Homework Assignment 3

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. Align the bat coronavirus RaTG13 Spike protein sequence to the aligned set of spike sequences we created in class as proteins. This is a challenging task, but all of the code to do this is shown in the week 3 notebook. You just have to put together the pieces. Here are the steps you could take to complete this problem:

    i) download the bat coronavirus RaTG13 sequence (last thing in Week 3 notebook)
    ii) read in the nucleotide alignments of covid genomes we made in class with (we called it "temp.aligned.fasta")
    iii) extract the bit of the alignment that is the spike ('s') protein
    iv) translate those sequences to proteins
    v) create a new alignment using `mafft` where you have now added the bat protein to the set of human-born SARs-CoV-2 sequences. `mafft` can align protein and DNA sequences, so no worries there.

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?