PURA Final Report: Lauren Sabo

Part 1: Project Outcome

The aim of my project was to perform structural and functional genome annotation across five African cichlid species using BRAKER3, incorporating species-specific RNA-Seq data and high-quality genome assemblies generated from PacBio HiFi reads. The goal of this project was to research gene content and structure across species to better understand evolutionary divergence and potential functional adaptations. The main objective was to identify lineage-specific genes and conserved elements that might inform speciation mechanisms in Lake Malawi cichlids.

Over the course of the semester, I successfully extracted high-molecular-weight DNA from blood samples using the Bionano SP-G2 Blood DNA isolation protocol, improving on earlier attempts with fin clips that yielded inconsistent results. I sequenced these samples on the PacBio HiFi platform, generating long-read data with sufficient depth and quality to enable reliable downstream assembly and annotation. I used genome assemblies in conjunction with RNA-Seq alignments to run BRAKER3 and begin curating gene models for each species. Through this, I then spent the majority of the semester curating a BLASTp pipeline, using the BRAKER3 data, to completely, functionally annotate the genomes presented.

This work contributes to ongoing efforts to build high-quality genomic resources for African cichlids, a species known for their rapid and extensive adaptive radiation. My analyses provide a foundation for identifying candidate genes associated with traits under selection and may help inform conservation strategies as well.

Part 2: Professional Development Outcomes

This semester has been one of the most formative of my undergraduate research career. I developed familiarity with advanced bioinformatics tools such as BRAKER3, as well as critical skills in genome assembly evaluation, annotation curation, and data troubleshooting. One of my proudest accomplishments was adapting to and resolving the challenge of the insufficient BLASTx pipeline by incorporating the BLASTp pipeline, which dramatically improved the frequency of usable, complete data.

Working closely with my mentor, Nikesh Kumar, I met the expectations we set in our mentor/mentee agreement: to maintain open communication, seek guidance when needed, and take initiative in solving problems. I maintained weekly progress updates and actively engaged in developing our experimental strategy. If I could repeat my project, I would have researched the BLASTp pipeline sooner, avoiding much confusion from earlier in the semester.

Looking ahead, my goals include deepening my understanding of comparative genomics and gene evolution, publishing my research, and continuing this work in graduate school. I aim to develop greater fluency in scripting for data parsing and visualization to streamline my workflows. Most importantly, I aim to retain the spirit of curiosity and persistence that has driven my progress thus far.