Evidence-based Annotation Revision to Genes Involved in the Virulence Factor Melanin Production in Fungal Pathogen, *Cryptococcus neoformans*

by

Alaa Fatima Mohammed

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.

Spring 2021

Evidence-based Annotation Revision to Genes Involved in the Virulence Factor Melanin Production in Fungal Pathogen, *Cryptococcus neoformans*

Alaa Fatima Mohammed

APPROVED:

_____

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

_____

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

_____

Dr. Stephen Wright, Second Reader
Professor, Biology Department

_____

Dr. John R. Vile
University Honors College Dean

ACKNOWLEDGEMENTS

I would like to thank everyone who played a role in my academic accomplishments because I know it could not have been possible without them. I want to thank Dr. Rebecca Seipelt-Thiemann for her mentorship and patience over the past 3 years. I am grateful for her allowing me to be part of her research team and having had the opportunity to conduct this research. I want to thank Dr. Stephen Wright and Dr. Dennis Mullen for their continuous guidance and taking the time to read my thesis. Most of all, I want to thank my family for supporting me with their love and encouragement every day.

ABSTRACT

Cryptococcus neoformans is a fungal pathogen that causes infection in immunosuppressed individuals. Melanin, a charged pigment polymer, is one of many contributing factors to the virulence of *C. neoformans*. Nine genes have been implicated in melanin production in *C. neoformans*. The goal of this study was to validate and/or improve computationally-predicted gene models for seven of these genes through the evaluation of transcriptome evidence. Improvements to gene models were curated in a collaborative annotation space and the data used to create refined super-transcript gene models included aspects of gene regulation, such as alternative transcription starts, alternative splicing, and alternative transcription termination. Computational gene models were validated for four genes (*LAC2*, *PKA*, *CLC1*, and *VAD1*) and computational models were improved for three genes (*LAC1, CHK1,* and *AGC/PKA*). This study, therefore, has produced a more accurate resource for the *Cryptococcus* community studying virulence factors.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

INTRODUCTION

Cryptococcus neoformans is a common opportunistic fungal pathogen that primarily infects immunosuppressed individuals such as HIV patients and organ transplant recipients (Vechi et al. 2019; Sabiiti & May, 2012; Loyse et al. 2013). *Cryptococcus neoformans* is most commonly associated with meningitis, but can infect almost any human organ including the brain by crossing the blood-brain barrier (Vechi et al. 2019; Yang et al. 2017). Approximately 1 million cases of cryptococcal meningitis are reported annually, with a mortality rate >75% and most cases occurring in sub-Saharan Africa (Yang et al. 2017; Sabiiti & May, 2012; Loyse et al. 2013; Rajasingham et al. 2017).

The primary form of contraction is through inhalation of desiccated yeast or spores from the environment which can result in no disease due to an effective host immune response, fungal latency/dormancy, or primary pulmonary infection followed by dissemination (Srikanta et al. 2014; Sabiiti & May, 2012). Once disseminated, the organism is carried in the bloodstream to other organs, including lung, skin, bone marrow, and most commonly the brain (Srikanta et al. 2014; Yang et al. 2017; Sabiiti & May, 2012). *Cryptococcus neoformans* has evolved many virulence factors to reproduce and survive the immune response of the human host including formation of a complex polysaccharide capsule, nutrient uptake, growth at 37°C, urease expression, and melanin production (Steenbergen & Casadevall, 2003; Sabiiti & May, 2012; Pettit et al. 2010).

The capsule, which is a large layer of polysaccharide surrounding the yeast, allows the pathogen to survive against phagocytosis and drying, and provides attachment to host (Sabiiti & May, 2012). Another virulence factor of *C. neoformans* is nutrient

uptake. For the pathogen to survive, it must compete with the host for certain nutrients like copper which plays a role in capsule and melanin synthesis (Sabiiti & May, 2012). Copper uptake is believed to be through the copper sensing/transport system. This is because there is a high concentration of Ctr4, the copper transport protein, found in macrophage-internalized and brain isolated cryptococci, suggesting that copper uptake is essential for *C. neoformans* intracellular survival and dissemination to the brain (Sabiiti & May, 2012). In the laboratory, the optimal growth temperature is 30°C, but *C. neoformans* can grow over a wide temperature range including the human body temperature of 37°C (98.6°F) (Pettit et al. 2010). Most clinical *C. neoformans* isolates are urease positive and have an increased pathogenicity compared to urease negative strains (Steenbergen & Casadevall, 2003). Urease is an enzyme that hydrolyzes urea to ammonia and carbamate which increases the pH level (Steenbergen & Casadevall, 2003). Melanin is a charged polymer pigment made by a laccase (Sabiiti & May, 2012). There is some speculation on how melanin enhances *C. neoformans* pathogenicity, but it has been extensively studied in recent years (Steenbergen & Casadevall, 2003). Melanized cells were more resistant to host immune response mechanisms such as phagocytosis and cell death by immune effector cells, monocytes and neutrophils, in vitro (Steenbergen & Casadevall, 2003; Williamson, 1997; Sabiiti & May, 2012). Melanin protects the yeast cell from phagocytosis in vivo and in vitro (Nosanchuk & Casadevall, 2006) and melanin has antioxidant properties that defend from oxidative killing by phagocytes (Sabiiti & May, 2012; Steenbergen & Casadevall, 2003). Finally, melanin confers resistance to some antifungal drugs by binding and absorbing the drug (Sabiiti & May, 2012; Nosanchuk & Casadevall, 2006).

Despite plentiful and steadily progressing research regarding the host-pathogen biology of cryptococcal infection, few studies have approached understanding and learning more about the fungus at the genomic level or systems biology. The first requirements for such an undertaking include a sequenced genome and an accurate and functional genome annotation. While computational predictions of gene structure and function are improving, accurate gene and genome annotation requires biological support via evidence of transcription in cells rather than computation recognition of sequence motifs. Additionally, for eukaryotes that exhibit alternative transcription starts, alternative splicing, and alternative transcription termination/polyadenylation, definition of multiple transcripts must be incorporated into the gene and genome model annotation.

To begin the process of providing a more accurately annotated genome for the *C. neoformans* research community as well as learn more about the organism's genome structure, gene structure, and gene regulation features, a structure genome annotation project was begun using an existing genome sequence, a computationally-predicted gene annotation, new computational prediction tools, and several sets of transcriptome data. As a first step in this process, the focus of this study was genes related to virulence factor melanin production: *LAC1*, *LAC2*, *PKA*, *AGC/PKA*, *CHK1*, *CLC1*, and *VAD1*. Two additional genes, *VPH1* and *RAD53*, were not included in this study because another researcher annotated the former, and the latter had an unclear genomic location.

LAC1 and *LAC2* are adjacent to one another and encode laccases, the enzyme that catalyzes a rate-limiting step in diphenol oxidation and melanin production. However, despite the two genes sharing 75% nucleotide identity, there are some differences between the genes (Pukkila-Worley et al. 2005). For instance, *LAC2* basal transcript

levels are significantly lower than *LAC1* levels (Pukkila-Worley et al. 2005). *PKA* (Protein Kinase A) and *AGC/PKA* play an important role in organizing phenotypic changes, such as melanin deposition in the cell wall, that directly impact disease development (Caza & Kronstad, 2019). Mutated *PKA* resulted in reduced melanin and less virulence (Kronstad et al. 2011). *CHK1*, Checkpoint Kinase-1, pertains to effector kinases involved in DNA damage repair (Jung et al. 2019). When *CHK1* was disrupted in *C. neoformans*, the virulence was attenuated, and it allowed for increased susceptibility to antifungal drugs (Jung et al. 2019). *CLC1* encodes chloride ion channel proteins for maintaining ion homeostasis and cellular machinery of melanin production (Walton et al. 2005). The *VAD1* (Virulence-Associated DEAD Box Protein-1) gene is an RNA helicase and regulates laccase transcription (Walton et al. 2005).

With these biologically important genes as the first illustration, the goal of this project was to use experimental RNA sequencing data and an improved gene prediction algorithm, Program to Assemble Spliced Alignment (PASA) (Humann et al. 2019), within the genome annotation platform GenSAS (Humann et al. 2019) to improve gene and transcript accuracy in a community-based curation space (Apollo) for genes implicated in melanin production in fungal pathogen, *Cryptococcus neoformans* (Humann et al. 2019).

MATERIALS AND METHODS

Cryptococcus neoformans Genome and Melanin-coding Genes

The genome sequence and first annotation for *C. neoformans var. grubii* H99 (GCA_000149245) was obtained from the EnsemblFungi database (Howe et al. 2019). A list of genes related to melanin production in *C. neoformans* was identified from the literature: *LAC1*, *LAC2*, *PKA*, *AGC/PKA*, *CHK1*, *CLC1*, and *VAD1* (Pukkila-Worley et al. 2005; Caza & Kronstad, 2019; Jung et al. 2019; Walton et al. 2005). The coordinates of the seven melanin associated genes in this study, were identified in EnsemblFungi (Howe et al. 2019) or FungiDB (Basenko et al. 2018) (Table 1).

