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

Chase Burton, Rebecca Seipelt-Thiemann
Biology Department and Honors College, Middle Tennessee State University

Introduction

- *Cryptococcus neoformans* is a fungal pathogen that is virulent in humans (1, 14) and can be acquired in the environment through inhalation of soil contaminated with bird feces (12).
- There are three variants of *C. neoformans*: *grubii*, *gatti*, and *neoformans*, with *grubii* being the most pathogenic to humans (14).
- *C. neoformans* can cause Cryptococcal meningitis, resulting in death if left untreated (5, 14, 15). Immunocompromised individuals are predominately affected by *C. neoformans* (1, 5, 14, 15).
- Capsule formation is one of the phenotypic factors that contributes to *C. neoformans*' virulence (11, 14). The large capsule is composed of polysaccharides and protects *C. neoformans* from its environment (17).
- At least nine genes are responsible for capsule formation in *C. neoformans* (Table 1) (3, 4, 5, 6, 8, 10, 11, 13, 16, 18).
- Despite growing knowledge about the biology of this pathogenic fungus, few genomic-level resources, other than a sequenced genome and a computationally predicted gene set, are currently available.

Purpose

The objective of this experiment is:

- To reannotate and improve the accuracy of the virulence-causing genes pertaining to *C. neoformans*' capsule in the genome as viewed by the Apollo workspace in the GenSAS gene annotation system using evidence and improved computational gene models.

GenSAS Background

- The Genome Sequence Annotation Server (GenSAS) is a computational genome annotation platform that supports structural and functional genomic annotation along with manual genomic curation (9).
- GenSAS utilizes Apollo, which generates computationally-predicted genomic sequences based on data uploaded to the program (7), allowing for researchers to view and edit the genomic sequences in the GenSAS workspace.

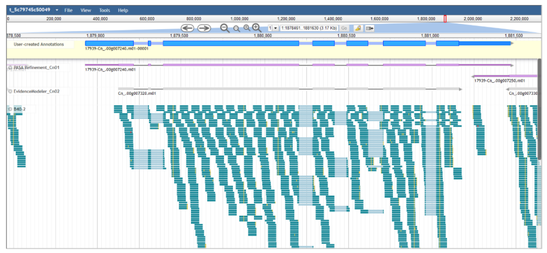

Figure 1. Overview of GenSAS with view of Apollo platform. Shown is Apollo through the GenSAS platform. At the top in blue, the gene model is in the Apollo workspace, which is in yellow, allowing for genomic editing. The pink model is generated using the Program of Assemble Spliced Alignments (PASA) algorithm that predicts gene structure. The dark and light green lines beneath the gene model represents data uploaded to GenSAS, with dark green representing exon data and light green representing intron data.

Methods

- The GenSAS annotation platform has been populated with the genome sequence, original sequence, and transcriptome (RNA sequencing data) from three experiments: seven clinical isolates and one reference strain (E. McClelland, personal communication), an estrogen and reference strain (J. Tucker, submitted), and an experiment with scytovirin and hormone-treated strain (R. Mcfeeters, personal communication).
- The new PASA algorithm has been used to improve their genome coordinates (Table 1) found on gene information databases, such as FungiDB.
- The transcriptome data, consisting of the three experiments, was used for reannotation and to compare with the *C. neoformans* var. *grubii* H99 genomic reference data present in GenSAS.
- Using these data, a super-transcript model was created for each of the genes using newly predicted computational models provided by PASA and analysis of specific sequencing transcriptome data to identify accurate exon borders, intron borders, alternative splicing, and approximate transcription starts and stops (Figure 2).

Table 1. Capsule Virulence Genes and Genomic Coordinates. Listed in the table are the nine genes that were the focus of this study. The coordinates listed are the chromosome and nucleotide region where each gene is located.

