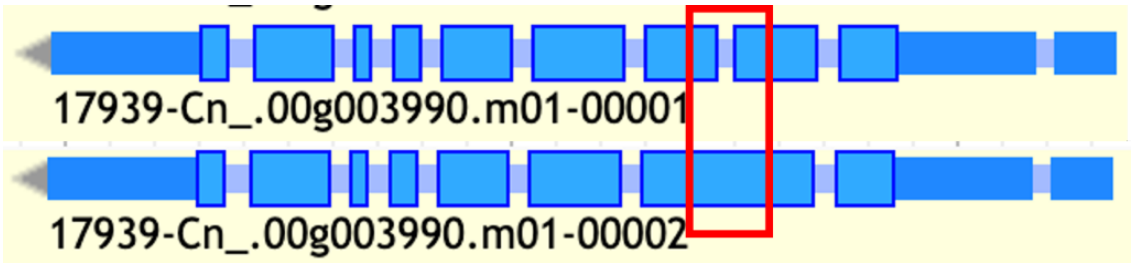

Figure 58. *PKA1* Gene Transcript Model Reannotation of Intron 3 Retention. The red rectangle indicates that intron 3 present in the original PASA gene transcript model (top) is retained in the modified gene transcript model (bottom), combining exons 3a and 4 to create exon 3b.

Figure 59. *PKA1* Gene Transcript Model Reannotation of Introns 7 and 9 Retention. The red rectangle to the right indicates that intron 7 present in the original PASA gene transcript model (top) is retained in the modified gene transcript model (bottom), combining exons 7a and 8 to create exon 7b. The red rectangle to the left indicates that intron 9 present in the original PASA gene transcript model (top) is retained in the modified gene transcript model (bottom), combining exons 9a and 10a to create exon 9b. These retentions occurred in the same evidence tracks simultaneously.

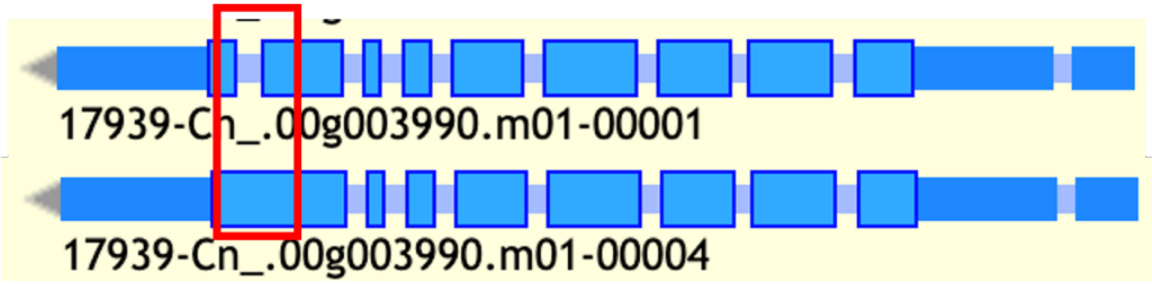

Figure 60. *PKA1* Gene Transcript Model Reannotation of Intron 9 Retention. The red rectangle indicates that intron 9 present in the original PASA gene transcript model (top) is retained in the modified gene transcript model (bottom), combining exons 9a and 10a to create exon 9b.

