BioE 131/231
Fall 2019
                                                                              Final Project

Overview

Congratulations, all! You have made it to the final project. From now until the end of the semester, you will analyze a set of genomes using the skills you have learned thus far and explore new types of analysis. These genomes come from the Salmonella reference collection (https://people.ucalgary.ca/~kesander/Kit_14A.html), which were isolated from different sources many years ago, but have not been fully analyzed. We are interested in learning more about these strains and their pathogenicity.

Week 1: Break into groups, obtain data, assemble genome, annotate genome.
Week 2: Propose your own analysis.
Week 3: No lab (Thanksgiving). Work on your own analysis.
Week 4: Group presentations during lab. Final report due.

The goal of this part of class is to let you work on a real, unpublished dataset, applying your own ideas of how best to conduct the analysis. Ideas for projects, deliverables, and due dates are listed below on a week-by-week basis.

Week 1

First off, find a group of four people (max five per group). Please let your GSI know who is in your new group so they can keep track. Your GSI will assign you a set of reads for the final project. Each group will be given a different set of reads.

Next, please assemble your genome into contigs using SPAdes as we did in lab 8/9 using the `-1` and `-2` flags for paired end reads. Remember to run your assembly using `-t 1` and `-m 16` to limit your CPU and memory usage (save some for everyone else). If those limits are too low, let your GSI know. This will take a while, so be sure to run it in `tmux`.

When you're done, calculate assembly statistics and plot a histogram of contig lengths from your genome assembly in iPython, along with the N50. Summarize your results from SeqMatch (do we have the strain we expect?) and RAST/BASys annotations. (If annotation isn't ready in time, submit this in Week 2). Your report is due the following Wednesday at 11:59 AM.

Week 2

By now, you should have an assembly and annotation of your genome. While you are waiting for your annotations, you can start brainstorming ideas for your own analysis. Before you leave today, make sure your GSI approves of your project. You will have two weeks (including Thanksgiving) to work on it before your final presentation.

A list of project ideas will be uploaded to bCourses shortly if it hasn't been already.

Upload a report summarizing the findings of your annotations and a description of the project you intend to work on by the following Wednesday at 11:59 AM.

Presentations

Prepare a ten minute PowerPoint presentation describing all of the results of your genome assembly and analysis. Everyone in your group should speak during the presentation. You will have 5 minutes for questions at the end.

Summarize the results of your assembly (e.g., N50, contig length histogram).
Summarize the results of your annotations.

Which analysis project did you choose?
What were some of the issues you ran into?
What were your results?
If you had more time, what additional experiments and analyses would you perform?

Written report

Submit a report (up to 5 pages, not included figures) along with your final presentation. This report should summarize your assembly results, your annotations, the methods that you used for your original analysis, and the findings of your analysis. Please be clear about the question you are trying to answer, how your chosen method will help you answer it, and any potential limitations of your results.