Genes	CNAG Identification	Coordinates	Source
CAP64	CNAG_02885	Chromosome 3, 530,546..532,814(-)	FungiDB
CAP59	CNAG_00721	Chromosome 1, 1,878,927..1,881,469(+)	FungiDB
CAP10	CNAG_07554	Chromosome 3, 1,190,190..1,192,419(-)	FungiDB
CAP60	CNAG_00600	Chromosome 1, 1,544,106..1,546,340(-)	FungiDB
CMT1	CNAG_03158	Chromosome 8, 1:211,670..214,236(+)	FungiDB
CAS1	CNAG_07937	Chromosome 13, 452,065..455,631(-)	FungiDB
CAS3	CNAG_03644	Chromosome 2, 1:423,088..425,925(+)	FungiDB
CAC1	CNAG_03202	Chromosome 8, 329,991..338,255(-)	FungiDB
PKA1	CNAG_00396	Chromosome 1, 1,039,957..1,042,941(-)	FungiDB

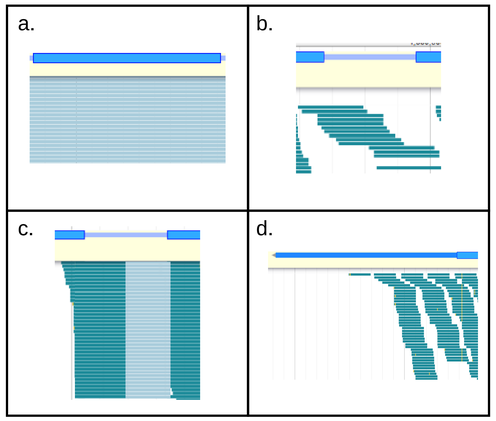

Figure 2. Examples of Areas Requiring Reannotation. (a) Exon skipping. The blue line represents an exon while light green area underneath represents the intron evidence suggesting that the exon is not translated. (b) Intron retention. The light blue portion of the blue line represents the intron and the dark green exon evidence present underneath blue line suggest the retention (c) Exon Size (Shortening). The intron represented by the light blue portion of the blue line has dark green exon evidence present underneath the blue line that suggests the left half of the intron is retained and translated as an exon. (d) Untranslated Region. This region is present at the beginning and end of the gene (blue line). In some instances, dark green exon evidence is present to support the shortening of the untranslated region.

Results

- Reannotation of gene models was conducted using through analysis of transcriptome data. Some genes required multiple transcripts to be created in response to the provided transcriptome data (Figure 3).
- Based on the transcripts that were created, super-transcripts were made to represent a combination of each of the transcript models (Figure 4). These models represent all possible combinations of exons that would be translated based on the transcriptome data.

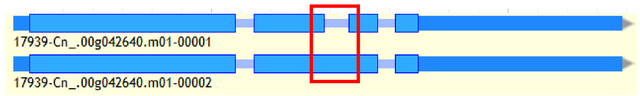

Figure 3. Transcript Reannotations of the *CMT1* Gene. The first gene model is the original model proposed by the PASA algorithm. The second gene model is the transcript model created through reannotation based on transcriptome data. The reannotation of the transcript shows the retention of intron 2, indicated by the red box.

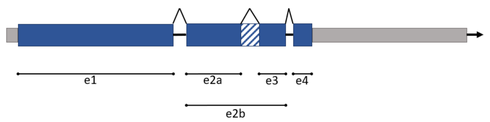

Figure 4. Super-Transcript Model of the *CMT1* Gene. The blue rectangles represent the exons that are translated. The blue hashed line area in exon e2b is an area that is translated or untranslated. The gray area is the untranslated region of the gene.

Conclusions

- Based on the transcriptome data, the nine genes required reannotations to their transcript models, which were collapsed to form super-transcript models of possible translation pathways for each of those genes in order to improve the accuracy of the gene models present within GenSAS.

Future Directions

Further research is needed:

- This genomic evidence can then reconstruct RNA and protein *in silico* using ExPASy, a bioinformatics tool.
- SMART and pfam protein databases can be used to search for protein domains in new isoforms allowing for functional predictions.

