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
- 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.

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Sequence</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>QIA98554.1</td>
<td>YNYKLPDDFTGCVIAMNNNLDSKVGGNYNYLYLRFLRKNLKPFFERDISTEIYQAGSTPC</td>
<td>QIA98554.1</td>
</tr>
<tr>
<td>QIV14984.1</td>
<td>YNYKLPDDFTGCVIAMNNNLDSKVGGNYNYLYLRFLRKNLKPFFERDISTEIYQAGSTPC</td>
<td>QIV14984.1</td>
</tr>
<tr>
<td>QIV14996.1</td>
<td>YNYKLPDDFTGCVIAMNNNLDSKVGGNYNYLYLRFLRKNLKPFFERDISTEIYQAGSTPC</td>
<td>QIV14996.1</td>
</tr>
<tr>
<td>QIV15008.1</td>
<td>YNYKLPDDFTGCVIAMNNNLDSKVGGNYNYLYLRFLRKNLKPFFERDISTEIYQAGSTPC</td>
<td>QIV15008.1</td>
</tr>
<tr>
<td>QIU78731.1</td>
<td>YNYKLPDDFTGCVIAMNNNLDSKVGGNYNYLYLRFLRKNLKPFFERDISTEIYQAGSTPC</td>
<td>QIU78731.1</td>
</tr>
<tr>
<td>QIU78743.1</td>
<td>YNYKLPDDFTGCVIAMNNNLDSKVGGNYNYLYLRFLRKNLKPFFERDISTEIYQAGSTPC</td>
<td>QIU78743.1</td>
</tr>
<tr>
<td>QIU78755.1</td>
<td>YNYKLPDDFTGCVIAMNNNLDSKVGGNYNYLYLRFLRKNLKPFFERDISTEIYQAGSTPC</td>
<td>QIU78755.1</td>
</tr>
<tr>
<td>QID21068.1</td>
<td>YNYKLPDDFTGCVIAMNNNLDSKVGGNYNYLYLRFLRKNLKPFFERDISTEIYQAGSTPC</td>
<td>QID21068.1</td>
</tr>
<tr>
<td>QIV15140.1</td>
<td>YNYKLPDDFTGCVIAMNNNLDSKVGGNYNYLYLRFLRKNLKPFFERDISTEIYQAGSTPC</td>
<td>QIV15140.1</td>
</tr>
<tr>
<td>QIU78719.1</td>
<td>YNYKLPDDFTGCVIAMNNNLDSKVGGNYNYLYLRFLRKNLKPFFERDISTEIYQAGSTPC</td>
<td>QIU78719.1</td>
</tr>
<tr>
<td>QIQ49762.1</td>
<td>YNYKLPDDFTGCVIAMNNNLDSKVGGNYNYLYLRFLRKNLKPFFERDISTEIYQAGSTPC</td>
<td>QIQ49762.1</td>
</tr>
<tr>
<td>QIX13867.1</td>
<td>YNYKLPDDFTGCVIAMNNNLDSKVGGNYNYLYLRFLRKNLKPFFERDISTEIYQAGSTPC</td>
<td>QIX13867.1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Sequence</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>QIA98554.1</td>
<td>NGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVLSFELLHATVCGBKSTNLVKNKCVN</td>
<td>QIA98554.1</td>
</tr>
<tr>
<td>QIV14984.1</td>
<td>NGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVLSFELLHATVCGBKSTNLVKNKCVN</td>
<td>QIV14984.1</td>
</tr>
<tr>
<td>QIV14996.1</td>
<td>NGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVLSFELLHATVCGBKSTNLVKNKCVN</td>
<td>QIV14996.1</td>
</tr>
<tr>
<td>QIV15008.1</td>
<td>NGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVLSFELLHATVCGBKSTNLVKNKCVN</td>
<td>QIV15008.1</td>
</tr>
<tr>
<td>QIU78731.1</td>
<td>NGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVLSFELLHATVCGBKSTNLVKNKCVN</td>
<td>QIU78731.1</td>
</tr>
<tr>
<td>QIU78743.1</td>
<td>NGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVLSFELLHATVCGBKSTNLVKNKCVN</td>
<td>QIU78743.1</td>
</tr>
<tr>
<td>QIU78755.1</td>
<td>NGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVLSFELLHATVCGBKSTNLVKNKCVN</td>
<td>QIU78755.1</td>
</tr>
<tr>
<td>QID21068.1</td>
<td>NGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVLSFELLHATVCGBKSTNLVKNKCVN</td>
<td>QID21068.1</td>
</tr>
<tr>
<td>QIV15140.1</td>
<td>NGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVLSFELLHATVCGBKSTNLVKNKCVN</td>
<td>QIV15140.1</td>
</tr>
<tr>
<td>QIU78719.1</td>
<td>NGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVLSFELLHATVCGBKSTNLVKNKCVN</td>
<td>QIU78719.1</td>
</tr>
<tr>
<td>QIQ49762.1</td>
<td>NGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVLSFELLHATVCGBKSTNLVKNKCVN</td>
<td>QIQ49762.1</td>
</tr>
<tr>
<td>QIX13867.1</td>
<td>NGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVLSFELLHATVCGBKSTNLVKNKCVN</td>
<td>QIX13867.1</td>
</tr>
</tbody>
</table>

RBD Sequence Consistent with Wrap et al.
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

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

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


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