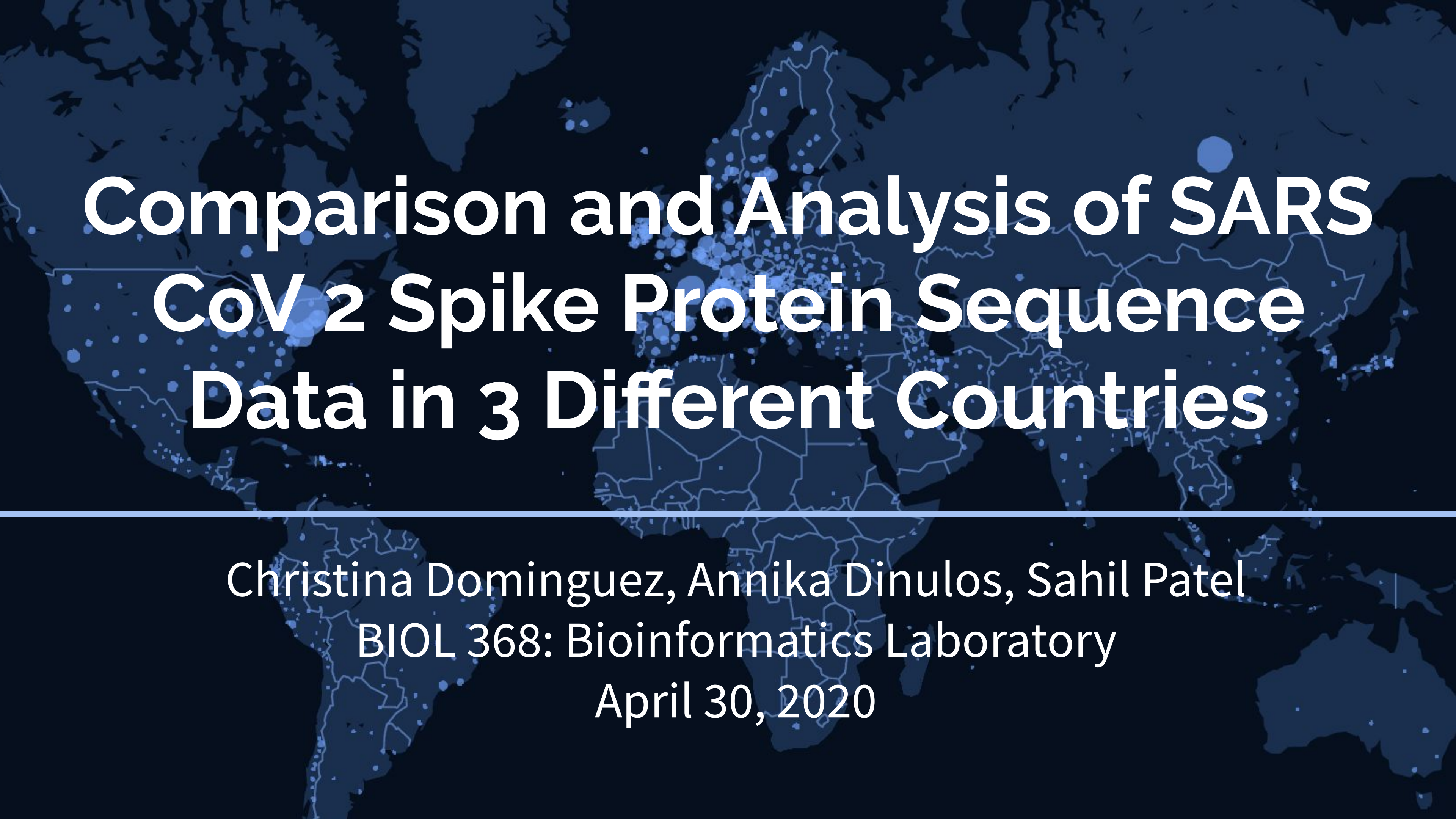


Comparison and Analysis of SARS CoV 2 Spike Protein Sequence Data in 3 Different Countries

Christina Dominguez, Annika Dinulos, Sahil Patel
BIOL 368: Bioinformatics Laboratory
April 30, 2020

Outline

- SARS-Cov-2 is rapidly spreading globally
- Multiple sequence alignment and phylogenetic trees were created
- Spike Protein sequences show little to no difference in each of the 3 countries
- Multiple Sequence Alignment of RBDs is consistent within the 12 Sequences
- Structure remained the same in each of the 12 sequences
- RBD remained unchanged in all 3 countries
- Selective pressure in SARS CoV 2 does not call for mutation
- Future directions could expand on sequence/structure relationship

