Title: Using Phylogenetic trees and substitutions to Capture Relationship of SARS-Cov-2 Variants

Names: Liz Wyman Z1884762
           Kleo Bano Z1940978
           Chris Troyer Z1945059
           Roberto Rivas Z1906735

Emails: Z1884762@students.niu.edu
           Z1940978@students.niu.edu
           Z1945059@students.niu.edu
           Z1906735@students.niu.edu

Main Question: How do variants of SARS-COV-2 (Omicron, Alpha, etc) compare to each other?

Background & Significance:
The COVID19 pandemic created a newfound necessity for research in SARS-COV-2 viruses and variants. Since the start of the pandemic more and more variants have been discovered and sequenced. We will use different algorithms to create multiple phylogenetic trees. This will illustrate the different evolutionary relationships between variants. Using multiple algorithms will allow us to create a clear picture of the evolutionary relationship and compare algorithm outputs.

Data:
We will use SARS-Cov-2 genomes that can be downloaded from the NCBI website (data files are prepared by Dr. Hou). Each genome has around 30K DNA bases. We will first of all conduct all vs. all genome comparison and obtain the similarity measurement between every pair of genomes. The similarity is a numeric value between 0 and 1.

The study:
A comparison of phylogenetic trees will be used. We will calculate branch lengths for each algorithm and compare trees. We will map indels, substitutions and compare how those values change or modify the output on a specific phylogenetic tree. In the case of low substitution rate, we expect the branches to be close together. In the case of high substitution rate, we expect the branches to be far apart from one another.

Programming language & system: Java, Linux