Identifying gene duplications under positive selection in *Wyeomyia smithii*

Julian Aileru[1], Jade Takakuwa[1], Peter Armbruster[3], Bill Bradshaw[4], Christina Holzapfel[4], Liz Cooper[1,2]

University of North Carolina at Charlotte[1], North Carolina Research Center[2], Georgetown University[3], University of Oregon[4]

Introduction

- *Wyeomyia smithii* is a mosquito that has intrigued researchers due to its uncommon behavior of polymorphic biting behavior and unique life history.
- Recent genome assembly and annotation efforts have revealed a significant number of duplication events within the 904.6MB *Wyeomyia* Genome.
- Our study focuses on examining the Dn/Ds ratio for gene duplicates within *Wyeomyia*, shedding light on the selective pressures that have driven genetic divergence.

Data Preparation

- Orthologous Groups for *Wyeomyia* Genes were identified using Orthofinder.

- Fasta files for orthologous sequences containing two or more *Wyeomyia* genes were identified and filtered for genes with identical or similar lengths.

- Selected Orthogroup sequence files were then aligned using clustal-omega and converted to a Codon Alignment using PAL2NAL.

- The Codon Alignment files were then converted to AXT format and then KaKs_Calculator 3.0 was run on each Codon Alignment File.

Test Statistic

Using KaKsCalculator3.0, with the Model Averaging Method, pairwise Dn/Ds mutation rates were calculated for gene duplicates. Additionally, the topGo package in R was used for GO enrichment analysis.

$$Dn/Ds = \frac{\text{rate of non-synonymous substitutions}}{\text{rate of synonymous substitutions}}$$

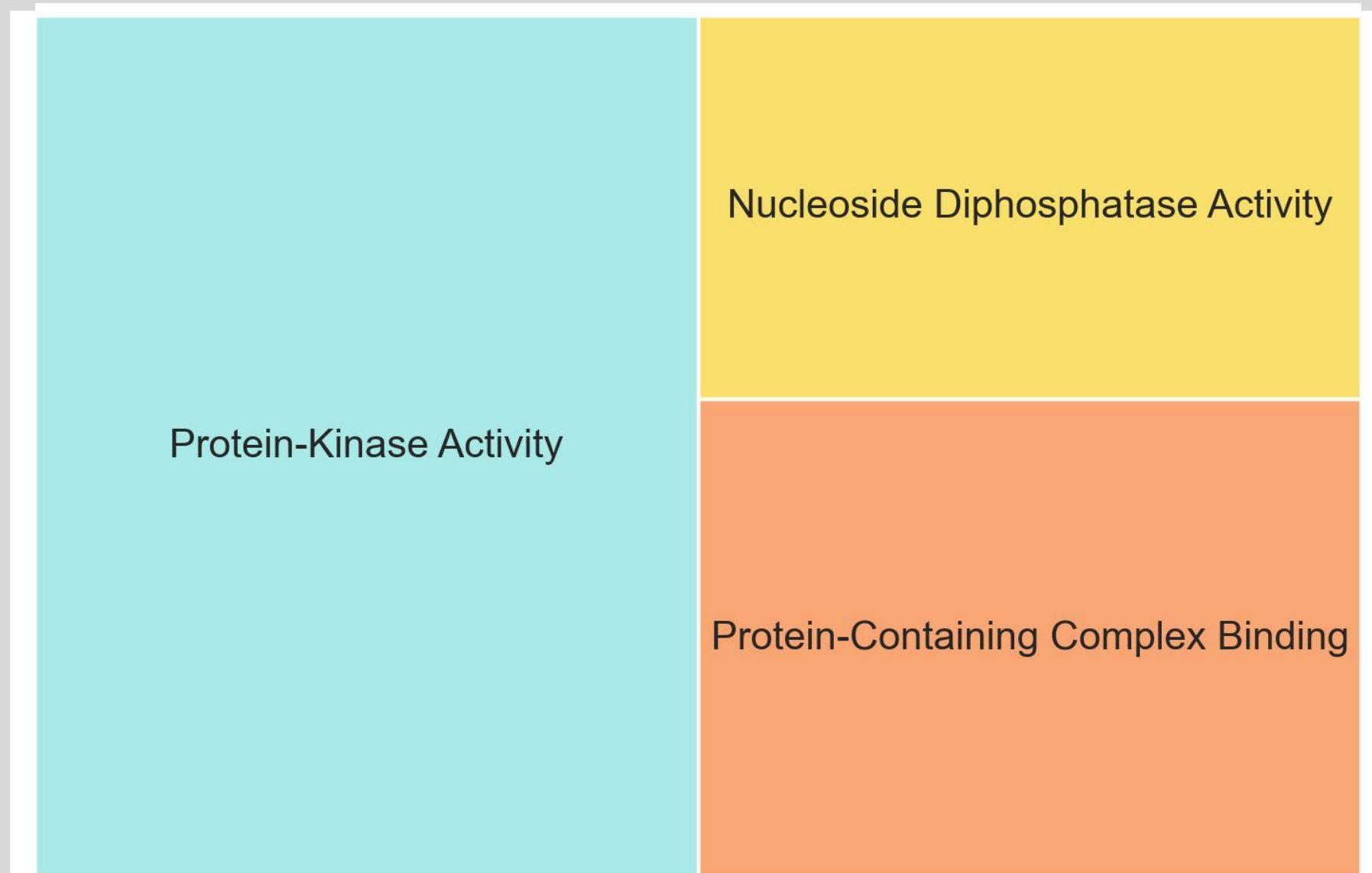Positive Selection:

