Structural Genome Annotation using Transcriptome Data for Two Eukaryotic Species

Madonna Ghobrial[1], Marcela Tello-Ruiz[2], Cristina Fernandez-Marco[2], Doreen Ware[2], and Rebecca Seipelt-Thiemann[1]

[1]Biology Department and Honors College, Middle Tennessee State University; [2] Cold Spring Harbor Laboratory

Introduction and Aims

- Decreasing prices and technology improvements have allowed scientists to sequence hundreds of genomes across all kingdoms.
- Knowing the genome sequence is only the first step in gaining a better understanding of how the genome and genes work.
- The next steps in genomic science will be to annotate the structural components, and then the functional aspects of the genome.
- *Zea mays* is a species of corn which fuels vehicles and is found in various products of everyday use.
- *Cryptococcus neoformans* is a fungal pathogen that causes cryptococcal meningitis in immunocompromised individuals.
- This project aimed to use different bioinformatics platforms to contribute to the structural genome annotation for two eukaryotic species: *Zea mays* and *Cryptococcus neoformans.*

Materials & Methods

Zea mays project

Corn transcriptome data was visualized in the Apollo genome viewer/annotation platform with current gene models for genes related to plant immunity in corn

⇩

Gene models and data for 12 genes were evaluated for areas of incongruence such as untranslated regions (UTR), intron retention, differing exon borders, or alternative splice sites

⇩

Areas of incongruence were recorded, and a new gene model was then generated in the annotation space

Cryptococcus neoformans project

Transcriptome data, genome sequence, and the current genome annotation were uploaded to a genome annotation platform, GenSAS

⇩

Informed computational gene models using a new algorithm (PASA) were generated

⇩

Gene models for six genes in the *C. neoformans* HOG pathway were also evaluated for areas of incongruence

⇩

Areas of incongruence were recorded, and revised gene models were generated

Results

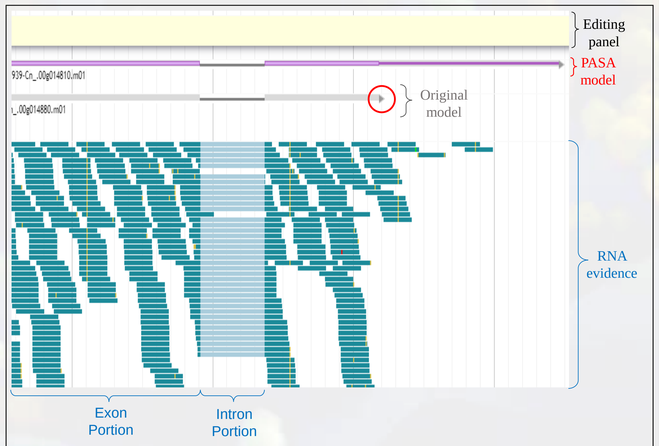

Figure 1. Apollo Software. The Apollo genome annotation and visualization portal was used to examine RNA sequencing evidence and current gene models. Direction of transcription follows the direction of the arrow (red circle). All areas of gene transcript are labeled.

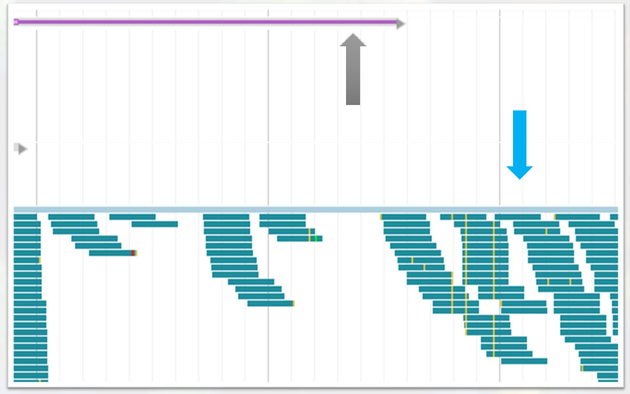

Figure 2. Examples of RNA Evidence for Gene Model Revision of UTR: The supporting RNA evidence (blue arrow) is being transcribed downstream of the PASA model (grey arrow), suggesting a larger 3' UTR than PASA predicted.

