

Biol/CS 383 - Bioinformatics - Spring 2019

Module 2 - 16S Metagenomics - Capstone

Due: The PDF should be uploaded to Moodle before class on Weds., March 27

This assignment is your opportunity to collect, organize, analyze, and present the metagenomic work you have done with the local soil dataset. While your capstone will include many of the pieces you have assembled to-date, it is an independent piece of work requiring additional analysis, visualizations, and the like. See below for some hints/ideas about where you can take your analyses.

You will design and develop your capstone alone or with a partner, drawing on the materials and knowledge you developed working on the metagenomics module.

Your capstone assignment is a short scientific-format paper with the following sections:

Introduction - what is the broader context of what you are presenting? What are your hypotheses/predictions? (suggest 3 paragraphs: 1st is big picture, 2nd is targeted toward the context of your question, 3rd is your hypotheses/predictions).

Materials/Methods - state clearly where the data came from and how you obtained it. This also includes how you analyzed it (both the workflow and the analysis). What are the scientific protocols/processes that are employed? Basic details of the relevant parts of the experimental design should be included and are available from the “meet the data” - and if you are using the 2012/13 data, you can cite the relevant publication. Importantly, cite the software and workflows you used!

Results - statistical analyses, professional quality figures/tables, diagrams and the like, along with a verbal description of your results. Note, your figures/tables need appropriate captions.

Discussion - an interpretation of your results; be sure to put things into the broader context (the same that you used in the introduction...were your hypotheses/predictions supported?). I recommend 3 paragraphs - 1st is to state, broadly, whether your hypotheses were supported or not, the 2nd is to interpret your results and discuss the nuances of the data and analysis, and the 3rd is to put all of your results back into the broader context that you began the paper with (how do your results inform that broad topic?).

References - you should cite all methods and software used. You should also use scientific papers to help establish the broader context. You should use the MLA style for your references (see: <http://library.earlham.edu/c.php?g=82591&p=533379>). We expect 5-10 references (5 is a bare minimum).

Items to consider:

- The value of diagrams, e.g. one that shows the workflow from field -> lab -> sequencer -> Mothur -> R -> Science (that is readings -> data -> information). Maps are often useful for geocoded data.
- The Mothur MiSeq SOP has lots of ideas for how to further analyze data. Consider your questions and develop some of these additional analyses to help answer your questions. Also note the citation for the Mothur SOP (not to mention the citations for R and vegan).
- Be sure to identify a clear question/hypothesis/prediction! Organize your paper around your hypothesis. Think about the context of these samples, what can you learn about the context, what questions does it suggest?
- Look at papers, you will need to read some and cite some anyway. What types of analyses or graphs did they use that you liked?
- Use collaboration tools such as Google Drive documents, sheets, etc. to organize your work. If you are passing around an MS Word document via email and wondering who has the current copy you are unlikely to be successful at this or any other group writing project.

Schedule:

- You will have time during class on this coming Thursday and next Tuesday to work on this.