Table 1. Seven *C. neoformans* Melanin-Production Genes Used in This Study

Gene Name	CNAG ID[1]	Coordinates[1]	Function
LAC1 ***(Laccase-1)***	02087	Chromosome 6: 1,232,467 - 1,234,881	Encodes for acyl-CoA-dependent ceramide synthase (Howe et al. 2019)
LAC2 ***(Laccase-2)***	02086	Chromosome 6: 1,235,018 - 1,237,477	Encodes for acyl-CoA-dependent ceramide synthase (Howe et al. 2019)
PKA ***(Protein Kinase A)***	04162	Chromosome 9: 166,848 - 169,568	Organizes phenotypic changes, such as capsule and melanin, that directly impact disease development (Caza & Kronstad, 2019)
AGC/PKA ***(AGC/Protein Kinase A)***	00396	Chromosome 1: 1,039,957 - 1,042,941	Organizes phenotypic changes, such as capsule and melanin, that directly impact disease development (Caza & Kronstad, 2019)

CHK1 ***(Checkpoint Kinase-1)***	03167	Chromosome 8: 230,927 - 233,605	Regulates DNA repair genes and influences susceptibility to drugs (Jung et al. 2019)
CLC1 ***(Chloride Channel-1)***	07647	Chromosome 6: 1,051,970 – 1,056,522	Encode proteins for cellular machinery of melanin production (Walton et al. 2005)
VAD1 ***(Virulence-Associated DEAD Box Protein-1)***	01537	Chromosome 11: 225,540 - 228,664	Role in laccase transcription (Walton et al. 2005)

[1] Systematic name (CNAG ID number) and coordinates were identified using EnsemblFungi (Howe et al. 2019) and FungiDB (Basenko et al. 2018).

RNA Sequencing Data

Transcriptome data from three experiments were used in this study. The first set of data consisted of three replicates of seven clinical strains from four male and three female patients in Botswana and the reference strain H99S (E. McClelland, personal communications, March 18, 2019). The second set consisted of strains H99S *C. neoformans* treated with scytovirin, scytovirin and testosterone, or untreated (R. McFeeders, personal communications, March 18, 2019). The third set of data consisted of H99S treated with testosterone, estrogen, or ethanol (Tucker et al. 2020).

Table 2. Transcriptome RNA Sequencing Data Names and Sources.

Strain	Treatment[1]	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	
B30	N/A	Clinical Isolate Patient: Botswana Female	
B40	N/A	Clinical Isolate Patient: Botswana Male	
B43	N/A	Clinical Isolate Patient: Botswana Female	
B45	N/A	Clinical Isolate Patient: Botswana Male	
B58	N/A	Clinical Isolate Patient: Botswana Male	
H99S	N/A	Clinical Isolate Reference Strain (H99S_Clinical)	
H99S	N/A	Control Treatment Time: 4 h	R. McFeeters, personal communication, March 18, 2019
H99S	100 µg scytovirin	Treatment Time: 4 h (H99S_Sc100_RM)	
H99S	100 µg scytovirin and testosterone	Treatment Time: 4 h (H99S_Sc100_T_RM)	
H99S	Testosterone	(H99S_T_JT)	J. Tucker, personal communication, March 18, 2019
H99S	Estrogen	(H99S_E_JT)	

H99S	Ethanol	(H99S_v_JT)	

[1] Samples represent three biological replicates each.

GenSAS and Apollo

GenSAS (Genome Sequence Activation Server), a platform for collaborative structural genome annotation, was the primary tool used in this project for collaborating and visualizing the *C. neoformans* sequence and evidence (Humann et al. 2019). Transcriptome data, genome sequence, and an initial genome annotation were loaded into a project. As a first revision step, these data were used with a Program to Assemble Spliced Alignments (PASA) to generate an improved genome annotation (Haas et al. 2008). All these data and evidence were then visualized using Apollo, a genome annotation editor (Figure 1). Genomic regions, gene structures, and RNA evidence were examined to identify areas of incongruence supported by evidence in terminal and internal exon structure, intron structure, alternative splicing, and putative coding regions. Mismatches and incongruences were recorded. New revised gene transcript models were curated in the Apollo user space based on these lines of evidence and support (Humann et al. 2019).

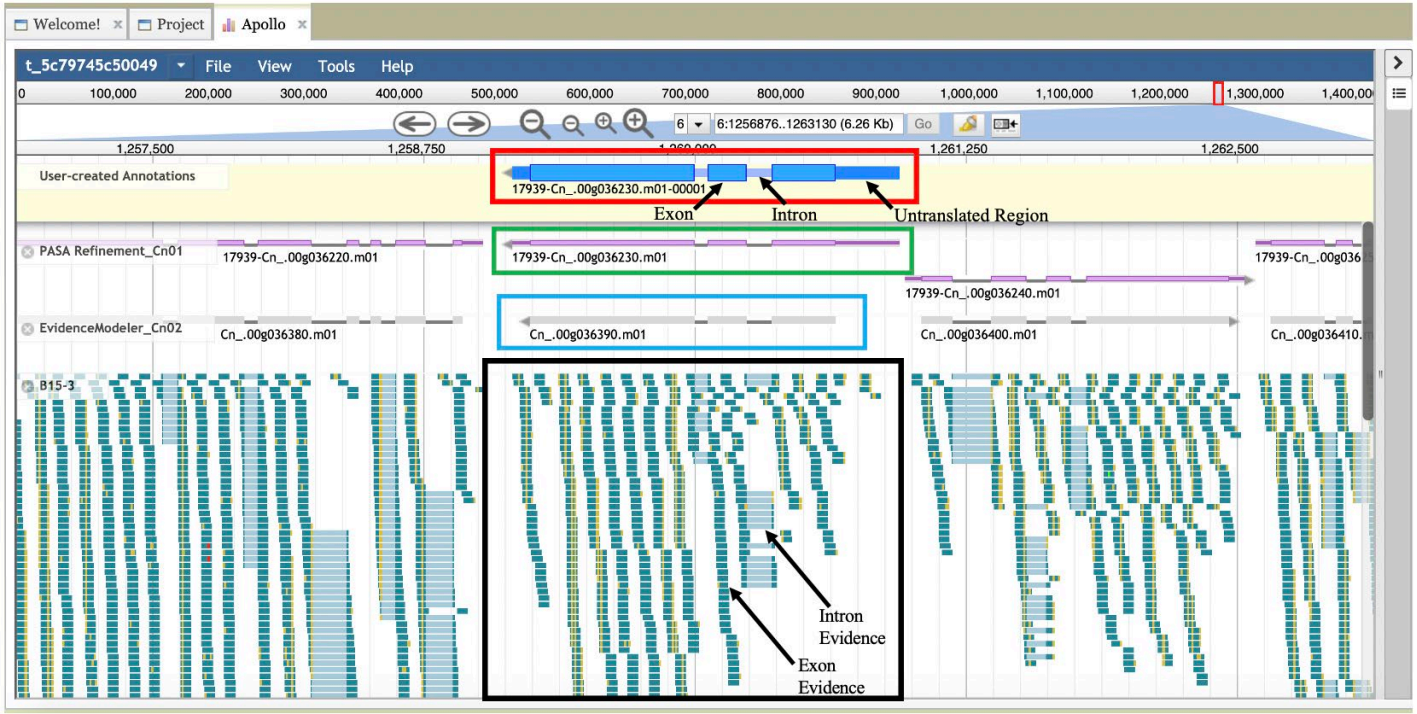

Figure 1. Apollo Genome Annotation Workspace Viewed in GenSAS. The red and green boxes highlight the same PASA Refinement_Cn01 gene transcript model but the red box shows the model in the annotation user workspace. The intron, exon, and untranslated regions of the gene are indicated with arrows in the red box. The blue box highlights the original gene transcript model

uploaded into GenSAS, called EvidenceModeler_Cn02. The black box shows the RNA transcriptome data, B15-3, for the selected gene with arrows indicating intron and exon evidence. The intron evidence is shown as light blue and exon evidence as dark green.

RESULTS

The purpose of this study is to correct the gene structure annotations for the seven genes involved in the virulence, melanin production in *C. neoformans* based on evidence within Apollo in GenSAS (Table 1) (Humann et al. 2019).

Comparing Gene Transcript Models

The Apollo user space is a collaborative genome annotation editor within GenSAS for curating gene transcript models based on analysis and evidence (Figure 1) (Humann et al. 2019). The transcriptome RNA data was analyzed for incongruencies and annotations were created for the gene transcript models based on comparison with: PASA Refinement_Cn01 (Figure 2c), EvidenceModeler_Cn02 (Figure 2d), and FungiDB or EnsemblFungi gene transcript models (Figure 2a). The new gene transcript models were created using the PASA model and adding annotations accordingly based on evidence in the Apollo user workspace (Figure 2b). Reannotations were made for exon skipping, intron retention, exon length variations, and length variations in untranslated regions in the gene transcript models (Figure 3).