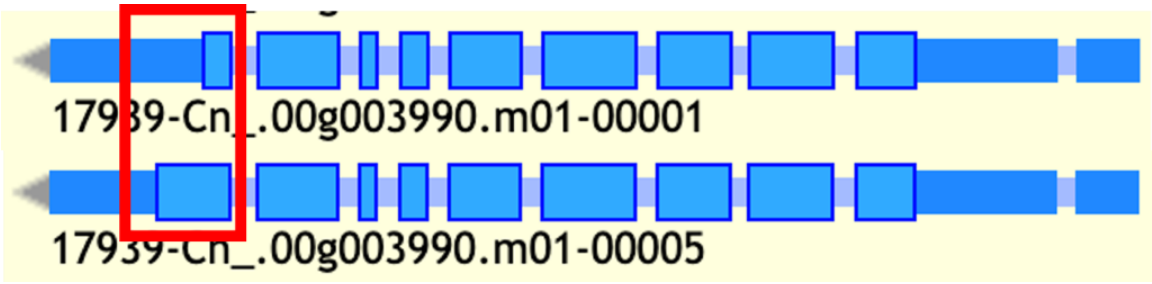

Figure 61. *PKA1* Gene Transcript Model Reannotation in the Increase in Length of Exon 10a. The red rectangle indicates the increase in the length of exon 10a between the original PASA gene transcript model (top) and the modified gene transcript model (bottom) to form exon 10b. This decreases the length of the 3' untranslated region.

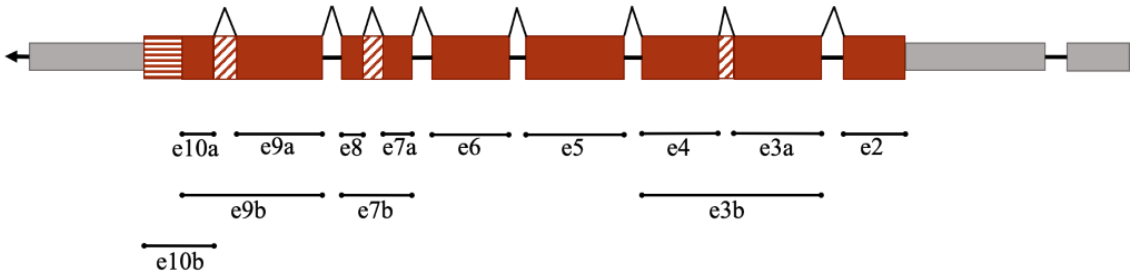

Figure. 62. *PKA1* Super-Transcript Gene Model. The RNA transcriptome data that support the reannotations for *PKA1* were used to create the gene transcript model displaying all possible paths. The retention of introns 3, 7, and 9 is represented by the boxes with red and white diagonal lines, forming exons 3b, 7b, and 9b. The increase in the length of exon 10a is represented by the box with red and white horizontal lines, forming exon 10b.

CONCLUSIONS

The ability to effectively and efficiently adapt to various conditions to favor the environment of its human host through various pathways makes *C. neoformans* important to understand and to find prevention for infection within immunocompromised individuals (Alspaugh 2015). The capsule is an important structure to *C. neoformans*, protecting it from the environment in unfavorable conditions (Alspaugh 2015). Being able to identify and understand the genes that cause the formation of the capsule in *C. neoformans* through their structure, function, and expression will allow for better understanding the virulence of *C. neoformans*.

In this study, improved gene transcript models for genes involved in capsule formation were created using RNA transcriptome data (E. McClelland, R. McFeeters, & J. Tucker, personal communication, March 18, 2019), the PASA algorithm (Haas et al. 2003), and the genome sequence for *C. neoformans* var. g*rubii* H99S for eight capsule-forming genes (Table 1): *CAP64*, *CAP59*, *CAP10*, *CAP60*, *CMT1*, *CAS1*, *CAS3*, and *PKA1*. Using RNA transcriptome data (Table 2), the computational gene transcript models proposed by the PASA algorithm were improved and reannotated for inconsistencies regarding exon skipping, intron retention, the increase or decrease in the length of exons, and the increase or decrease in the length of the 5' and 3' untranslated regions (Table 3). Twenty-three areas of inconsistencies were found and corrected in the newly created gene transcript models for the genes: twenty intron retentions and three increases in exon lengths (decreases in length of two 3' regions and one 5' region). With these reannotations, the accuracy of the gene transcript models was improved, and super-

transcript models displaying alternative splicing were created for each of the eight *C. neoformans* genes associated with capsule formation to better represent the gene and transcript structures and translation pathways for these genes.

