Final Project
Practical Introduction to Metagenomics
AS 410.734
Due date (<u>strict</u>): December 15th at noon

The Final Project is a practical one that makes up a significant portion (30%) of your grade and is meant to exercise your ability to devise and perform a metagenomics experiment using the tools and techniques learned throughout the semester.

I recognize that you all have different interests in the field and within the course material presented this semester. I want this to be an opportunity for you to further that interest in a practical project. To that end, you have quite a lot of freedom in choosing your own project for this. You will need to write a proposal for your project, which will be part of the grade of the final project itself.

Fear not, non-creative types. If the idea of crafting your own project and executing it in just under a month fills you with overwhelming anxiety, there are sample projects defined below as well.

<u>Project proposal (due Nov. 24th)</u>:

You must write a one-page, detailed proposal of your project and submit it before 11:59 PM on Saturday, November 24th. It should represent something of approximately the same depth and workload as the mid-term project I defined. The proposal should include:

1. Brief background on your study's topic.
2. Description of analysis you want to perform. What question does it attempt to answer?
3. Description of your input data. (scale, formats, etc.)
4. What process will you follow, and what tools are required for your analysis? This includes any installations you will need to perform (or requests for me to perform) as well as custom scripts you plan on writing. This doesn't have to be absolute, I know things will come up, but I want to see that you've thought the analysis out at least initially.
5. How will you present the results? Data tables, graphical chart types, web interface, etc.

We've covered a wide array of topics related to metagenomics throughout the semester. I can see from the Discussion posts that many of you tend to be more active and engaged on some topics compared with others. I've seen a lot of posts where you've suggested additional steps that could be performed on an analysis or (even more fun) questioned the authors' assertions in some studies. Now is a chance to explore these, or devise something completely different based on the topics we've covered.

It's pretty rare that I have rejected a project in my other courses – it's more often that I'll suggest additions or modifications to make it fit more with the course or meet project

requirements. I'll also be doing everything I can to respond to your submitted proposals within a day or two of their submission. For these reasons, you should consider it generally safe to get started on your project immediately upon submission if you want/need to get as much time in as possible before the due date.

Both the proposal and project submission itself will be done within Blackboard. Under "Lessons" you'll find areas for both.

The Write-Up:

Because I'm allowing such a wide range of topics and activities, proposed by you, for this final, I cannot provide a single description of the write-up requirements that will apply to everyone. Instead, I will provide general requirements below and then more specific ones individually with the e-mail you'll receive when I approve your project.

The target reader of your write-up should be someone who has only a cursory knowledge of what metagenomics is about. This means that your introduction should introduce metagenomics in general and discuss its importance. Then, you should provide background on your specific area of study for the project with references to previous work.

Again, I'll send details of the sections expected in your specific project with your acceptance e-mail, but there are some sections most should have, such as Methods, Results, Data/Code Availability, Discussion, and References.

The paper should be at least 3 pages in length, usually 5-6, and code should be attached as separate files (or left on the DIAG file system) rather than pasted within the document. PDF format is preferred, especially if you employ advanced formatting or embedded images.

Example Projects:

1.   Comparison of a newly-published tool/method with previous ones. This field changes rapidly and you will almost certainly be able to find new things published even during the course of the semester on the topics we've covered. You can try this latest and greatest thing and compare it with its predecessors. While this may seem to be one of the easiest ways to get through the final project, my expectations here will be pretty high. If you choose this route, your comparison should include a summary of the tools/approaches already in existence, executions of the new tool, and detailed comparisons of the results with the older ones. The write-up here should be of the detail and quality where you could spend another few weeks on the project and actually publish it for peer-review. (You don't have to, I'm just commenting on what I expect in the write-up.)

2.   Create a visual interface for data from a topic from the course. No, this wasn't something that was taught in this course, but I know many of you bring skills of your own to the course that you might be interested in applying. Every semester I've had at least one or two students obsessed with data visualization and I want to allow them to run with it. I know that it's difficult to create a great interface without really knowing the underlying data and methods, so this will come out in a good interface.

3.   Implement a RAST-like pipeline. Our speaker Dr. Meyer challenged you, saying that most of you should be able to implement the basics of the RAST pipeline in less than a week. "Try it!" he said, and he had his own students do it in a workshop. The basic premise here is to turn traditional annotation on its head and, instead of searching an entire set of gene predictions against large databases like NR, search a collection of annotated genes/profiles against your genome/microbiome of interest to annotate them. Then search any remaining gene predictions the traditional way. This one is for the brave.

4.   Old-school validation. One of the tenants of published research is the expectation of reproducible results. For bioinformatics papers this often means you should be able to download the author's dataset(s) and re-run the analysis described. You should try this with a dataset of manageable size and to validate a published analysis with an emphasis on how you might be able to improve upon with using updated tools/approaches.