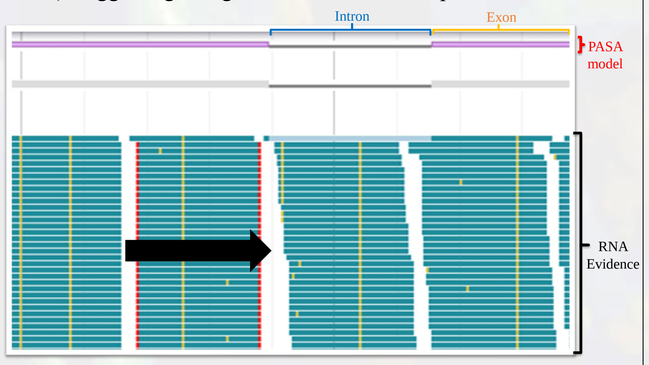

Figure 3. RNA Evidence for Intron Retention. There are RNA reads (black arrow) present for a location the PASA model suggests is an intron (blue bracket). This suggests that this is a retained intron.

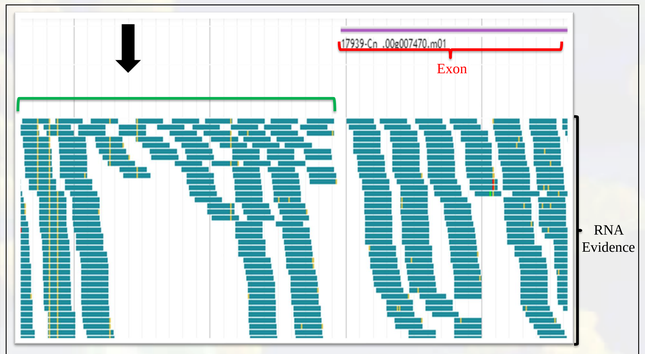

Figure 4. Example of RNA Evidence for Gene Model Revision of Differing Exon Borders. There is supporting RNA evidence (yellow arrow) that suggested the exon depicted by PASA (red bracket) is larger, suggesting an extension of the exon (green bracket).

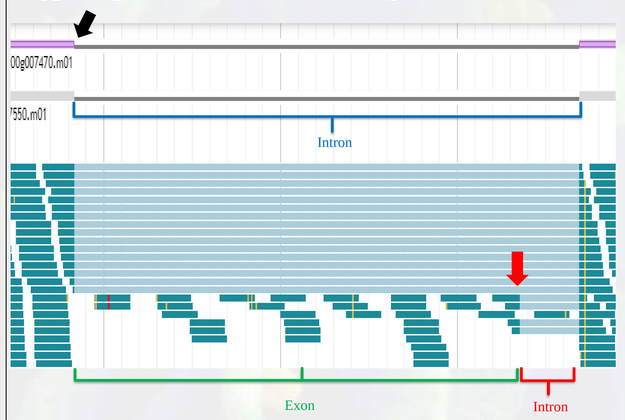

Figure 5. Example of Gene Model Revision for an Alternative 5' Splice Site. The PASA model shows this entire shaded area as intron (blue bracket). However, the RNA evidence shows that there may be an alternative 5' splice site (red arrow) rather than the splice site PASA shows (black arrow) which means the exon on the left is extended (green bracket) and the intron is shortened (red bracket).

