BioE 131/231
Fall 2018

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 microbial metagenome using the skills you have learned thus far. These metagenomes have been obtained from radioactive soil at different depths beneath the surface, and contain diverse soil microbes.

Week 1: Break into groups, obtain data, assemble contigs, quantify abundance of microbes.
Week 2: MG-RAST, propose your own analysis.
Week 3: No lab (Thanksgiving). Work on your own analysis.
Week 4: Group presentations during lab.

The goal of this part of class is to let you work on a real, unpublished dataset, applying your own ideas of how best to conduct the analysis.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 another pair of lab partners and form a new group of four people (max five per group, min three). Please let your GSI know who is in your new group so they can keep track. Your GSI will assign you a set of metagenomic reads for the final project. Each group will be given a different set of reads.

Next, please assemble your metagenome into contigs using the `metaspades` part of SPAdes. This is pretty similar to genome assembly, but you'll have to use the `--meta` flag. There is no need to uncompress the input read files--just pass them to SPAdes as is, using the `-1` and `-2` flags. Remember to run your assembly using `-p 2` 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 a `screen`.

While this is running, try to get one of these web services to quantify the abundance of different microbes in your metagenome:

MetaPhlAn2: http://huttenhower.sph.harvard.edu/galaxy/
Click "Get Data" to upload your FASTQ files, then click "MetaPhlAn2" and feed them through the pipeline.

One Codex: https://onecodex.com/
Register for an account, then use the `onecodex` command (from a terminal on bioe131.com) to upload your data. Login to the website to process it.

It's possible to run MetaPhlAn2 locally on bioe131.com if you can't get it to work over the web.

When you're done, plot a histogram of contig lengths from your metagenome assembly in iPython, along with the N50. Summarize the results of either MetaPhlAn or One Codex, giving us an idea of what kind of microbes were living in your soil sample. Your report is due the following Wednesday at 11:59 PM.

Week 2

By now, you should have an assembly of your metagenomic contigs. The next thing you'll want to do is annotate them. For this, you can use MG-RAST, a metagenome version of the same tool you used to annotate your genomes. It will take a while, so use this time to brainstorm 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 of no more than two pages summarizing the findings of your MG-RAST run and a description of the project you intend to work on by the following Wednesday at 11:59 PM.

Presentations

Prepare a ten minute PowerPoint presentation describing all of the results of your metagenome 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).
What were the most abundant microbes in your sample?
How do those abundances differ from those of your classmates?
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?