Homework Assignment 3

Question 1: In class we played with downloading and aligning sequences from the internet. In our exploration we found that the SARS-CoV-2 Spike locus has significant similarity to a bat coronavirus RaTG13 Spike protein. For your homework I'd like you to adapt the tools and code in the Week 3 notebook to do the following things

1. Download the bat coronavirus RaTG13 Spike protein sequence. Align it to the set of translated Spike proteins that you can get from the alignments we made in class. You will need to use `mafft` for this, but no worries– it can align DNA and proteins

2. Using that alignment above create a visualization of the number of differences between pairs of sequences– perhaps a heatmap as in class or a histogram. Is the bat coronavirus sequence diverged from the others?

3. Create a multiple alignment of sequences for the ORF8 locus using the SARS-CoV-2 genome sequences. How do levels of differentiation at this locus compare to Spike?