Literature Cited

1. Alspaugh JA. 2015. Virulence mechanisms and Cryptococcus neoformans pathogenesis. Fungal Genetics and Biology. 78:55–58. doi:10.1016/j.fgb.2014.09.004.
2. Alspaugh JA, Pukkila-Worley R, Harashima T, Cavallo LM, Funnell D, Cox GM, Perfect JR, Kronstad JW, Heitman J. 2002. Adenylyl cyclase functions downstream of the Gα protein Gpa1 and controls mating and pathogenicity of Cryptococcus neoformans. Eukaryotic Cell. 1(1):75–84. doi:10.1128/EC.1.1.75-84.2002.
3. Bose I, Reese AJ, Ory JJ, Janbon G, Doering TL. 2003. A yeast under cover: The capsule of Cryptococcus neoformans. Eukaryotic Cell. 2(4):655–663. doi:10.1128/EC.2.4.655-663.2003.
4. Buchanan KL, Murphy JW. 1998. What makes Cryptococcus neoformans a pathogen? Emerging Infectious Diseases. 4(1):71–83. doi:10.3201/eid0401.980109.
5. Chang YC, Chung KJK. 1994. Complementation of a Capsule-Deficient Mutation of Cryptococcus neoformans Restores Its Virulence. [accessed 2020 Jan 31]. http://mcb.asm.org/.
6. Chang YC, Penoyer LA, Kwon-Chung KJ. 1996. The Second Capsule Gene of Cryptococcus neoformans, CAP64, Is Essential for Virulence. [accessed 2020 Jan 31]. http://iai.asm.org/.
7. Dunn NA, Unni DR, Diesh C, Munoz-Torres M, Harris NL, Yao E, Rasche H, Holmes IH, Elsik CG, Lewis SE. 2019. Apollo: Democratizing genome annotation. PLoS Computational Biology. 15(2). doi:10.1371/journal.pcbi.1006790.
8. 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 Immunology & Medical Microbiology. 57(2):142–150. doi:10.1111/j.1574-695X.2009.00587.x. [accessed 2020 Jan 31]. https://academic.oup.com/femspd/article-lookup/doi/10.1111/j.1574-695X.2009.00587.x.
9. 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.
10. 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. Molecular Microbiology. 42(2):453–467. doi:10.1046/j.1365-2958.2001.02651.x. [accessed 2020 Jan 31]. http://doi.wiley.com/10.1046/j.1365-2958.2001.02651.x.
11. 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 Jan 31]. www.actabp.pl.
12. Mada PK, Alam MU. 2019. Cryptococcus (Cryptococcosis). StatPearls Publishing. [accessed 2020 Mar 6]. http://www.ncbi.nlm.nih.gov/pubmed/28613714.
13. Moyrand F, Chang YC, Himmelreich U, Kwon-Chung KJ, Janbon G. 2004. Cas3p belongs to a seven-member family of capsule structure designer proteins. Eukaryotic Cell. 3(6):1513–1524. doi:10.1128/EC.3.6.1513-1524.2004.
14. Nielsen K, Cox GM, Litvintseva AP, Mylonakis E, Malliaris SD, Benjamin DK, Giles SS, Mitchell TG, Casadevall A, Perfect JR, et al. 2005. Cryptococcus neoformans α strains preferentially disseminate to the central nervous system during coinfection. Infection and Immunity. 73(8):4922–4933. doi:10.1128/IAI.73.8.4922-4933.2005.
15. Odom A. 1997. Calcineurin is required for virulence of Cryptococcus neoformans. The EMBO Journal. 16(10):2576–2589. doi:10.1093/emboj/16.10.2576. [accessed 2020 Jan 31]. http://emboj.embopress.org/cgi/doi/10.1093/emboj/16.10.2576.
16. O'Meara TR, Alspaugh JA. 2012. The Cryptococcus neoformans capsule: A sword and a shield. Clinical Microbiology Reviews. 25(3):387–408. doi:10.1128/CMR.00001-12.
17. del Poeta M. 2004. Role of phagocytosis in the virulence of Cryptococcus neoformans. Eukaryotic Cell. 3(5):1067–1075. doi:10.1128/EC.3.5.1067-1075.2004.
18. Zaragoza O, Rodrigues ML, de Jesus M, Frases S, Dadachova E, Casadevall A. 2009. Chapter 4 The Capsule of the Fungal Pathogen Cryptococcus neoformans. Advances in Applied Microbiology. 68:133–216. doi:10.1016/S0065-2164(09)01204-0.