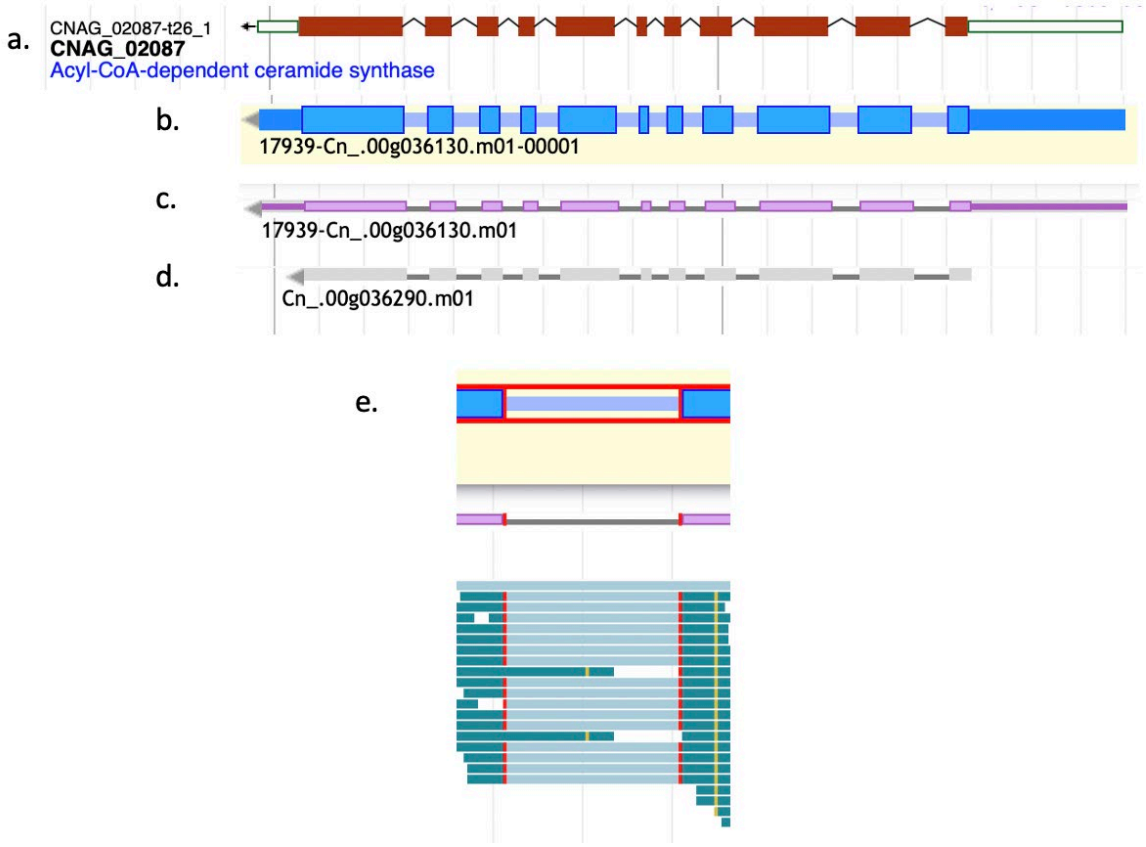

Figure 2. Example Gene Structures Used for Comparison. (a.) The *LAC1* gene structure proposed by FungiDB. (b.) The *LAC1* gene structure model proposed by PASA Refinement_Cn01 in the Apollo user workspace in GenSAS. (c.) The *LAC1* gene structure model proposed by PASA Refinement_Cn01. (d.) The *LAC1* gene structure model proposed by original EvidenceModeler_Cn02 in GenSAS. (e.) The RNA transcriptome data for a selected gene region showing intron evidence as light blue and exon evidence as dark green.

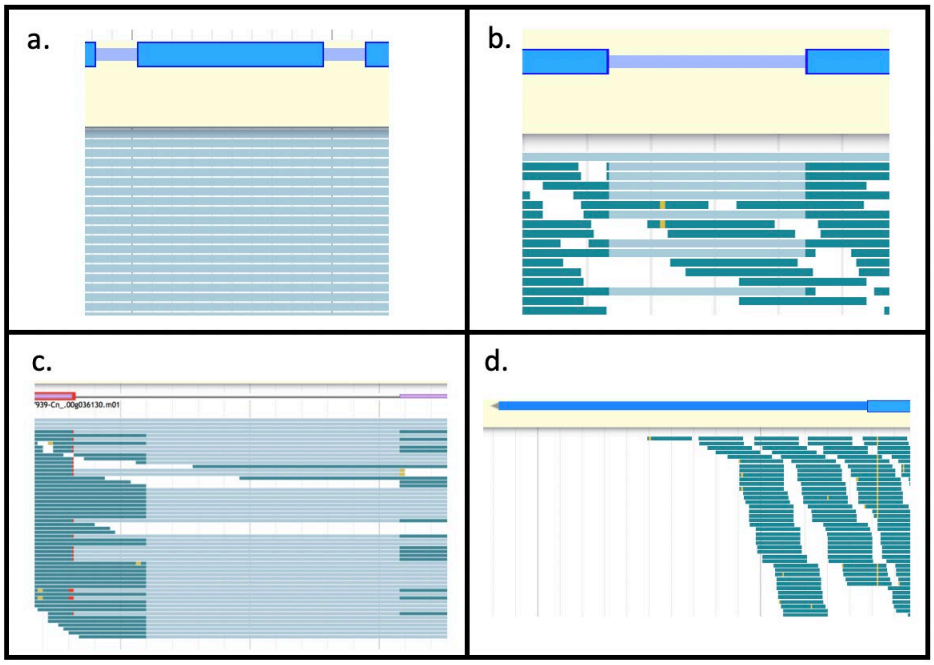

Figure 3. Incongruencies to Reannotate in Gene Structure. (a.) The above gene transcript model indicates an exon, but the light blue intron evidence below suggests exon skipping occurred here. (b.) The above gene transcript model indicated an intron, but the dark green exon evidence below suggests potential intron retention. There are roughly equal amounts of light blue intron and dark green exon evidence. (c.) The above gene transcript model indicated the exon length, but the dark green exon evidence below suggests the left exon is longer than the model proposes. (d.) The above gene transcript model indicates the untranslated region (thinner blue box) length, but the dark green exon evidence below suggests the exon is longer than proposed, resulting in the shortening of the untranslated region.

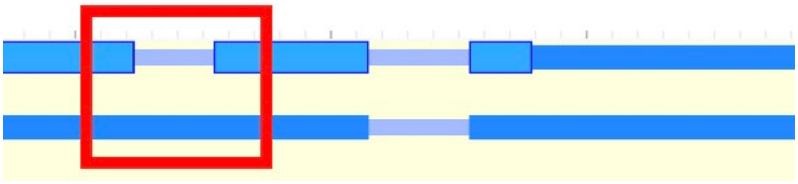

Figure 4. Example of Gene Reannotation. The red box shows a user modification made in the Apollo workspace. The top gene transcript model is the PASA Refinement_Cn01 model. The bottom model has the highlighted intron region retained because the RNA transcriptome data suggested this reannotation.

Model and Evidence Analysis

The initial comparison of the three gene transcript models (Figure 2a, c, d) showed the untranslated regions were absent in the original EvidenceModeler_Cn02 (Figure 2d) models. The untranslated regions were present in PASA Refinement_Cn01 (Figure 2c) models and FungiDB or EnsemblFungi gene transcript models (Figure 2a). The RNA transcriptome data (Table 2) were used to identify areas of incongruency in the PASA Refinement_Cn01 (Figure 2c) proposed gene models. Incongruencies needing reannotation were found in three of the seven virulent, melanin producing genes (Table 3). The incongruencies supported by RNA transcriptome data were indicated in new gene transcript models curated within the Apollo user space. The new gene transcript models were condensed into one super-transcript gene model for each of the three genes.

Table 3. Summary of Gene Reannotations. This is a summary of the genes, incongruencies, exon coordinates of reannotations, and supporting evidence from RNA transcriptome data.

Gene	Incongruency	Exon Coordinates	Data Track Evidence
LAC1	Retention of intron 1	Exon 1b: 6:1234136 – 1234881 (-) strand	B15-1, B43-2
LAC1	Retention of intron 1 and 2	Exon 1c: 6: 1233854 - 1234881 (-) strand	B15-2
LAC1	Retention of intron 1 and 2; exon 4 lengthened	Exon 1c: 6: 1233854 - 1234881 (-) strand Exon 4b: 6:1233702 - 1233804 (-) strand	B18-1
LAC1	Retention of intron 2 and 5	Exon 2b: 6:1233854 - 1234287 (-) strand Exon 5b: 6:1233526 - 1233648 (-) strand	B18-2
LAC1	Retention of intron 2, 5, 6, and 10	Exon 2b: 6:1233854 - 1234287 (-) strand	B18-3

		Exon 5c: 6:1233300 - 1233648 (-) strand Exon 10b: 6:1232467 - 1233009 (-) strand	
LAC1	Retention of intron 2	Exon 2b: 6:1233854 - 1234287 (-) strand	B45-3
CHK1	Retention of intron 9	Exon 9b: 8: 232933-233605 (+) strand	B30-2
AGC/PKA	Retention of intron 3	Exon 3b: 1:1041613-1042105 (-) strand	B30-3
AGC/PKA	Retention of introns 7 and 9	Exon 7b: 1:1040800-1040994 (-) strand Exon 9b: 1:1039957-1040749 (-) strand	B30-2
AGC/PKA	Retention of intron 9	Exon 9b: 1:1039957-1040749 (-) strand	B40-2, B58-1
AGC/PKA	Exon 10 lengthened	Exon 10b: 1:1039957- 1042105 (-) strand	B15-2, B18-1, B30-2, B30-3, B58-2

LAC1

The three reference gene transcript models for the *LAC1* gene were similar and only differed in the untranslated regions proposed by the original EvidenceModeler_Cn02 model (Figure 5). Based on the RNA transcriptome data, six reannotated models were curated in the Apollo user workspace starting with the PASA proposed model and adding modifications (Figure 12). The first model shows the retention of intron 1 that is supported by two data tracks (Figure 6). The second model shows the retention of introns 1 and 2 that is supported by one data track (Figure 7). The third model shows the retention of introns 1 and 2 with exon 4 increased in length (Figure 8). The fourth model shows the retention of introns 2 and 5 supported by one data track (Figure 9). The fifth model shows the retention of introns 2, 5, 6, and 10 (Figure 10). The final model shows the retention of only intron 2 (Figure 11). The six curated models were condensed into a super-transcript gene model (Figure 13).