$$Dn/Ds > 1 \quad \text{Positive Selection}$$

Negative Selection:

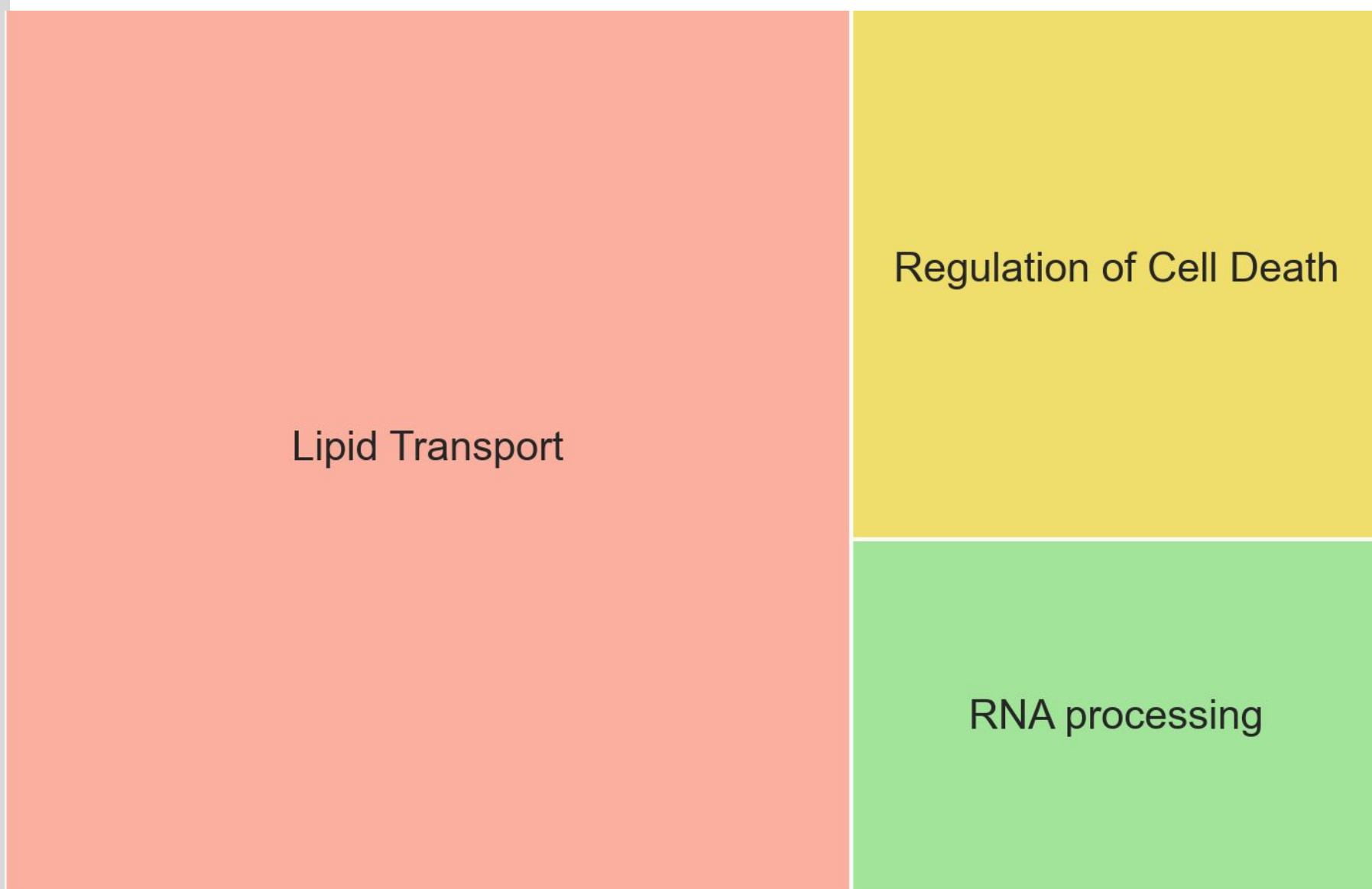$$Dn/Ds < 1 \quad \text{Negative Selection}$$

