**Lab 4: Bioinformatics**

1) What is the above sequence? What gene is it from? Who published the paper that the sequence is from?

Protein BLAST:

A screenshot of a cell phone

Description automatically generated

The sequence is the spike glycoprotein of SARS-CoV-2. The paper was published by Cai, et al. in *Science* in 2020 (Title: Distinct conformational states of SARS-CoV-2 spike protein).

2) In one sentence, why is this particular gene of interest to medical and public health practice?

(From introduction of paper) The spike protein is involved in the beginning stages of infection, and so knowledge of spike protein structure and function is important to vaccine and treatment development.

3) Using your knowledge from class and internet resources, describe (in your own words, no copy-pasting!) an alignment technique.

Description of alignment technique.

4) Download the file `all\_sequences.fasta`. Align all of the sequences. Describe the method by which you aligned the sequences.

Describe alignment method

5) Estimate the evolutionary history of these sequences by building a phylogeny with the multiple-sequence alignment resulting from step 4. Describe the method by which you built a phylogeny.

Describe phylogeny method

Tree (used MAFFT + neighbor joining methods) in PDF file.