Mutation at the 501 Position of SARS-CoV-2 Results in a Higher Binding Affinity to the Human ACE2 Receptor

Kam Taghizadeh, Fatima Alghanem
10/15/2020

Outline

- Interaction of SARS-CoV-2 and SARS-CoV to human ACE2
- Future evolution of SARS-CoV-2
- Mutation at position 501 leads to a higher infectious strain
- Our research question: Why does the viral strain become more infectious if there is a mutation at position 501?
- Our methods and results of creating a sequence alignment and phylogenetic tree
- Future directions we would take

Interaction of SARS-CoV-2 and ACE2

- SARS-CoV-2 spike protein receptor-binding domain (RBD) specifically recognizes its host receptor angiotensin-converting enzyme 2 (ACE2) (Wan et al, 2020)

- Research has shown that the specific host that gets infected is dependent on the binding affinity between the viral RBD and host ACE2 in the primary viral attachment step (Wan et al, 2020)

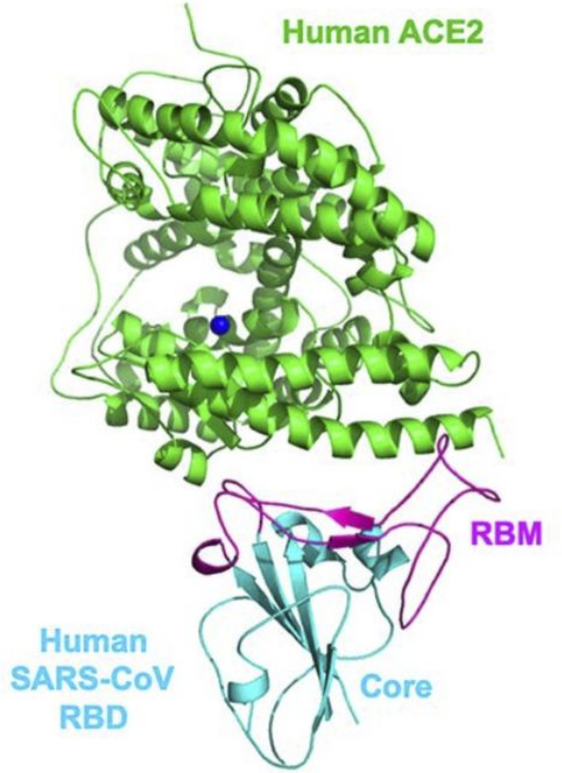
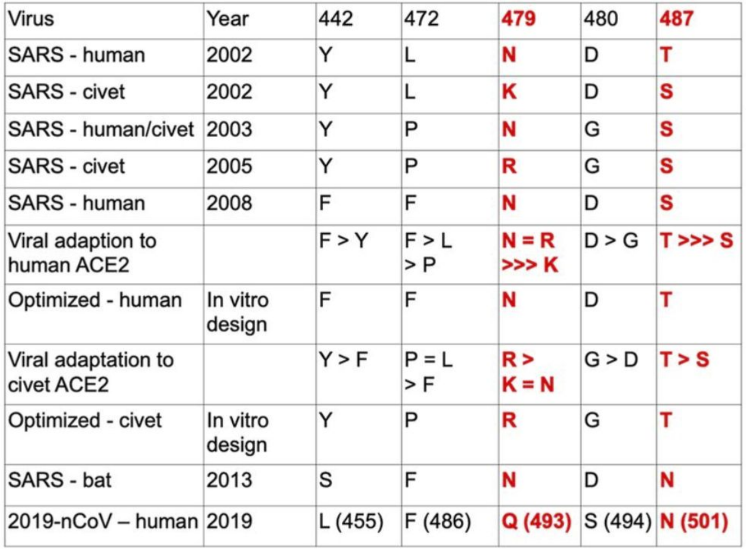


Figure 1: Human SARS-CoV: Human ACE2 (green), RBM (magenta), core (blue)

Interaction of SARS-CoV-2 and ACE2

- Two virus-binding hotspots have been discovered on human ACE2
  - Various mutations have occurred near these hotspots, including residue 487 in SARS-CoV, which corresponds to residue 501 in SARS-CoV-2 (Wan et al, 2020)
  - Residue 487 in SARS-CoV (501 in SARS-CoV-2) is known to enhance binding affinity to human ACE2 (Wan et al, 2020)
  - Mutation of residue 487 in SARS-CoV strengthens the structural stability and enhances the viral binding to human Hotspot Lys353 in ACE2 (Wan et al, 2020)



