

Lab 3 – Part 1

Genome Assembly and Analysis Report Guidelines

This report should have 2 sections. It will be worth 100 points. It should be no more than 5 pages. It is due at 11:59:00 on 4/28/17.

Section 1: *E. coli* assembly

You should describe the data and what you did. Be sure to include the following information.

- What portion of the graph is likely the genome? What are the two large sets of contigs at the bottom of the graph (around node 18 and around node 4)?
- What do you think the identity of this DNA is and how you determined that.
- What you think the most common path of this DNA is.
 - How did you determine this?
 - What do you think the less common paths represent?
- Were there regions you couldn't resolve? Why? What data would allow you to resolve these? If this were a previously unsequenced organism (a newly discovered fish for example) how would your confirmation techniques change?
- How would you go about trying to get the actual genome of this bacterial strain without any additional data? If you could get additional data, what would it be? Why?
- What do you think the small, unconnected contigs at the bottom are? Why do you think that (what data are you basing that on)?
- Include the fasta file for the region node 4→65

Section 2: *S. aureus* assembly

You should describe how we analyzed this genome. In particular you should include.

- A description (in your own words) of the N50 and N75. Why are they useful for comparing between different assemblies/assemblers on the same data or genome? Why are they not a good measure if you are comparing between different genomes?
- What is the N50 and N75 for this data?
- Does this assembly seem like it is assembling a genome that is the correct size for *S. aureus*?
- How is this assembly different (or not) than the *E. coli* assembly?
- Are there any extra-genomic DNA structures in this assembly (phage/plasmid/not junk)? How did you determine this?
- Starting at node 207-, at what nodes would any possible path you can trace end?

As always, this should be written in paragraph form, not as bullet points. Also, be sure to cite any sources and tools you used (yes, that include Bandage and BLAST).