The variation in alternative splicing and translation pathways suggested through the super-transcript models is specific to the conditions of the RNA transcriptome data (Table 2). These conditions resulted in the twenty intron retentions and three increases in exon length as stated previously. As a result, the proteins made through translation will vary depending on the conditions present for each gene.

Future Directions

In addition to the eight genes of focus in this study, there are other genes pertaining to capsule formation, including *CAC1*, *CAP1*, *CAS2*, *CAS4*, *CAS5*, *CAS6*, and *GPA1* (Bose et al. 2003). The improvement in the accuracy of the gene transcript models representing the eight genes of focus in this study will provide a better understanding and information in future studies pertaining to the virulence of *C. neoformans*. With further experimentation, these improved gene transcript models can allow for reconstruction and translation of RNA, determination of differences between proteins, and prediction of functional changes to proteins based on their domain functions.

LITERATURE CITED

Alspaugh JA. 2015. Virulence mechanisms and Cryptococcus neoformans pathogenesis. Fungal Genet Biol. 78:55–58. doi:10.1016/j.fgb.2014.09.004.

Basenko E, Pulman J, Shanmugasundram A, Harb O, Crouch K, Starns D, Warrenfeltz S, Aurrecoechea C, Stoeckert C, Kissinger J, Roos D, Hertz-Fowler C. 2018. FungiDB: An Integrated Bioinformatic Resource for Fungi and Oomycetes. J Fungi. 4(1):39. doi:10.3390/jof4010039. [accessed 2020 Aug 29]. http://www.mdpi.com/2309-608X/4/1/39.

Bose I, Reese AJ, Ory JJ, Janbon G, Doering TL. 2003. A yeast under cover: The capsule of Cryptococcus neoformans. Eukaryot Cell. 2(4):655–663. doi:10.1128/EC.2.4.655-663.2003.

Buchanan KL, Murphy JW. 1998. What makes Cryptococcus neoformans a pathogen? Emerg Infect Dis. 4(1):71–83. doi:10.3201/eid0401.980109.

Chang YC, Kwon-Chung KJ. 1994. Complementation of a capsule-deficient mutation of Cryptococcus neoformans restores its virulence. Mol Cell Biol. 14(7):4912–4919. doi:10.1128/mcb.14.7.4912. [accessed 2020 Feb 1]. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC358863/.

Chang YC, Penoyer LA, Kwon-Chung KJ. 1996. The Second Capsule Gene of Cryptococcus neoformans, CAP64, Is Essential for Virulence. [accessed 2020 Feb 1]. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC174025/.

Grijpstra J, Tefsen B, van Die I, de Cock H. 2009. The Cryptococcus neoformans cap10 and cap59 mutant strains, affected in glucuronoxylomannan synthesis, differentially activate human dendritic cells. FEMS Immunol Med Microbiol. 57(2):142–150. doi:10.1111/j.1574-695X.2009.00587.x. [accessed 2020 Feb 1]. https://academic.oup.com/femspd/article-lookup/doi/10.1111/j.1574-695X.2009.00587.x.

Haas BJ, Delcher AL, Mount SM, Wortman JR, Smith RK, Hannick LI, Maiti R, Ronning CM, Rusch DB, Town CD, Salzberg S, White O. 2003. Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies. Nucleic Acids Res. 31(19):5654–5666. doi:10.1093/nar/gkg770. [accessed 2020 Aug 29]. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC206470/.

