Midterm Exam

*Applications in Bioinformatics – Spring 2016*

This is a take Home Exam due on March 21th, 2017 by Midnight

**“On my honor, I have neither given nor received aid on this assignment, and I pledge that I am in compliance with the VCU Honor System.”**

**By placing your name here you agree to the VCU Honor Pledge:**

**Your Name:**  (No Name, No Grade)

# Introduction

We’ve spent the semester thus far learning about flu viruses and the technical details of Python by experimenting with creating tools that execute external programs and perform some fairly trivial tasks like parsing. Now it’s time to take the next step. You’ll still be responsible for designing a pipeline for a particular purpose, but that’s no longer the end goal; rather, now you’ll use the pipeline you create to assist you in performing further analysis of a flu virus dataset!

# The Problem

You already know that every year since 1997 the WHO has recommended annual reformulations of vaccine strains. You have received an email from the WHO that states due to budget cuts they won’t be able to perform assays on as many flu virus strains this year. Specifically, the WHO has asked you to use your bioinformatics skills and expertise to provide conclusive evidence to help them narrow down what H3N2 strains they should focus their efforts on. You identify that the best way to do this would be to compare all of the HA & NA proteins segments from H3N2 strains obtained this year and compare that first against the current vaccine strain to first determine if a vaccine change should occur, then determine if there is a better strain in the new sequences that would better cover the emerging flu strains.

**The Details**

You may make use of any of the programs we utilized so far in order to accomplish this. As a general overview your first step will involve comparing the current H3N2 vaccine strain to the new sequences, then compare each sequence in the new sequences to see if there is a sequence that could provide better coverage. It will be important to remember the difference between query coverage and percent identity and you’ll like need to make use of both Clustal and blast to accomplish this task. Utilizing all the programs available to you will provide evidence to strengthen your arguments for the answers to the questions below.

# The Questions

1. Looking at your results do you feel it is necessary to change the H3N2 Vaccine Strain? [Provide details from your results as to why or why not]. If yes, then provide the strain that you believe would help provide better immunity toward the emerging flu strains.
2. Based on the pipeline programming, justify the parameters used to parse the data. (Hint: Why does using a 97% threshold value give you valid results? Or does it?)
3. Write a flowchart or a step by step description of procedure used to solve the exam. [ This should include detailed steps so that we may duplicate your methods. E.G. Step 1, I ran blast and clustal to align and analyze the current vaccine strain against the emerging strains. command: python blast.py vaccine.fasta prot emerging.fasta blastresults.txt and python clustal.py ... Step 2, ...]

# Deliverables

1. Your commented script, as a yourname\_midterm.py file (70% of the grade)
2. Your answers to the questions above in .doc format (30% of the grade)

**As always, please submit them through the electronic curator in Blackboard.**