Neutral Selection:

$$Dn/Ds = 1 \quad \text{Neutral Selection}$$

Expansion & Selection

Positive Selection	Expansion - Yes	Expansion - No
Yes	343	1563
No	497	3047

Table 1: Contingency Table for Positive Selection and Expansion

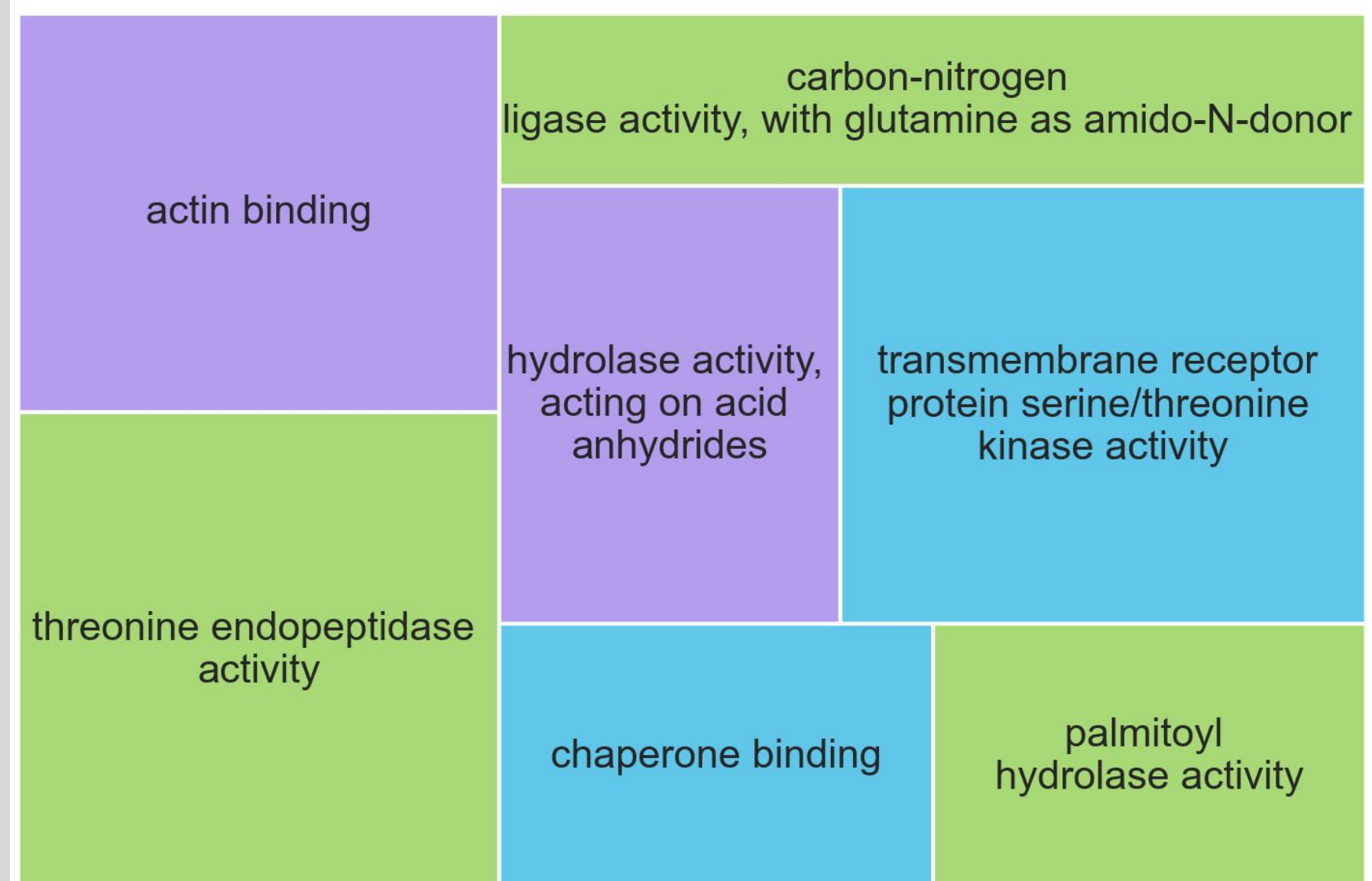| Test Statistic	Degrees of Freedom	p-value
χ^2	1	0.0001263

