Abstract: Identifying Pipelines for *D. Melanogaster* Microbiome Sequence Data

Austen Miller

June 25, 2024

The diet of animals can significantly impact the diversity of their microbiomes, which in turn impacts their health. Studying the composition of these communities in model organisms, such as *Drosophila Melanogaster*, can provide insights into how they might function within humans. By analyzing the variable regions of the 16s rRNA gene, bacterial community composition can be observed down to the level of species. Members of the lab of Dr. Imilce Rodriguez have sequenced the 16s rRNA gene of the bacterial communities of *D. melanogaster* flies varying in age after exposure to 50% ethanol 0, 1, or 2 times. Oxford Nanopore Technology allows full length sequencing of regions like the 16s gene but also generates quite large amounts of data. Our objective is to identify the best bioinformatic pipeline to analyze Nanopore full-length 16s rRNA sequence data to obtain information not only about bacteria abundance and composition but also about alpha and beta diversity. We are working with the Boqueron HPCF to provide processing power for the large data sets. We will be trying the NanoRTax pipeline, as well as developing our own using other tools for bacterial taxonomy classification and R for sample diversity analysis. The preliminary results on the project have included challenges with the Boqueron system due to incompatibilities between the requirements of the system and those of the NanoRTax program. This problem is exacerbated by the absence of a full-time IT team for real-time support. Whether we succeed in resolving the challenges of working with NanoRTax, or use another pipeline, it will provide a useful tool for data analysis of ONT sequence data in these kinds of experiments.