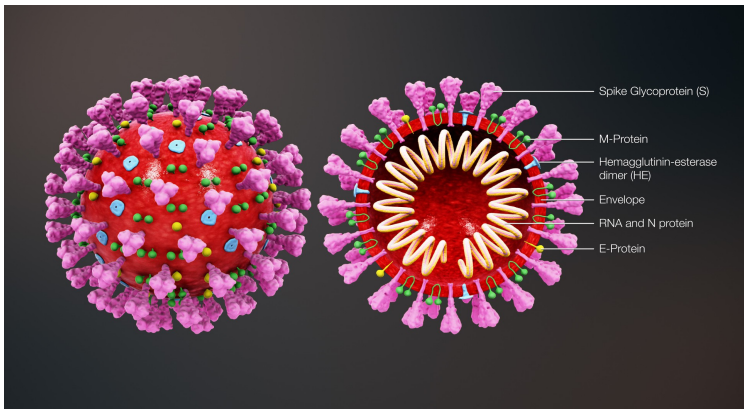
Hello everyone,

Welcome to our upcoming summer camp from June 22 – July 3 on Intermediate Coding for Medicine. You all learned Python in our introductory camps or elsewhere. Over the next ten days, you will see how your coding skills can be applied to understand the living world.

Here is our overall plan. I have a pool of problems from diverse areas of biology, and you need to apply your computational skills to answer them. Each one teaches you different concepts at the intersection of biology and genomics. **You can work at your own pace to solve as many problem as possible.** During the process, you may have to learn additional bioinformatics programs, and I will go over them when you work on those specific problems.

Here I list the primary problems first (with a proposed timeline) and then extra ones for those who finish these faster.

Day 1-2: Finding Information from the Coronavirus Genome



In the first two sessions, we will go over the basics of genome, nucleotides, proteins and transcription/translation. You will learn about the Biopython library developed to make genome analysis easy. You will practice your skills on the publicly available coronavirus genomes.

I have a separate but related problem for those, who completed this in the introductory module. Check “malaria parasite” in the “extra problems” section.

Day 3-4: Whale Evolution

Here you will learn computational tools - BLAST, MUSCLE and FastTree. We covered some of these in the introductory class, but here we do it in more detail.

The biological example we will pick for this exercise is whale evolution. I request all of you to watch the video https://www.youtube.com/watch?v=-OCMx2VuP1U. The evidence in the video is based on fossils, but can we study history using the DNA data?

We will learn that our bodies are also full of "fossils", and those fossils can be used to figure out where the whales came from.

Day 5, Week 2 Project: Asian Flush

From day 5 onward, we will explore the human genome, and see how single change in DNA can lead to the physical effects in the body. Asian flush syndrome describe some people's face turning red after drinking alcohol. This is caused by a single mutation in DNA prevalent in Chinese, Japanese and Korean people. We will find out why a single mutation causes such dramatic reaction.

Prior to the class, I have two short requests for each of you.

First, during the camp, we will be communicating using a message board - http://coding4medicine.com:2197. **If you do not have an account from an earlier class, please create one with your email address as username and let me know.** We need to approve the account, and then you will be able to access the forum.

Second, please watch the whale evolution documentary mentioned earlier, if you have not seen it already. It is not necessary for the class, but does provide context for what we plan to do on third/fourth days.

I look forward to meeting you all,

Dr. Manoj Pratim Samanta
June, 2020

Extra Problems

I am posting the summary of these problems here. The details will be shared, when you start working on any of them.

1. Malaria Parasite

Malaria is caused by a parasite, which is not a virus or bacteria but a unicellular eukaryote. Its genome is unusual, and you will find that out by visually inspecting it. Then we will write Python code to understand the implication of the differences in genome analysis.

2. Genome Assembly Jigsaw Puzzle

Genomes are assembled from small pieces of nucleotides, and the process is not different from solving a jigsaw puzzle. You will see an example of an equivalent problem and develop your own code to solve it.

3. How Different is *E. coli* from *E. coli*?

We will compare two E. coli genomes to answer this question. The new program we will use is Mummer. Also, you will learn about the computational (algorithmic) challenges in solving problems like this.

4. Color Blindness

Color blindness in humans is caused by single mutations (SNP) in genes for light-absorbing proteins. We will extract those genes from the human genome and find out the differences made by SNPs.

5. Genome Evolution and Gene Expansion

How are new genes created? We will look into the expansion of aldehyde dehydrogenase genes in the human genome to investigate.

6. Rnase P

In addition to protein-coding genes, our genomes contain RNA sequences. The rules for finding them are quite different than translating three-letter genetic code. You will solve one such problem here and explore the **regular expression** library in Python to do so.