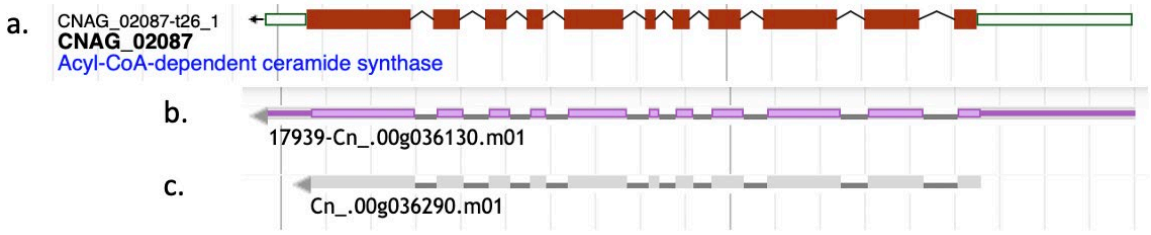

Figure 5. Reference Gene Models for *LAC1*. (a.) The *LAC1* gene structure proposed by FungiDB. (b.) The *LAC1* gene structure model proposed by PASA Refinement_Cn01. (c.) The *LAC1* gene structure model proposed by original EvidenceModeler_Cn02 in GenSAS.

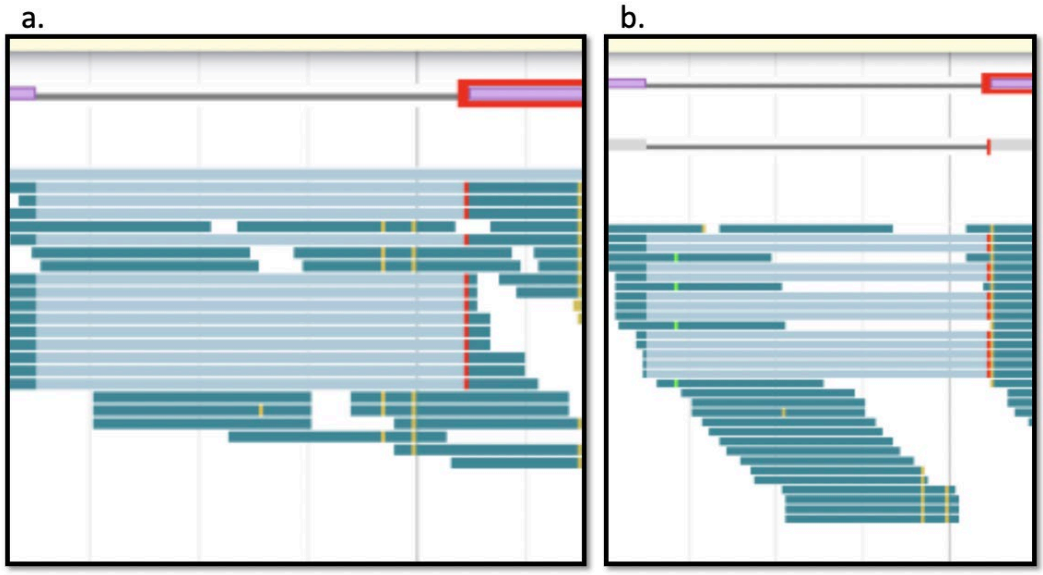

Figure 6. *LAC1* Evidence for Intron 1 Retention. The dark green exon evidence below intron 1 suggests retention. The RNA transcriptome data tracks indicating retention are (a.) B15-1 and (b.) B43-2.

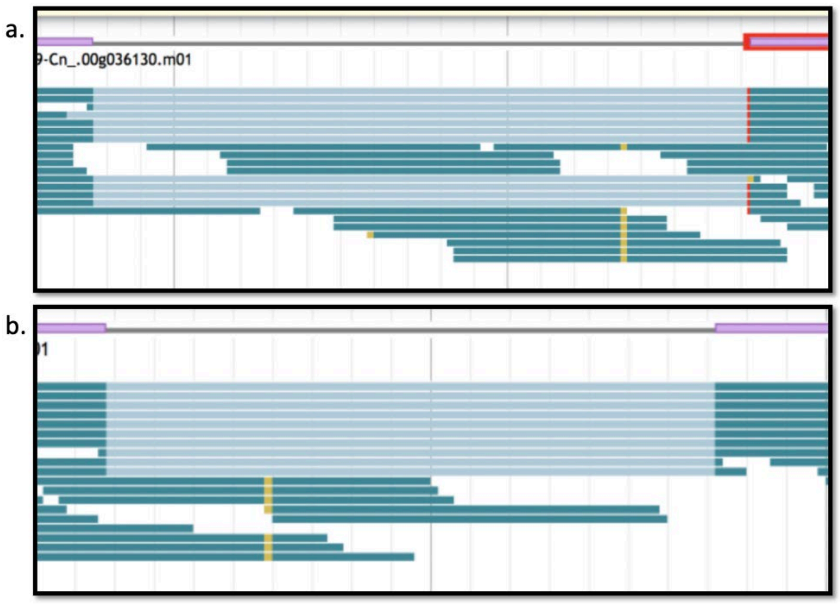

Figure 7. *LAC1* Evidence for Introns 1 and 2 Retention. (a.) The dark green exon evidence below intron 1 suggests retention. (b.) The dark green exon evidence below intron 2 suggests retention. The data track for this evidence is B15-2.

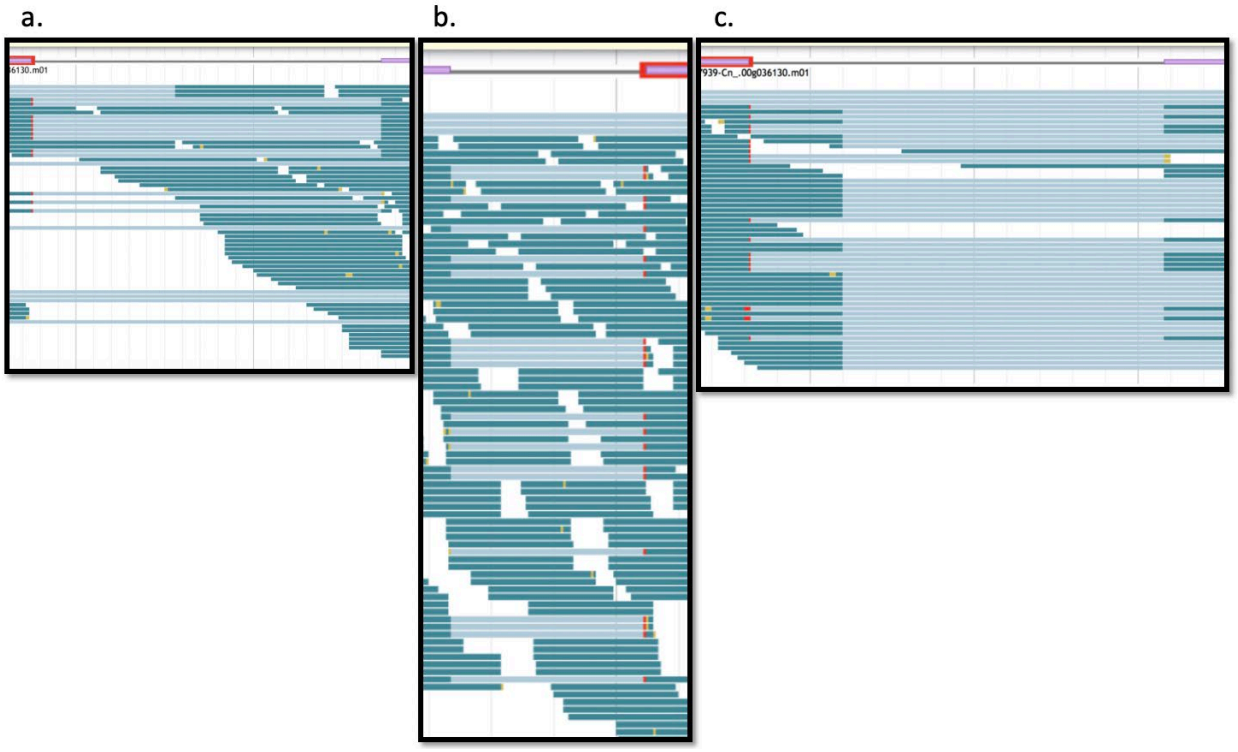

Figure 8. *LAC1* Evidence for Retention of Introns 1 and 2 with Exon 4 Lengthened. (a.) The dark green exon evidence below intron 1 suggests retention. (b.) The dark green exon evidence below intron 2 suggests retention. (c.) The dark green exon evidence below intron 3 suggests exon 4 is longer than PASA proposed. The evidence for these reannotations is RNA transcriptome data from the B18-1 strain.