Howe KL, Contreras-Moreira B, De Silva N, Maslen G, Akanni W, Allen J, Alvarez-Jarreta J, Barba M, Bolser DM, Cambell L, Carbajo M, Chakiachvili M, Christensen M, Cummins C, Cuzick A, Davis P, Fexova S, Gall A, George N, Gil L, Gupta P, Hammond-Kosack K, Haskell E, Hunt S, Jaiswal P, Janacek S, Kersey P, Langridge N, Maheswari U, Maurel T, McDowall M, Moore B, Muffato M, Naamati G, Naithani S, Olson A, Papatheodorou I, Patricio M, Paulini M, Pedro H, Perry E, Preece J, Rosello M, Russell M, Sitnik V, Staines D, Stein J, Tello-Ruiz M, Trevanion S, Urban M, Wei S, Ware D, Williams G, Yates A, Flicek P. 2019. Ensembl Genomes 2020-enabling non-vertebrate genomic research. Nucleic Acids Res. 48(D1):D689–D695. doi:10.1093/nar/gkz890. [accessed 2020 Aug 29]. http://plants.ensembl.org.

Humann JL, Lee T, Ficklin S, Main D. 2019. Structural and functional annotation of eukaryotic genomes with GenSAS. In: Methods in Molecular Biology. Vol. 1962. Humana Press Inc. p. 29–51. [accessed 2020 Aug 29]. https://doi.org/10.1007/978-1-4939-9173-0_3,.

Janbon G, Himmelreich U, Moyrand F, Improvisi L, Dromer F. 2001. Cas1p is a membrane protein necessary for the O-acetylation of the Cryptococcus neoformans capsular polysaccharide. Mol Microbiol. 42(2):453–467. doi:10.1046/j.1365-2958.2001.02651.x. [accessed 2020 Feb 1]. http://doi.wiley.com/10.1046/j.1365-2958.2001.02651.x.

Karkowska-Kuleta J, Rapala-Kozik M, Kozik A. 2009. Fungi pathogenic to humans: molecular bases of virulence of Candida albicans, Cryptococcus neoformans and Aspergillus fumigatus. [accessed 2020 Feb 1]. http://psjd.icm.edu.pl/psjd/element/bwmeta1.element.bwnjournal-article-abpv56p211kz.

Mada PK, Alam MU. 2019. Cryptococcus (Cryptococcosis). StatPearls Publishing. [accessed 2020 Mar 7]. http://www.ncbi.nlm.nih.gov/pubmed/28613714.

Moyrand F, Chang YC, Himmelreich U, Kwon-Chung KJ, Janbon G. 2004. Cas3p belongs to a seven-member family of capsule structure designer proteins. Eukaryot Cell. 3(6):1513–1524. doi:10.1128/EC.3.6.1513-1524.2004.

Nielsen K, Cox GM, Litvintseva AP, Mylonakis E, Malliaris SD, Benjamin DK, Giles SS, Mitchell TG, Casadevall A, Perfect JR, Heitman J. 2005. Cryptococcus neoformans α strains preferentially disseminate to the central nervous system during coinfection. Infect Immun. 73(8):4922–4933. doi:10.1128/IAI.73.8.4922-4933.2005.

O'Meara TR, Andrew Alspaugh J. 2012. The Cryptococcus neoformans capsule: A sword and a shield. Clin Microbiol Rev. 25(3):387–408. doi:10.1128/CMR.00001-12.

Odom A. 1997. Calcineurin is required for virulence of Cryptococcus neoformans. EMBO J. 16(10):2576–2589. doi:10.1093/emboj/16.10.2576. [accessed 2020 Feb 1]. http://emboj.embopress.org/cgi/doi/10.1093/emboj/16.10.2576.

Del Poeta M. 2004. Role of phagocytosis in the virulence of Cryptococcus neoformans. Eukaryot Cell. 3(5):1067–1075. doi:10.1128/EC.3.5.1067-1075.2004.

Zaragoza O, Rodrigues ML, De Jesus M, Frases S, Dadachova E, Casadevall A. 2009. Chapter 4 The Capsule of the Fungal Pathogen Cryptococcus neoformans. Adv Appl Microbiol. 68:133–216. doi:10.1016/S0065-2164(09)01204-0.