

## **Clinical and Research Genomics Assignment #5**

### **Metagenomics Analysis and Swab Report**

---

**Assignment: Familiarize yourself with metagenomics and microbiome analysis using data from the PathoMap project and then analyze your own swab sample collected earlier in the course.**

**Due Date: 11:59PM on May 17<sup>th</sup>**

---

### **Downloading Data**

Go to link:

<http://physiology.med.cornell.edu/faculty/mason/lab/clinicalgenomics/problemsets.html>

To complete this assignment you will need the following files:

- PATHOMAP\_AB012 – Abandoned South Ferry Station
- PATHOMAP\_Culture01 – Cultured Sample (same station as PATHOMAP\_P00046)
- PATHOMAP\_P00046 – Subway Sample
- PATHOMAP\_GC012 – Gowanus Canal

### **Running OneCodex**

Register for an account with OneCodex here: <https://app.onecodex.com>

Then upload your file (.fasta or .fastq.gz). Once the file is uploaded and the algorithm has run you can view the results by clicking the 'Compare Analyses' tab.

Study the results and explore the various visualization options. You will use these results to answer the questions below.

### **Running BLAST (Basic Local Alignment Search Tool)**

To run BLAST follow the tutorial posted here:

<https://www.dropbox.com/s/assxif7k347d067/How%20to%20Use%20BLAST.pptx?dl=0>

You should use the '\_1000.fasta' version of the files since the regular files are too large and take too long to run on the BLAST web-based interface.

Explore the results. You will use these results to answer the questions below.

---

## Short Answer Questions

1. Summarize the results from the different tools. Do you see any similarities? Any differences? Pick a sample and list 5 species you found interesting. Give the species name as well as a brief description of the organism.
  2. We found interesting molecular echoes that make the abandoned station and Gowanus samples stand out against the rest of our dataset. Are there any organisms that you found only in the Gowanus? How about any organisms only in the abandoned station, but not in the other samples? Any explanation as to why/how they were introduced there?
  3. Compare and contrast the culture and sequencing methods. What are the pros/cons of each? Compare the results of Culture01 and P00046, they were taken from the same station, but one swab was cultured then sequenced, the other was sequenced. Do you find any organisms found in both samples?
- 

## Essay Questions

4. Put on the hat of metagenomics expert. You now have a massive metagenomic dataset of nearly 1500 samples collected from New York City's subway system that represent all species in a city. What sort of analysis would you like to perform and research question would you like to delve into?
- 

## Swab Analysis Report

Submit a report addressing the following questions:

- Where did you swab?
- Do the organisms predicted match what you would expect?
- What is the complexity of the sample?
- Is there evidence for pathogenic microbes? Evaluate the strength of that evidence.
- If your sample did not yield enough DNA or a library, what reasons may have contributed to that?
- Your genome and omes: Given what you have learned in the course, would you want to have your genome sequenced? Why or why not?

(If insufficient data was generated for your sample, answer the same questions, as appropriate, using the undetermined reads. Is there evidence for organisms that you would expect from the environment you swabbed within the undetermined reads?).

For any questions, please contact Chandrima Bhattacharya ([chb4004@med.cornell.edu](mailto:chb4004@med.cornell.edu)), Ebrahim Afshinnkoo ([eba2001@med.cornell.edu](mailto:eba2001@med.cornell.edu)), or Professor Mason ([chm2042@med.cornell.edu](mailto:chm2042@med.cornell.edu))