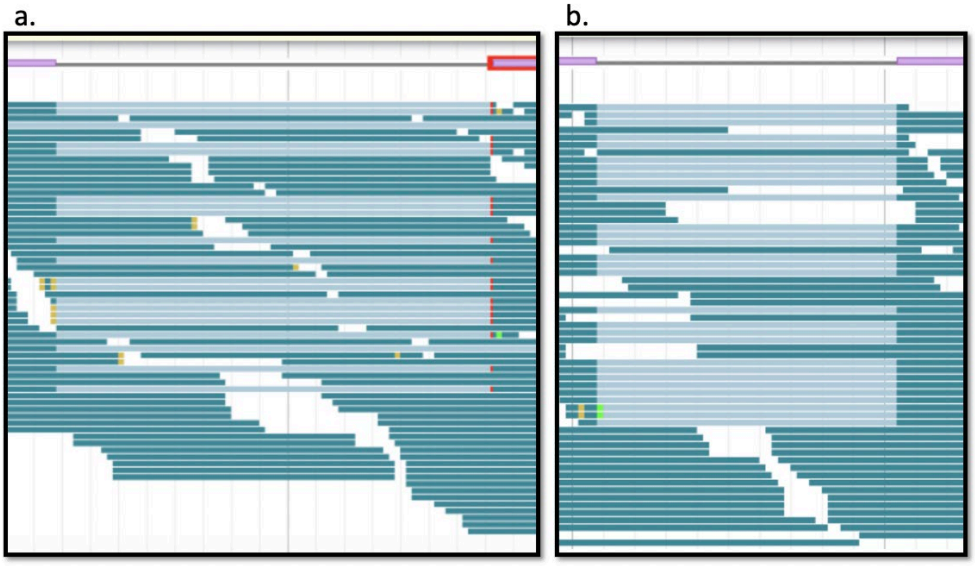

Figure 9. *LAC1* Evidence for Introns 2 and 5 Retention. (a.) The dark green exon evidence below intron 2 suggests retention. (b.) The dark green exon evidence below intron 5 suggests retention. The evidence for these reannotations is RNA transcriptome data from the B18-2 strain.

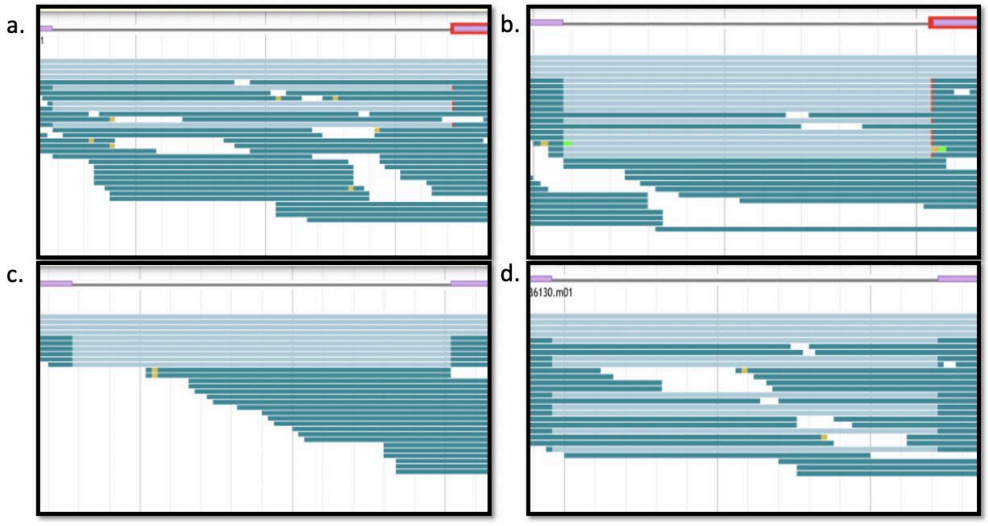

Figure 10. *LAC1* Evidence for Introns 2, 5, 6, and 10 Retention. (a.) The dark green exon evidence below intron 2 suggests retention. (b.) The dark green exon evidence below intron 5 suggests retention. (c.) The dark green exon evidence below intron 6

suggests retention. (d.) The dark green exon evidence below intron 10 suggests retention. The evidence for these reannotations is RNA transcriptome data from the B18-3 strain.

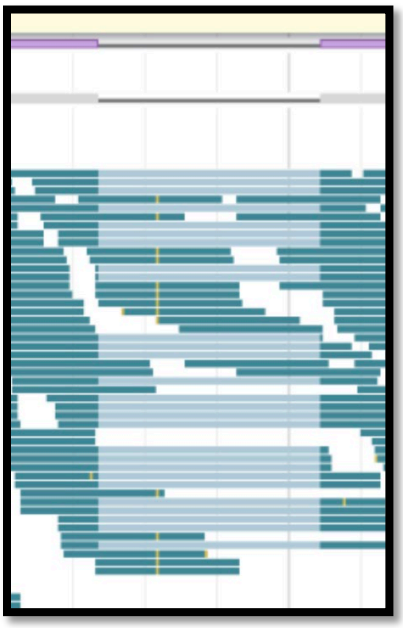

Figure 11. *LAC1* Evidence for Intron 2 Retention. The dark green exon evidence below intron 2 suggests retention. The evidence for this reannotation is RNA transcriptome data from B45-3.

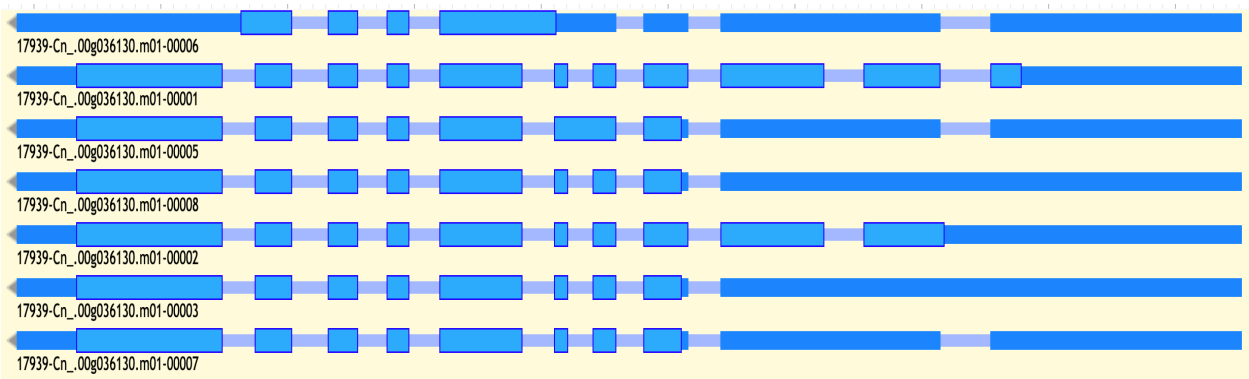

Figure 12. Six *LAC1* Curated Models and PASA Model. The *LAC1* incongruencies supported by RNA transcriptome data (Table 3) were modified within the Apollo user space to curate six potential models. The second gene transcript model from the top,

labeled 17939-Cn_.00g036130.m01-0001, is the original PASA proposed model with no reannotations (Figure 5b).

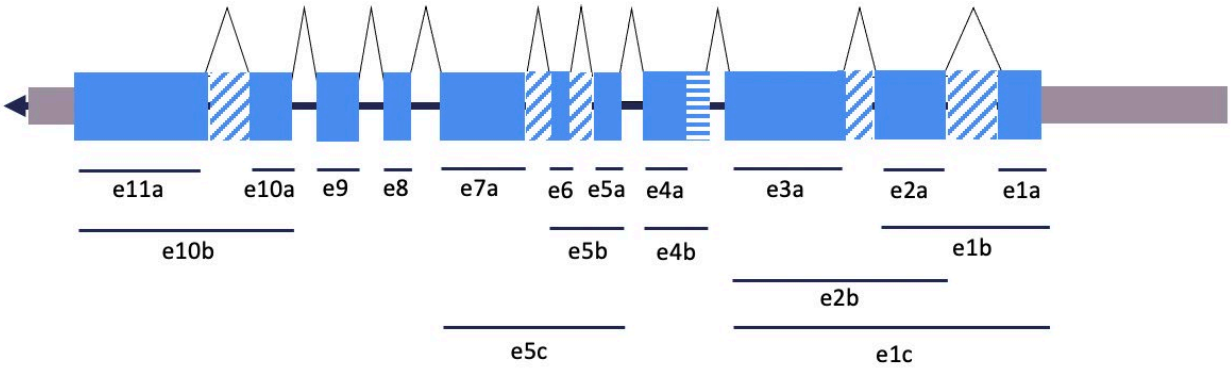

Figure 13. *LAC1* Super-transcript Gene Model. A super-transcript gene model was made to condense the six reannotated gene models (Figure 12) into one to show all the possible pathways. The retention of introns 1, 2, 5, 6, and 10 are indicated by the regions with diagonal stripes, forming exon 1b, 2b, 1c, 5b, 5c, and 10b. The increased length of exon 4 is indicated by the horizontal striped region, forming exon 4b.