Enriched Gene Categories
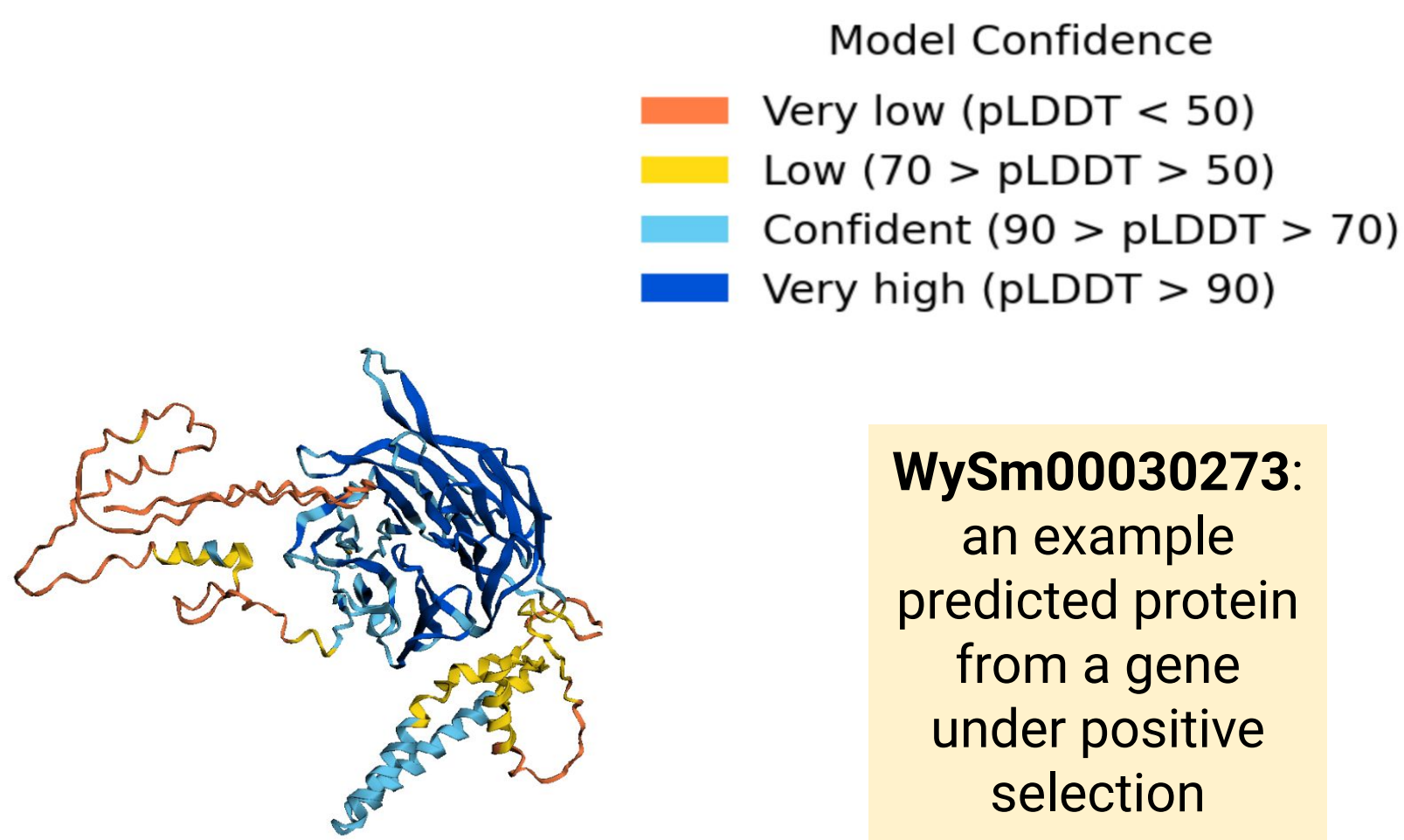

Enriched Categories in Positive Selection (MF)

Enriched Categories in Positive Selection (BP)

Enriched Categories in Positive Selection and Expansion

Predicted Gene Structure

Model Confidence
- Very low (pLDDT < 50)
- Low (70 > pLDDT > 50)
- Confident (90 > pLDDT > 70)
- Very high (pLDDT > 90)

WySm00030273: an example predicted protein from a gene under positive selection

Conclusions

- Gene Ontology Analysis revealed multiple different functional categories under positive selection in the gene duplicates of *Wyeomyia*
- The significant association between positive selection and expansion suggests that many of the expanded gene families in *Wyeomyia* may have also undergone selection for neofunctionalization or subfunctionalization, which could be a crucial mechanism for their unique adaptation.
- Results from a previous study comparing the gene expression of avid biters and obligate nonbiters revealed significant expression of actin and RNA processing genes in nonbiters

References

