Spring 2018

Improving the Genome Assemblies of Multiple Dropsohila and Bacterio[hage Species

Margaret Gaggioli
Washington University in St. Louis

Improving the Genome Assemblies of Multiple *Drosophila* and Bacteriophage Species

Margaret Gaggioli

Mentor: Sarah C. R. Elgin

Sequencing technologies have advanced rapidly in recent years, substantially reducing the cost of whole genome sequencing. However, these technologies are not perfect—manual inspections and targeted resequencing are required to produce high quality assemblies. During Summer 2017, we improved the genome assemblies of 25 bacteriophage and two Drosophila species (*Drosophila ficusphila*, *Drosophila eugracilis*). The Drosophila species are being used in motif finding through phylogenetic footprinting. For the phage project, we characterized the physical ends and resolved low quality regions within the assemblies so that the high-quality sequences can be used in gene annotations and submitted for publication. The first step for both projects involved the use of Consed to identify consensus errors, low quality regions, and gaps within the draft assemblies. We then attempted to resolve these problematic regions using bioinformatics and wet-bench techniques. Initial assessment revealed 34 gaps in the Muller F and D elements of the two Drosophila species. Many of these gaps are located within repetitive regions and might have resulted from misassembly. Using Consed, we resolved one gap by rearranging the available Illumina and 454 sequencing reads. The remaining gaps are caused by missing data and require additional sequencing to resolve. We used Consed to design primers on both sides of each gap and the region was then amplified using PCR. If gel electrophoresis analysis determined that the desired PCR product had been generated, then the samples were sent for sequencing. Between both species, we resolved two gaps using this approach but four attempted gaps remain unresolved. There were no gaps in the phage projects and we were able to resolve consensus errors and low-quality regions using bioinformatics and wet-lab techniques. Thus the phage genomes are finished, but more remains to be done in the Drosophila projects.