CHK1

The three reference gene transcript models for the *CHK1* gene were similar and only differed in the untranslated regions proposed by the original EvidenceModeler_Cn02 model (Figure 14). Based on the RNA transcriptome data, one reannotated model was curated in the Apollo user workspace starting with the PASA proposed model and adding modifications (Figure 16). The reannotated model shows the retention of intron 9 that is supported by one data track (Figure 15). The curated model and PASA model were condensed into a super-transcript gene model (Figure 17).

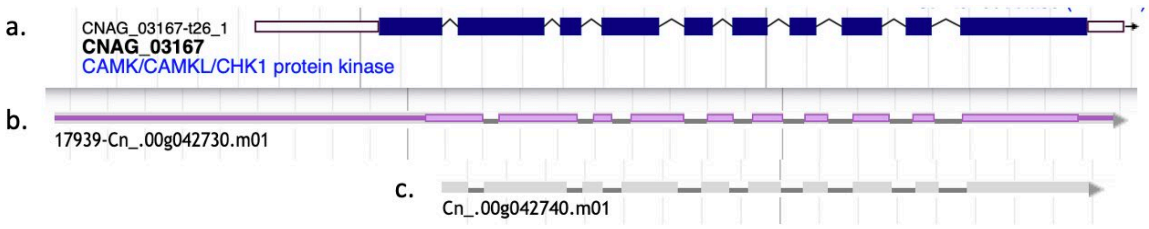

Figure 14. Reference Gene Models for *CHK1*. (a.) The *CHK1* gene structure proposed by FungiDB. (b.) The *CHK1* gene structure model proposed by PASA Refinement_Cn01. (c.) The *CHK1* gene structure model proposed by original EvidenceModeler_Cn02 in GenSAS.

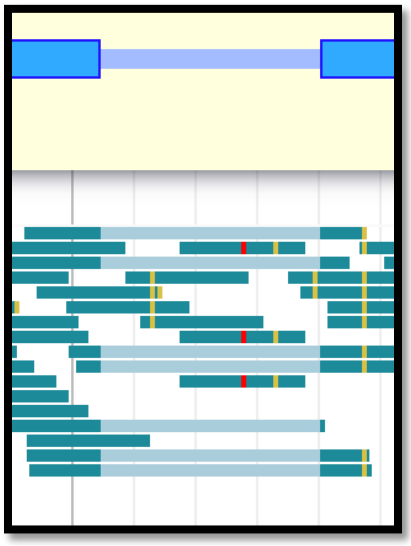

Figure 15. *CHK1* Evidence for Intron 9 Retention. The dark green exon evidence below intron 9 suggests retention. The evidence for this reannotation is RNA transcriptome data from B30-2.

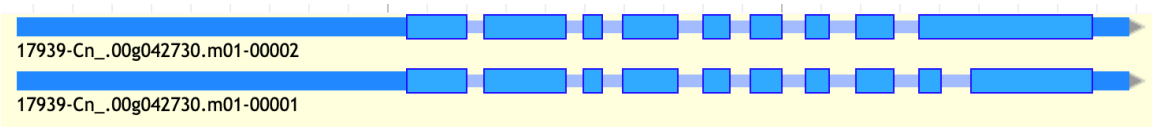

Figure 16. One *CHK1* Curated Model and PASA Model. The *CHK1* incongruency supported by RNA transcriptome data (Table 3) was modified within the Apollo user space to curate another potential model. The bottom gene transcript model, labeled 17939-Cn_.00g042730.m01-0001, is the original PASA proposed model with no reannotations (Figure 14b).

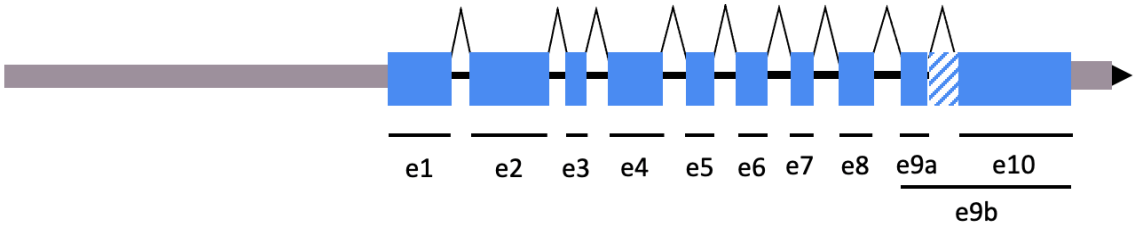

Figure 17. *CHK1* Super-transcript Gene Model. A super-transcript gene model was made to condense the reannotated gene model and PASA model into one to show the possible pathways (Figure 16). The retention of intron 9 is indicated by the region with diagonal stripes, forming exon 9b.

AGC/PKA

The three reference gene transcript models for the *AGC/PKA* gene differed in the untranslated regions (Figure 18). Unlike the FungiDB model, the PASA Refinement_Cn01 and EvidenceModeler_Cn02 proposed gene models do not include an intron within the 5' untranslated region (Figure 18). Based on the RNA transcriptome

data, four reannotated models were curated in the Apollo user workspace starting with the PASA proposed model and adding modifications (Figure 23). The first model shows the retention of intron 3 that is supported by one data track (Figure 19). The second model shows the retention of introns 7 and 9 that is supported by one data track (Figure 20). The third model shows the retention of intron 9 that is supported by two data tracks (Figure 21). The fourth model shows exon 10 increased in length (Figure 22). The four curated models were condensed into a super-transcript gene model (Figure 24).

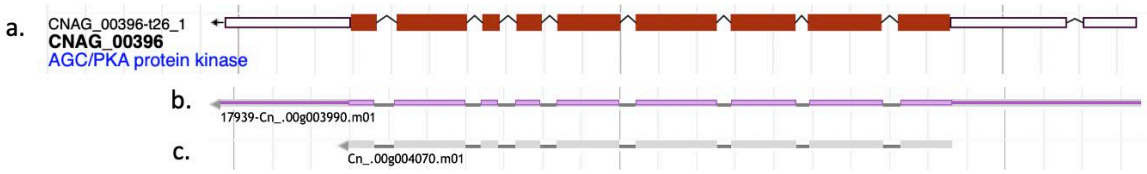

Figure 18. Reference Gene Models for *AGC/PKA*. (a.) The *AGC/PKA* gene structure proposed by FungiDB with the intron in the 5' untranslated region. (b.) The *AGC/PKA* gene structure model proposed by PASA Refinement_Cn01. (c.) The *AGC/PKA* gene structure model proposed by original EvidenceModeler_Cn02 in GenSAS.

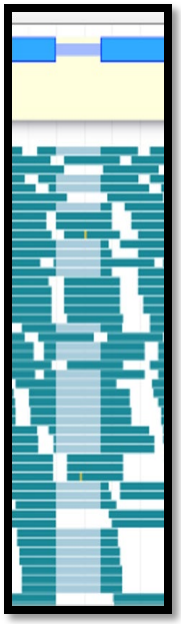

Figure 19. *AGC/PKA* Evidence for Intron 3 Retention. The dark green exon evidence below intron 3 suggests retention. The evidence for this reannotation is RNA transcriptome data from B30-3.

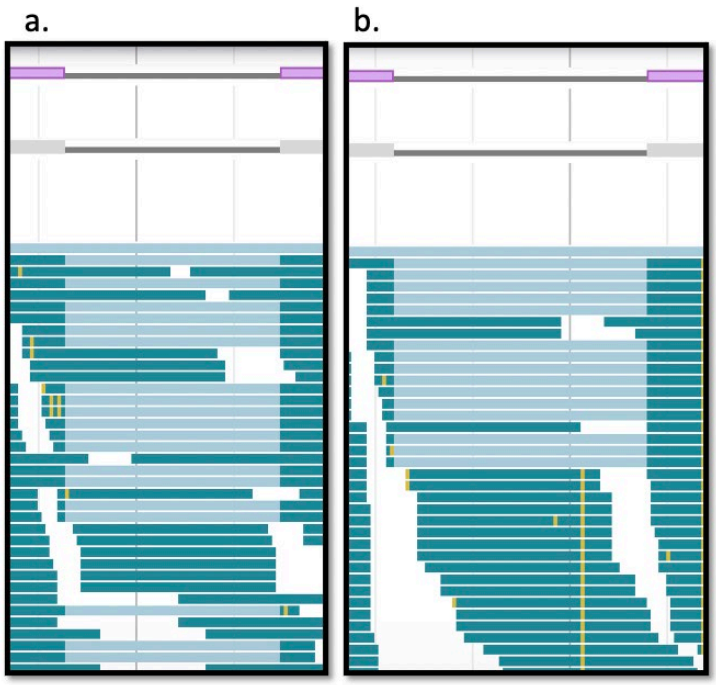

Figure 20. *AGC/PKA* Evidence for Retention of Introns 7 and 9. (a.) The dark green exon evidence below intron 7 suggests retention. (b.) The dark green exon evidence below intron 9 suggests retention. The evidence for these reannotations is RNA transcriptome data from B30-2.