Outline

- **SARS-CoV-2 is rapidly spreading globally**
- Multiple sequence alignment and phylogenetic trees were created
- Spike Protein sequences show little to no difference in each of the 3 countries
- Multiple Sequence Alignment of RBDs is consistent within the 12 Sequences
- Structure remained the same in each of the 12 sequences
- RBD remained unchanged in all 3 countries
- Selective pressure in SARS CoV 2 does not call for mutation
- Future directions could expand on sequence/structure relationship

SARS-CoV-2 is Rapidly Spreading Globally

- SARS-CoV-2 cause of COVID-19
  - USA: 1.06M confirmed infections (30%), 123K recovered, 61K deaths (37%)
  - Spain: 210K confirmed infections, 108K recovered, 24K deaths
  - South Korea:
- Coronavirus spike (S) glycoproteins mediates entry into host cells
  - Surface-exposed, therefore main target of antibodies
- cryo-EM structures of the SARS-CoV-2 S ectodomain trimer can provide a roadmap for designing vaccines and/or inhibitors of virus infection
- Evolution and mutation dynamics of SARS-CoV-2 need to be monitored closely
  - Varied virulence and immune characteristics have already emerged

Question

Has the SARS CoV 2 S glycoprotein undergone significant mutation as it continues to spread globally? What does it say about the virus, evolutionarily? and how might this affect vaccine development?

Outline

- SARS-Cov-2 is rapidly spreading globally
- **Multiple sequence alignment and phylogenetic trees were created**
- Spike Protein sequences show little to no difference in each of the 3 countries
- Multiple Sequence Alignment of RBDs is consistent within the 12 Sequences
- Structure remained the same in each of the 12 sequences
- RBD remained unchanged in all 3 countries
- Selective pressure in SARS CoV 2 does not call for mutation
- Future directions could expand on sequence/structure relationship

Multiple Sequence Alignment and Phylogenetic Trees Were Created

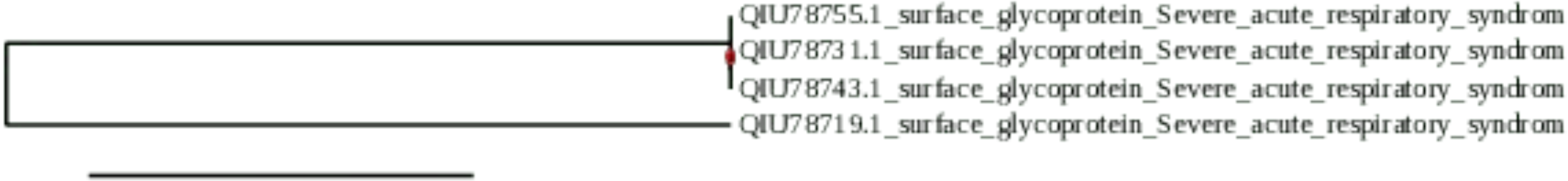
- Selected USA, Spain, and South Korea
  - Relatively Equidistant from one another
  - Similar Latitudes
  - Different Population Densities
- 4 Strains were selected at random from each of the countries
  - Derived from https://www.ncbi.nlm.nih.gov/genbank/sars-cov-2-seqs/
- Spike protein sequence data was aligned and used to make Phylogenetic Trees
  - Using http://www.phylogeny.fr/simple_phylogeny.cgi
- RBDs were compared with one another to search for inconsistencies
- Phylogenetic Trees were observed and interpreted
- Analysis was used to draw conclusions, answer question, and consider future explorations

Outline

- SARS-Cov-2 is rapidly spreading globally
- Multiple sequence alignment and phylogenetic trees were created
- **Spike Protein sequences show little to no difference in each of the 3 countries**
- Multiple Sequence Alignment of RBDs is consistent within the 12 Sequences
- Structure remained the same in each of the 12 sequences
- RBD remained unchanged in all 3 countries
- Selective pressure in SARS CoV 2 does not call for mutation
- Future directions could expand on sequence/structure relationship

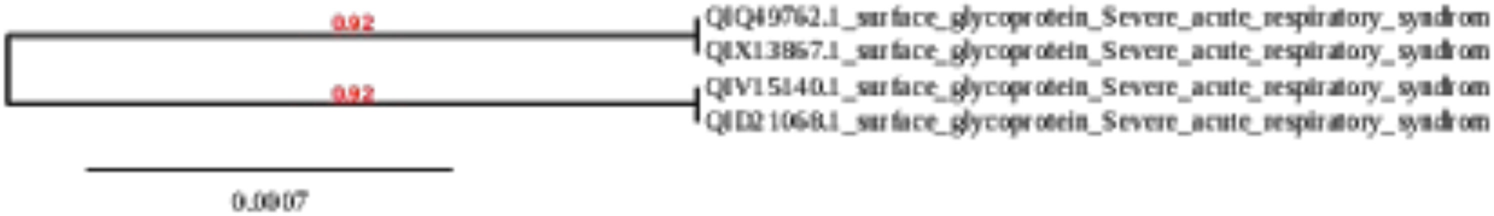
Spain Spike Protein Sequences Do not Show Difference in RBD Sequence

