Clinical and Research Genomics Assignment #5
From Lecture_13-15 (April 19th): Microbiome and Metagenome Characterizations and Cross-Species Analysis

_____

Assignment: Analyze and contextualize raw sequence data from the PathoMap project to characterize the microbiome and metagenome of your samples.
Due Date: 5:00PM on May 1st

This assignment has two sets of questions.

_____

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? For each sample 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. The findings of plague and anthrax in the dataset created intense interest and curiosity, not surprisingly. In response we posted a blog that highlighted further research into these findings. Read these three posts:
"The Long Road from Data to Wisdom, and from DNA to Pathogen"
http://www.pathomap.org/2015/03/10/long-road-data-wisdom-dna-pathogen/.
Then:http://read-lab-confederation.github.io/nyc-subway-anthrax-study/.
Any finally:
http://www.microbe.net/2015/12/23/taxonomic-forensics-in-metagenomics-with-in-silico-marker-panels/
What do you think of the discussion? In your opinion, what challenges face the field of metagenomics?

5. 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?

Please hand the assignment on the day of the lecture, or email if you cannot attend.
For any questions, please contact Ebrahim Afshinnekoo (eba2001@med.cornell.edu), Alexa McIntyre (abm237@cornell.edu), Priyanka Vijay (prv2004@med.cornell.edu), or Professor Mason (chm2042@med.cornell.edu).