Jonathan Taylor

Genome Assembly Final Report

To start off this report, I am going to explain the process to get all the information needed and the process to get the data and assembly. I had several problems with getting my entire assembly together and was unable to completely a full assembly. I believe it was due to using the windows OS instead of the mac OS software. I intend to explain what to do to get a full assembly to show that I know what I am doing and understand it, since I do not have an assembly complete.

The assembler used in this used in this process was abyss. The start of the assembling process is to run the fastq files in the assembly prep script. After getting these files, it is important to look at the fasta html files. Looking at these files will help tell where there needs to be trimming done and the quality of the assembly files. After ensuring the quality is good, it is time to do the trimming. The trimming is done by running the trimming file on spruce. The final step is to actually assembly the assembly. By tweaking the assembly assembler file with different k-mer value, it will help show which specific part of the assembly is going to be used. To figure which k-mer lengths to use, it is good to make a graph comparing N50 to k-mer values, Max BP values to k-mer values, and total BP of sequence to k-mer values. This should give a completed assembly after being ran.

The final step in the assembly process is to upload the final assembly to rast.nmpdr.org to be analyzed. The data should give information about the subsystem content in the genome. This helps show how the genome is going to be broken up within the organism being studied. It will help tell if the identification of the organism chosen is actually correct by comparing it with other organisms with closely related genetics. This concludes the assembly process.

Being that this was my first time ever doing any type of DNA sequencing and genome sequencing through bioinformatics, I faced several challenges. I did not realize that it was even possible to analyze data this way before I took this class. I do feel like I gained a fair amount of knowledge about bioinformatics throughout the class. Overall, this was one of the more interesting classes I’ve ever taken and hope to be able to continue learning more about bioinformatics in the future.