Zdobnov, E. M., & Apweiler, R. (2001). InterProScan–an integration platform for the signature-recognition methods in InterPro. Bioinformatics, 17(9), 847-848.
Emms, D. M., & Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biology, 20(1), 238.
Suyama, M., Torrents, D., & Bork, P. (2006). PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic Acids Research, 34(suppl_2), W609-W612.
Wang, D. P., Wan, H. L., Zhang, S., & Yu, J. (2010). Gamma-MYN: a new algorithm for estimating Ka and Ks with consideration of variable substitution rates. Biology Direct, 5(1), 1-7.
AJumper, J., Evans, R., Pritzel, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. Nature, 596(7873), 583-589.
De Bie, T., Cristianini, N., Demuth, J. P., & Hahn, M. W. (2006). CAFE: a computational tool for the study of gene family evolution. Bioinformatics, 22(10), 1269-1271.
Alexa, A., & Rahnenführer, J. (2010). topGO: Enrichment Analysis for Gene Ontology. R package version 2.38.1.
Bradshaw, William E., et al. "Evolutionary transition from blood feeding to obligate nonbiting in a Mosquito." *Proceedings of the National Academy of Sciences*, vol. 115, no. 5, 2017, pp. 1009–1014, https://doi.org/10.1073/pnas.1717502115.