Identifying Bioinformatic Pipelines for Analyzing Nanopore16s Gut Microbiome Sequence Data

Austen Miller, Kerianalee Cintron, Jan Bilbao Del Valle, Carlos Quiñones, Patricia Pujols, Imilce A. Rodriguez-Fernandez

College of Natural Sciences, Department of Biology, University of Puerto Rico Rio Piedras

In animals, multiple factors such as diet, age, and alcohol can significantly impact the diversity of their gut 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. The bacterial 16s rRNA gene can be sequenced to characterize the changes in gut bacteria composition and abundance. Our lab sequenced the 16s rRNA gene of the gut bacterial communities of young and old *D. melanogaster* flies 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.