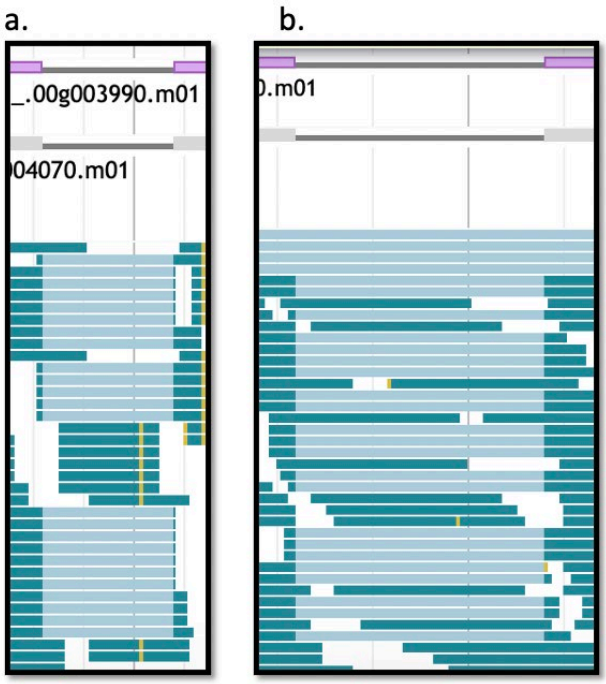

Figure 21. *AGC/PKA* Evidence for Retention of Intron 9. The dark green exon evidence below intron 9 suggests retention. The RNA transcriptome data tracks indicating retention are (a.) B40-2 and (b.) B58-1.

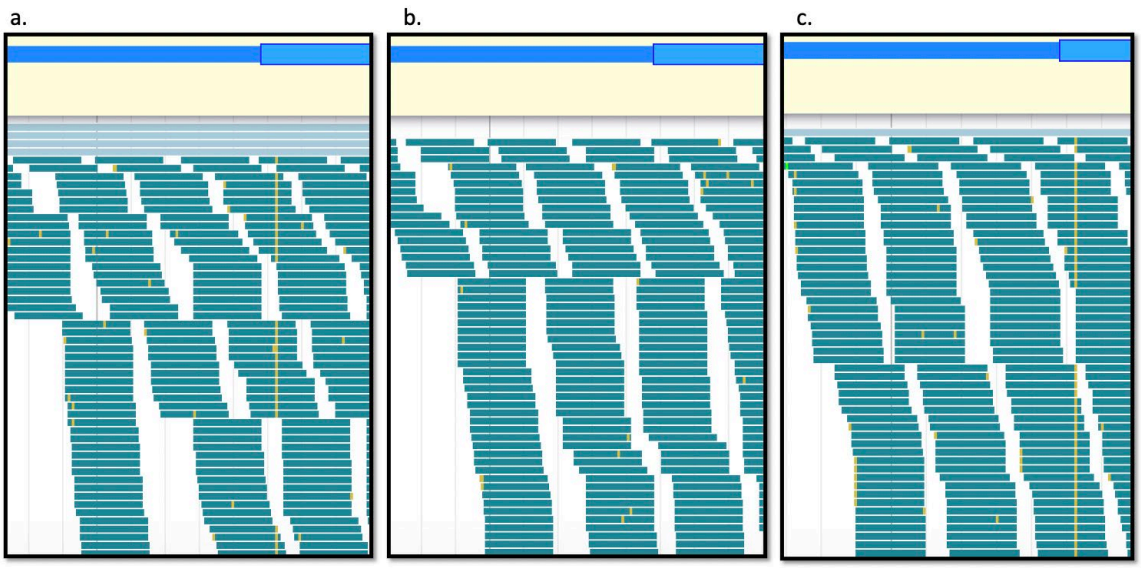

Figure 22. *AGC/PKA* Evidence for Lengthening Exon 10. The dark green exon evidence below the 3' untranslated region suggests exon 10 is longer than PASA proposed. The RNA transcriptome data tracks indicating lengthening are (a.) B15-2, (b.) B18-1, and (c.) B30-2. Data tracks B30-3 and B58-2 suggest the lengthening but are not shown.

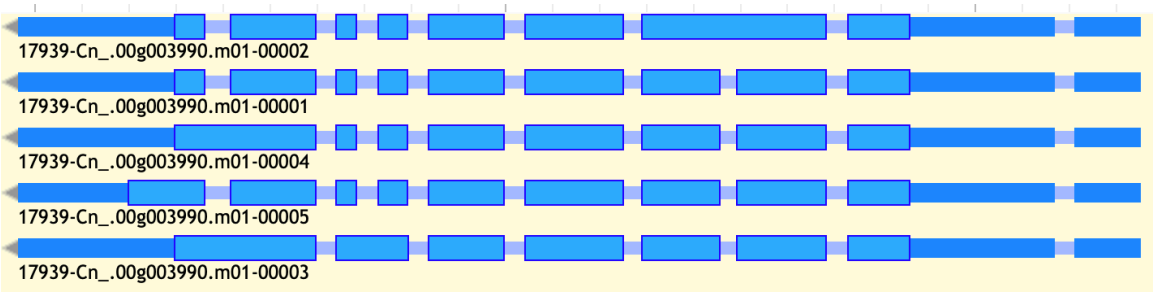

Figure 23. Four *AGC/PKA* Curated Models and PASA Model. The *AGC/PKA* incongruencies supported by RNA transcriptome data (Table 3) were modified within the Apollo user space to curate four potential models. The second gene transcript model from the top, labeled 17939-Cn_.00g003990.m01-0001, is the original PASA proposed model with the 5' untranslated region intron reannotation (Figure 18b).

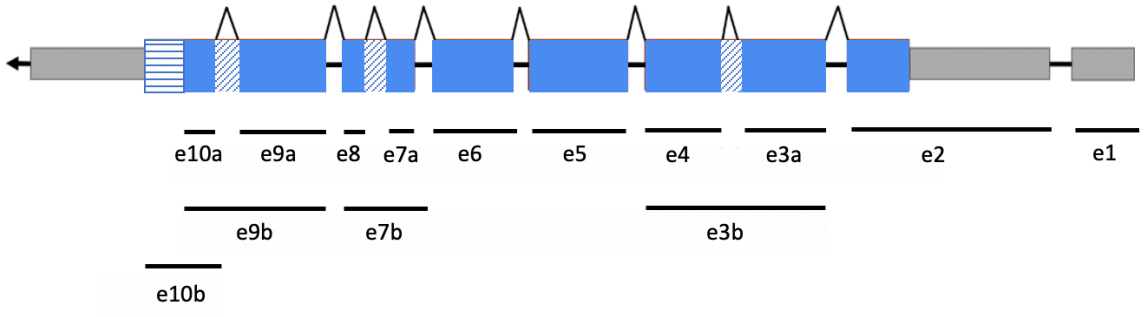

Figure 24. *AGC/PKA* Super-transcript Gene Model. A super-transcript gene model was made to condense the four reannotated gene models (Figure 23) into one to show all the possible pathways. The retention of introns 3, 7, and 9 are indicated by the regions with diagonal stripes, forming exon 3b, 7b, and 9b. The increased length of exon 10 is indicated by the horizontal striped region, forming exon 10b.

CONCLUSIONS

Studying pathogens at a genomic level allows for a better understanding on how to treat them effectively with different methods. For the virulence factor melanin, having an accurate genome could be beneficial for tracking potential new strains and drug resistance. In this study, improved gene transcript models were made for three of the seven melanin producing genes using manual evaluation of RNA transcriptome data (E. McClelland, R. McFeeters, & J. Tucker, personal communication, March 18, 2019) in combination with an improved genome generated via the PASA algorithm (Haas et al. 2003). PASA suggested gene models were first generated and then improved based on the transcriptome data. Eleven areas of incongruence were identified and reannotated to create new super gene transcript models that incorporate changes in terminal and internal exon structure, intron structure, alternative splicing, and putative coding regions. Changes included ten intron retentions and one area of increased exon length. The reannotations were combined to make one super-transcript model for each of the three genes with differences: *LAC1*, *CHK1*, and *AGC/PKA*. The super-transcript gene models show all the possible transcripts including alternative transcription starts, alternative splicing, and alternative transcription termination/polyadenylation.

The many virulence factors of *C. neoformans* allows for flexibility and adaptability to a wide range of conditions, making evasion of the innate human immune system more plausible (Yang et al. 2017). Gaining a better understanding of these virulence factors is important to prevent infection in immunocompromised people and those with less access to the necessary antifungal drugs in lower-income countries (Loyse et al. 2013). Melanin production enhances *C. neoformans* ability to resist antifungal drugs

and phagocytosis (Sabiiti & May, 2012; Nosanchuk & Casadevall, 2006). Studying the structure and function of melanin coding genes in *C. neoformans* will allow for a better understanding of the virulence at the genomic level.

Future Directions

The improved gene transcript models can be used to determine the differences in proteins translated and predict functional changes based on the presence or absence of protein domains. Additionally, while this study focused on seven genes related to melanin production other genes related to melanin production are known including *RAD53* and *VPH1* (Jung et al. 2019; Walton et al. 2005). Future experiments should include similar analysis of the remaining genes, other virulence factor genes, and finally the entire genome to provide a fully evidence based structural genome annotation that can then be used to begin a study for functionally annotating the entire genome. Having a fully annotated genome at the structural and functional level will enable the research community to plan and execute evidence-based experiments to quickly learn more about this pathogen.