- **QIU78719.1 spike protein sequence differs in two amino acids:**
  - QIU78719 is glycine, other three sequences are aspartic acid
  - QIU78719 noted as “X”, others three sequences are lysine



USA Spike Protein Sequences Do not Show Difference in RBD Sequence

- **Spike protein sequences cluster together based on single amino acid difference:**
    - QIQ49762 and QIX13867 are glycine
    - QID21068 and QIV15140 are aspartic acid



Spike Protein Sequences Phylogenetic Tree Does not Differ in RBD Sequence

- QIU78719, QIQ49762, and QIX13867 group together due to glycine instead of aspartic acid at single position
- Does not occur in RBD sequence

Outline

- SARS-Cov-2 is rapidly spreading globally
- Multiple sequence alignment and phylogenetic trees were created
- Spike Protein sequences show little to no difference in each of the 3 countries
- **Multiple Sequence Alignment of RBDs is consistent within the 12 Sequences**
- Structure remained the same in each of the 12 sequences
- RBD remained unchanged in all 3 countries
- Selective pressure in SARS CoV 2 does not call for mutation
- Future directions could expand on sequence/structure relationship

Multiple Sequence Alignment of RBD is Consistent Within 12 Sequences

RBD Sequence Consistent with Wrap et al.

```
QIA98554.1      CTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISN
QIV14984.1      CTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISN
QIV14996.1      CTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISN
QIV15008.1      CTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISN
QIU78731.1      CTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISN
QIU78743.1      CTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISN
QIU78755.1      CTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISN
QID21068.1      CTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISN
QIV15140.1      CTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISN
QIU78719.1      CTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISN
QIQ49762.1      CTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISN
QIX13867.1      CTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISN
                ************************************************************
```

```
QIA98554.1      YNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPC
QIV14984.1      YNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPC
QIV14996.1      YNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPC
QIV15008.1      YNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPC
QIU78731.1      YNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPC
QIU78743.1      YNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPC
QIU78755.1      YNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPC
QID21068.1      YNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPC
QIV15140.1      YNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPC
QIU78719.1      YNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPC
QIQ49762.1      YNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPC
QIX13867.1      YNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPC
                ************************************************************
```

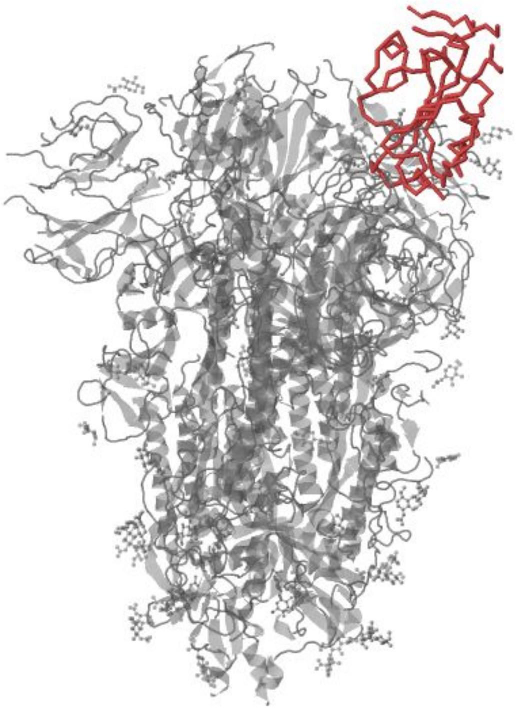
```
QIA98554.1      CVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQTGKIAD
QIV14984.1      CVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQTGKIAD
QIV14996.1      CVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQTGKIAD
QIV15008.1      CVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQTGKIAD
QIU78731.1      CVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQTGKIAD
QIU78743.1      CVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQTGKIAD
QIU78755.1      CVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQTGKIAD
QID21068.1      CVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQTGKIAD
QIV15140.1      CVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQTGKIAD
QIU78719.1      CVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQTGKIAD
QIQ49762.1      CVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQTGKIAD
QIX13867.1      CVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQTGKIAD
                ************************************************************
```

```
QIA98554.1      NGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPATVCGPKKSTNLVKNKCVN
QIV14984.1      NGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPATVCGPKKSTNLVKNKCVN
QIV14996.1      NGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPATVCGPKKSTNLVKNKCVN
QIV15008.1      NGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPATVCGPKKSTNLVKNKCVN
QIU78731.1      NGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPATVCGPKKSTNLVKNKCVN
QIU78743.1      NGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPATVCGPKKSTNLVKNKCVN
QIU78755.1      NGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPATVCGPKKSTNLVKNKCVN
QID21068.1      NGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPATVCGPKKSTNLVKNKCVN
QIV15140.1      NGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPATVCGPKKSTNLVKNKCVN
QIU78719.1      NGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPATVCGPXKSTNLVKNKCVN
QIQ49762.1      NGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPATVCGPKKSTNLVKNKCVN
QIX13867.1      NGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPATVCGPKKSTNLVKNKCVN
                *********************************************** ************
```