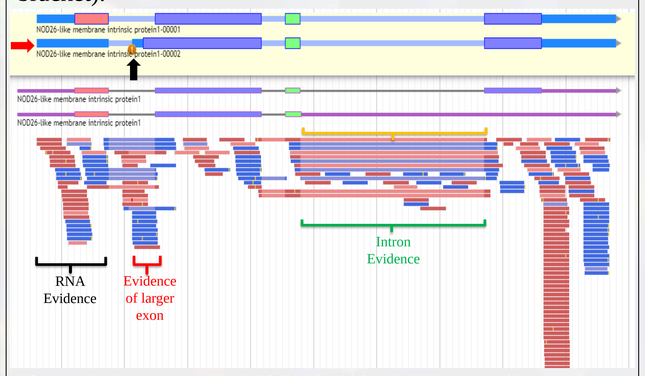

Figure 6. *Zea mays* Zm00001d016237. RNA reads (black bracket) suggested a shortened UTR (red arrow). RNA evidence (red bracket) shows the existence of a 3' alternative splice site (black arrow), resulting in a larger exon. The PASA model depicts one large exon at the 3' end (yellow bracket), however RNA evidence suggested the addition of an intron (green bracket).

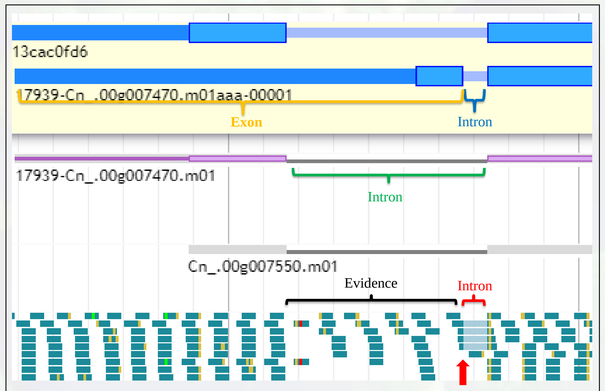

Figure 7. *C. neoformans AQP1* gene RNA evidence of Alternative 5' Splice Site. The RNA sequencing evidence showed reads (black bracket) and a smaller region of intron (red bracket) for an area the PASA model found to be one large intron (green bracket). This suggested there was an alternative 5' splice site (red arrow), generating a larger exon (yellow bracket) and a shorter intron (blue bracket).

Conclusions

•Out of the 12 *Zea mays* immunity related genes evaluated for the first portion of the project, only one gene (Zm00001d016237**)** showed adequate RNA evidence for manual annotation and curation. This gene contained 3 areas of incongruence: UTR, 3' alternative splice site, and the addition of an intron.

•A total of six *Cryptococcus neoformans* HOG pathway genes were annotated: *AQP1*, *ENA1*, *NHA1*, *HRK1*, *SRX1*, and *ATF1*.

•A total of thirteen areas of incongruence were found and edited within the *C. neoformans* project: one 5' UTR, five 3' UTR, two retained introns, one differing exon border, one alternative 5' splice site, one addition of an exon, one addition of an intron, and one gene model being split.

•The new revised gene models and transcripts for the *Zea mays* immunity-related gene as well as the six *C. neoformans* virulence-related HOG pathway genes provide not only more accurate information to plan and execute experiments but may also aid in better understanding each gene's regulation and expression, as well as the encoded protein's function.

•In turn, more accurate experimental results may lead to a clearer understanding of each protein's role in the *Zea mays* genome as well as the virulence of *Cryptococcus neoformans*.

•Future steps would include a BLAST search to confirm and identify structure of each gene and the encoded protein's function.

Literature Cited

- Bahn, Y.S., Jung, K.W. 2013. Stress Signaling Pathways for the Pathogenicity of Cryptococcus. Eukaryotic Cell. 12(12):1564–1577.
- Howe, KL., *et al*. 2019. Ensembl Genomes 2020-enabling non-vertebrate genomic research. Nucleic Acids Research.
- Humann, J.L., *et al*. 2019. Structural and Functional Annotation of Eukaryotic Genomes with GenSAS. Gene Prediction: Methods and Protocols. Ed. Martin Kollmar. New York: Humana Press. 29-51.