LITERATURE CITED

Basenko EY, Pulman JA, Shanmugasundram A, Harb OS, Crouch K, Starns D, Warrenfeltz S, Aurrecoechea C, Stoeckert CJ, Kissinger JC, et al. 2018. FungiDB: An integrated bioinformatic resource for fungi and oomycetes. Journal of Fungi [Internet]. [cited 2020 Feb 11]; 4(1):39. Available from: https://www.mdpi.com/2309-608X/4/1/39 doi: 10.3390/jof4010039

Caza M, Kronstad JW. 2019. The cAMP/Protein Kinase A pathway regulates virulence and adaptation to host conditions in Cryptococcus neoformans. Frontiers in Cellular and Infection Microbiology [Internet]. [cited 2020 Feb 7]; 2019(9): 212 Available from: https://doaj-org.ezproxy.mtsu.edu/article/c8bca4d3d8084b018cd19812f282bc3c doi:10.3389/fcimb.2019.00212

Haas BJ, Salzberg SL, Zhu W, Pertea M, Allen JE, Orvis J, White O, Buell CR, Wortman JR. 2008. Automated eukaryotic gene structure annotation using EVidence Modeler and the Program to Assemble Spliced Alignments. Genome Biology [Internet]. [cited 2020 Feb 7]; 9(1). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2395244/ doi:10.1186/gb-2008-9-1-r7

Howe KL, Contreras-Moreira B, De Silva N, Maslen G, Akanni W, Allen J, Alvarez-Jarreta J, Barba M, Bolser DM, Cambell L, et al. 2019. Ensembl genomes 2020-enabling non-vertebrate genomic research. Nucleic Acids Research [Internet]. [cited 2020 Feb 7]; 2 48(D1): D689–D695. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6943047/ doi: 10.1093/nar/gkz890

Humann JL, Lee T, Ficklin S, Main D. 2019. Structural and functional annotation of eukaryotic genomes with GenSAS. Methods in Molecular Biology Gene Prediction [Internet]. [cited 2020 Feb 7];1962:29–51. Available from: https://link.springer.com/protocol/10.1007/978-1-4939-9173-0_3#citeas. doi:10.1007/978-1-4939-9173-0_3

Jung K-W, Lee Y, Huh EY, Lee SC, Lim S, Bahn Y-S. 2019. Rad53- and Chk1-Dependent DNA damage response pathways cooperatively promote fungal pathogenesis and modulate antifungal drug susceptibility. MBio [Internet]. [cited 2020 Feb 7];10(1). Available from: https://mbio-asm-org.ezproxy.mtsu.edu/content/10/1/e01726-18. doi:10.1128/mbio.01726-18

Kronstad JW, Hu G, Choi J. 2011. The cAMP/Protein Kinase A Pathway and Virulence in *Cryptococcus neoformans*. Mycology [Internet]. [cited 2020 Feb 7]; 39(3), 143-150. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3385117/ doi:10.5941/myco.2011.39.3.143

Letunic I, Doerks T, Bork P. 2015. SMART: recent updates, new developments and status in 2015. Nucleic Acids Research [Internet]. [cited 2020 Feb 7]; 43(D1): D257-260. Available from: https://academic.oup.com/nar/article/43/D1/D257/2439521 doi:10.1093/nar/gku949

Loyse A, Thangaraj H, Easterbrook P, Ford N, Roy M, Chiller T, Govender N, Harrison TS, Bicanic T. 2013. Cryptococcal meningitis: improving access to essential antifungal medicines in resource-poor countries. The Lancet Infectious Diseases [Internet]. [cited 2020 Feb 7];13(7):629–637. Available from: http://eds.b.ebscohost.com.ezproxy.mtsu.edu/eds/detail/detail?vid=6&sid=12dd6afd-7e71-43b6-89d1-5296f141f1b1@pdc-v-sessmgr01&bdata=JnNpdGU9ZWRzLWxpdmUmc2NvcGU9c2l0ZQ==#AN=000321030500023&db=edswsc. doi:10.1016/s1473-3099(13)70078-1

Nosanchuk JD, Casadevall A. 2006. Impact of melanin on microbial virulence and clinical resistance to antimicrobial compounds. Antimicrobial Agents and Chemotherapy [Internet]. [cited 2020 Feb 7];50(11):3519–3528. Available from: https://aac-asm-org.ezproxy.mtsu.edu/content/50/11/3519. doi:10.1128/aac.00545-06

Pettit R, Repp K, Hazen K. 2010. Temperature affects the susceptibility of Cryptococcus neoformans biofilms to antifungal agents. Medical Mycology [Internet]. [cited 2020 Feb 7];48(2):421–427. Available from: https://www.ncbi.nlm.nih.gov/pubmed/?term=10.1080%2F13693780903136879 doi:10.1080/13693780903136879

Pukkila-Worley R, Gerrald QD, Kraus PR, Boily M-J, Davis MJ, Giles SS, Cox GM, Heitman J, Alspaugh JA. 2005. Transcriptional network of multiple capsule and melanin genes governed by the Cryptococcus neoformans Cyclic AMP cascade. Eukaryotic Cell [Internet]. [cited 2020 Feb 7];4(1):190–201. Available from: https://ec-asm-org.ezproxy.mtsu.edu/content/4/1/190. doi:10.1128/ec.4.1.190-201.2005

Rajasingham R, Smith RM, Park BJ, Jarvis JN, Govender NP, Chiller TM, Denning DW, Loyse A, Boulware DR. 2017. Global burden of disease of hiv-associated cryptococcal meningitis: An updated analysis. The Lancet Infectious Diseases [Internet]. [cited 2021 Mar 29];17(8): 873-881. Available from: https://pubmed.ncbi.nlm.nih.gov/28483415/. doi:10.1016/s1473-3099(17)30243-8

Sabiiti W, May RC. 2012. Capsule independent uptake of the fungal pathogen Cryptococcus neoformans into brain microvascular endothelial cells. PLoS ONE [Internet]. [cited 2020 Feb 7];7(4):1297–1313. Available from: https://search-proquest-com.ezproxy.mtsu.edu/docview/1112717366?pq-origsite=360link. doi:10.1371/journal.pone.0035455

Srikanta D, Santiago-Tirado FH, Doering TL. 2014. Cryptococcus neoformans: historical curiosity to modern pathogen. Yeast [Internet]. [cited 2020 Feb 7];31(2):47–60. Available from: http://eds.a.ebscohost.com.ezproxy.mtsu.edu/eds/detail/detail?vid=2&sid=6134474f-a288-4983-b0b7-09d53e3ec9f0@sessionmgr4008&bdata=JnNpdGU9ZWRzLWxpdmUmc2NvcGU9c2l0ZQ==#AN=edsgcl.357683785&db=edsgao. doi:10.1002/yea.2997

Steenbergen JN, Casadevall A. 2003. The origin and maintenance of virulence for the human pathogenic fungus Cryptococcus neoformans. Microbes and Infection [Internet]. [cited 2020 Feb 7];5(7):667–675. Available from: https://www.ncbi.nlm.nih.gov/pubmed/12787743 doi:10.1016/s1286-4579(03)00092-3

Tucker JS, Guess TE, McClelland EE. The Role of Testosterone and Gibberellic Acid in the Melanization of *Cryptococcus neoformans*. Front Microbiol [Internet]. [cited 2021 Mar 29];2020 Aug 13;11:1921. Available from: https://pubmed.ncbi.nlm.nih.gov/32922377/ doi: 10.3389/fmicb.2020.01921.

Vechi HT, Theodoro RC, Oliveira ALD, Gomes RMODS, Soares RDA, Freire MG, Bay MB. 2019. Invasive fungal infection by Cryptococcus neoformans var. grubii with bone marrow and meningeal involvement in a HIV-infected patient: a case report. BMC Infectious Diseases [Internet]. [cited 2020 Feb 7];19(1):220. Available from: https://www.ncbi.nlm.nih.gov/pubmed/?term=10.1186%2Fs12879-019-3831-8 doi:10.1186/s12879-019-3831-8

Walton FJ, Idnurm A, Heitman J. 2005. Novel gene functions required for melanization of the human pathogenCryptococcus neoformans. Molecular Microbiology [Internet]. [cited 2020 Feb 7];57(5):1381–1396. Available from: https://www.ncbi.nlm.nih.gov/pubmed/?term=10.1111%2Fj.1365-2958.2005.04779.x doi:10.1111/j.1365-2958.2005.04779.x

Williamson PR. 1997. Laccase and melanin in the pathogenesis of Cryptococcus neoformans. Frontiers in Bioscience [Internet]. [cited 2020 Feb 7];2(5):99-107. Available from: https://www.ncbi.nlm.nih.gov/pubmed/?term=10.2741%2Fa231 doi:10.2741/a231

Yang CL, Wang J, Zou LL. 2017. Innate immune evasion strategies against Cryptococcal meningitis caused by Cryptococcus neoformans (Review). Experimental and Therapeutic Medicine [Internet]. [cited 2020 Feb 7];14(6): 5243-5250. Available from: https://www.ncbi.nlm.nih.gov/pubmed/?term=10.3892%2Fetm.2017.5220 doi:10.3892/etm.2017.5220