Outline

- SARS-Cov-2 is rapidly spreading globally
- Multiple sequence alignment and phylogenetic trees were created
- Spike Protein sequences show little to no difference in each of the 3 countries
- Multiple Sequence Alignment of RBDs is consistent within the 12 Sequences
- **Structure of RBD remained the same in each of the 12 sequences**
- RBD remained unchanged in all 3 countries
- Selective pressure in SARS CoV 2 does not call for mutation
- Future directions could expand on sequence/structure relationship

The structure of the RBD Remains the Same Within 12 Sequences



Receptor Binding Domain of SARS-CoV-2 S

Sequences obtained from 3 countries showed no difference in sequence alignment.

The structure of the receptor binding domain is identical for all sequences.

Outline

- SARS-Cov-2 is rapidly spreading globally
- Multiple sequence alignment and phylogenetic trees were created
- Spike Protein sequences show little to no difference in each of the 3 countries
- Multiple Sequence Alignment of RBDs is consistent within the 12 Sequences
- Structure remained the same in each of the 12 sequences
- **RBD remained unchanged in all 3 countries**
- Selective pressure in SARS CoV 2 does not call for mutation
- Future directions could expand on sequence/structure relationship

Receptor Binding Domain of Spike Protein is Unchanged Between 3 Countries

- The spike protein sequences for the SARS-CoV-2 were obtained from South Korea, Spain, and the United States.
- The receptor binding domain (RBD) within the sequence is known to interact with human ACE-2 receptor
- Multiple sequence alignment showed no differences in the sequences pertaining to the RBD.
- This indicates that the spike protein RBD has not evolved as SARS-CoV-2 travels across the globe.

Outline

- SARS-Cov-2 is rapidly spreading globally
- Multiple sequence alignment and phylogenetic trees were created
- Spike Protein sequences show little to no difference in each of the 3 countries
- Multiple Sequence Alignment of RBDs is consistent within the 12 Sequences
- Structure remained the same in each of the 12 sequences
- RBD remained unchanged in all 3 countries
- **Selective pressure in SARS CoV 2 does not call for mutation**
- Future directions could expand on sequence/structure relationship

Less Selective Pressure for SARS-CoV-2 Mutation

- SARS-CoV-2 is highly infectious due to the tight binding affinity of the RBD (Wrapp et. al)
- There is less selective pressure for the RBD to mutate because it is highly transmissible.
- As humans start to develop antibodies mutations within the amino acid sequence are likely.
- It is important to globally monitor the genome data as SARS-CoV-2 continues to spread over time.

Outline

- SARS-Cov-2 is rapidly spreading globally
- Multiple sequence alignment and phylogenetic trees were created
- Spike Protein sequences show little to no difference in each of the 3 countries
- Multiple Sequence Alignment of RBDs is consistent within the 12 Sequences
- Structure remained the same in each of the 12 sequences
- RBD remained unchanged in all 3 countries
- Selective pressure in SARS CoV 2 does not call for mutation
- **Future directions could expand on sequence/structure relationship**

Future Directions Could Expand on Sequence Structure Relationship

- It is important to collect and analyze global data throughout time to monitor potential changes in the RBD sequence as humans develop antibodies.
  - It is also important to look at viral response to a future vaccine
- Other directions could focus on comparing the sequences of SARS-CoV RBD to SARS-CoV-2 RBD.
  - Wrapp et. al indicated that differences may be present within the sequences which could impact the structure of the RBD
- Unfavorable interactions between (F486-T82 and L455-Q493) could mutate with more selective pressure (Wan et. al).

References

Wan, Y., Shang, J., Graham, R., Baric, R. S., & Li, F. (2020). Receptor recognition by the novel coronavirus from Wuhan: an analysis based on decade-long structural studies of SARS coronavirus. Journal of virology, 94(7). DOI: 10.1128/JVI.00127-20.

Wrapp, D., Wang, N., Corbett, K. S., Goldsmith, J. A., Hsieh, C. L., Abiona, O., ... & McLellan, J. S. (2020). Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. Science, 367(6483), 1260-1263.

Thank you to Dr. Dahlquist,
Loyola Marymount University Biology
Department, and the BIOL 368 class