Identifying Pipelines for *D. Melanogaster* Microbiome Sequence Data

Austen Miller

June 14, 2024

Diverse communities of microorganisms, including bacteria, live within the guts of all animals. This symbiotic relationship is crucial for the health of the host by assisting in digestion and even affecting the brain (Slavin 2013; Cryan and Dinan 2012). The host’s diet can significantly impact these microscopic communities, including altering the microbial species diversity and composition. Studying the composition of these communities in model organisms can provide insights into how they might function within humans.

Full-length 16s rRNA sequencing uses the 16s ribosomal subunit sequence present in all bacteria to analyze the taxonomic composition of bacterial communities, such as the microbiomes of animals. This is possible because ribosomal sequences mutate relatively slowly due to their importance in biological functioning. However there are 9 known variable regions in the 16s rRNA gene in which that sequence is able to distinguish bacteria at the genus and species level. The similarity of the 16s gene within different bacteria with only a few changes makes it useful as a tool to analyze community composition when sequenced.

In this particular case, we will examine the effects of alcohol exposure on the microbiome of the fruit fly species *Drosophila melanogaster*. *D. melanogaster* is useful as a model organism due to its relative ease of work, short generation times, and many genes analogous to humans. The lab of Dr. Imilce Rodriguez is interested in studying how alcohol can affect the gut microbiome using *Drosophila* as a model. Using Nanopore sequencing technology, members in her lab sequenced the 16s rRNA of the bacteria communities present in the guts of young and old flies exposed to 50% ethanol vapor 0, 1 or 2 times.

Oxford Nanopore technology allows scientists to sequence long reads and thus the full-length 16s rRNA gene (~1500 bp). This is in contrast to other technologies, such as Next-Generation Sequencing (NGS) Illumina, where much shorter sequences (~500 bp) are generated. These longer reads allow for increased specificity in taxonomic classification of bacteria since all nine variable regions are sequenced. However, full-length 16S rRNA sequencing generates large amounts of data, which can be complicated to process and require substantial computational resources for analysis. Standardization across different laboratories and studies is essential but challenging to achieve. Nanopore sequencing is advantageous for various reasons, including being cheaper than Illumina, having long reads, which allows us for species-level resolution of bacteria communities, and, despite inaccuracy compared to Illumina, the company is developing new chemistry that has allowed to reduce error rate close to 1% (Zhang et al., n.d.).

Unfortunately, the existing pipelines for full-length 16s sequence data analysis are not always consistently updated. The rapid advance of technology can make the old analysis tools obsolete and difficult to replicate. The goal of my summer project is to identify the best bioinformatic pipeline to analyze Nanopore full-length 16s rRNAseq in our lab to obtain information not only about bacteria abundance and composition but also about alpha and beta diversity. I will use the 16s rRNAseq data previously generated by the lab (explained above).

One of the pipelines we will be trying is the NanoRTax pipeline. NanoRTax allows “...a nextflow-based pipeline for bacterial taxonomy classification and sample diversity analysis of nanopore full-length 16S rRNA amplicon reads” (Rodríguez-Pérez, Ciuffreda, and Flores 2022). We will also be working on developing our own pipelines using known tools for bacterial taxonomy classification and R for sample diversity analysis. The computational skills gained in the IQBIO program’s carpentries will be quite useful in working with these programs and understanding their results.

Possible challenges that we will be presented with are trouble with the computational tools being worked with, the details of *D. melanogaster* biology, and missing technical terms while working in Spanish and English. To overcome these obstacles, I will communicate with the PI and lab mates l and look for resources using the internet.

References

Cryan, John F., and Timothy G. Dinan. 2012. “*Mind-Altering Microorganisms: The Impact of the Gut Microbiota on Brain and Behaviour*.” Nature Reviews. Neuroscience 13 (10): 701–12. <https://doi.org/10.1038/nrn3346>.

Slavin, Joanne. 2013. “*Fiber and Prebiotics: Mechanisms and Health Benefits*.” Nutrients 5 (4): 1417–35. <https://doi.org/10.3390/nu5041417>.

Delahaye, Clara, and Jacques Nicolas. 2021. “*Sequencing DNA with Nanopores: Troubles and Biases.*” *PLoS ONE* 16 (10): e0257521.<https://doi.org/10.1371/journal.pone.0257521>.

Zhang, Tianyuan, Hanzhou Li, Silin Ma, Jian Cao, Hao Liao, Qiaoyun Huang, and Wenli Chen. n.d. “The Newest Oxford Nanopore R10.4.1 Full-Length 16S rRNA Sequencing Enables the Accurate Resolution of Species-Level Microbial Community Profiling.” *Applied and Environmental Microbiology* 89 (10): e00605-23. <https://doi.org/10.1128/aem.00605-23>.

Rodríguez-Pérez, Héctor, Laura Ciuffreda, and Carlos Flores. 2022. “NanoRTax, a Real-Time Pipeline for Taxonomic and Diversity Analysis of Nanopore 16S rRNA Amplicon Sequencing Data.” *Computational and Structural Biotechnology Journal* 20:5350–54.<https://doi.org/10.1016/j.csbj.2022.09.024>.