Virus	Year	442	472	**479**	480	**487**
SARS - human	2002	Y	L	**N**	D	**T**
SARS - civet	2002	Y	L	**K**	D	**S**
SARS - human/civet	2003	Y	P	**N**	G	**S**
SARS - civet	2005	Y	P	**R**	G	**S**
SARS - human	2008	F	F	**N**	D	**S**
Viral adaption to human ACE2		F > Y	F > L > P	**N = R >>> K**	D > G	**T >>> S**
Optimized - human	In vitro design	F	F	**N**	D	**T**
Viral adaptation to civet ACE2		Y > F	P = L > F	**R > K = N**	G > D	**T > S**
Optimized - civet	In vitro design	Y	P	**R**	G	**T**
SARS - bat	2013	S	F	**N**	D	**N**
2019-nCoV – human	2019	L (455)	F (486)	**Q (493)**	S (494)	**N (501)**

Figure 2: Comparison of the viral strains, note position 487 and its correspondence to position 501 in 2019-nCoV-human

Outline

- Interaction of SARS-CoV-2 and SARS-CoV to human ACE2
- Mutation at position 501 leads to a higher infectious strain
- Our research question: Why does the viral strain become more infectious if there is a mutation at position 501?
- Our methods and results of creating a sequence alignment and phylogenetic tree
- Future directions we would take

Future Evolution of SARS-CoV-2

- Asn501 is among the top twenty residues that will most likely mutate and have an adversarial effect on infectivity

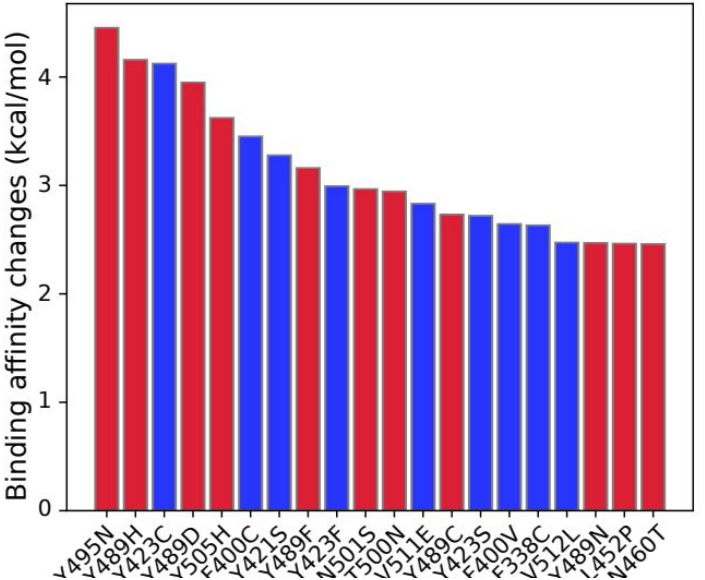
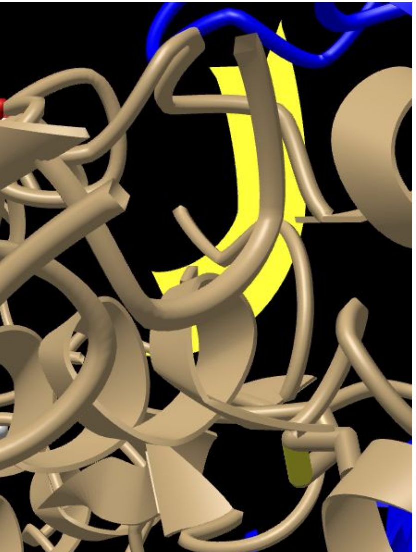
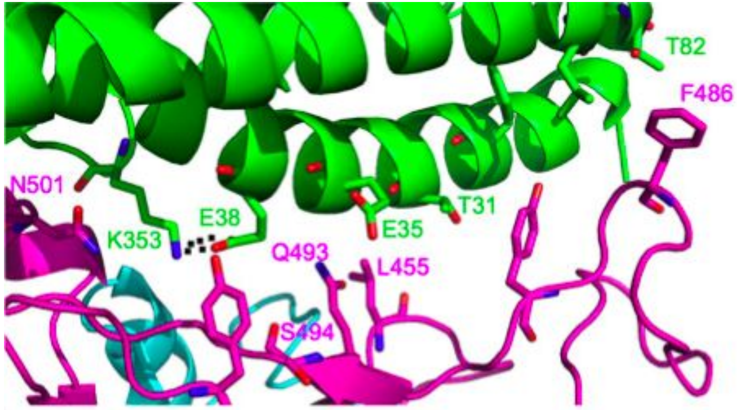


Figure 3: Top 20 amino acids most likely to mutate (Red=mutations on RBD, blue=mutations away from RBD) https://arxiv.org/pdf/2005.14669.pdf

Research Question- Why does the viral strain become more infectious if there is a mutation at position 501?

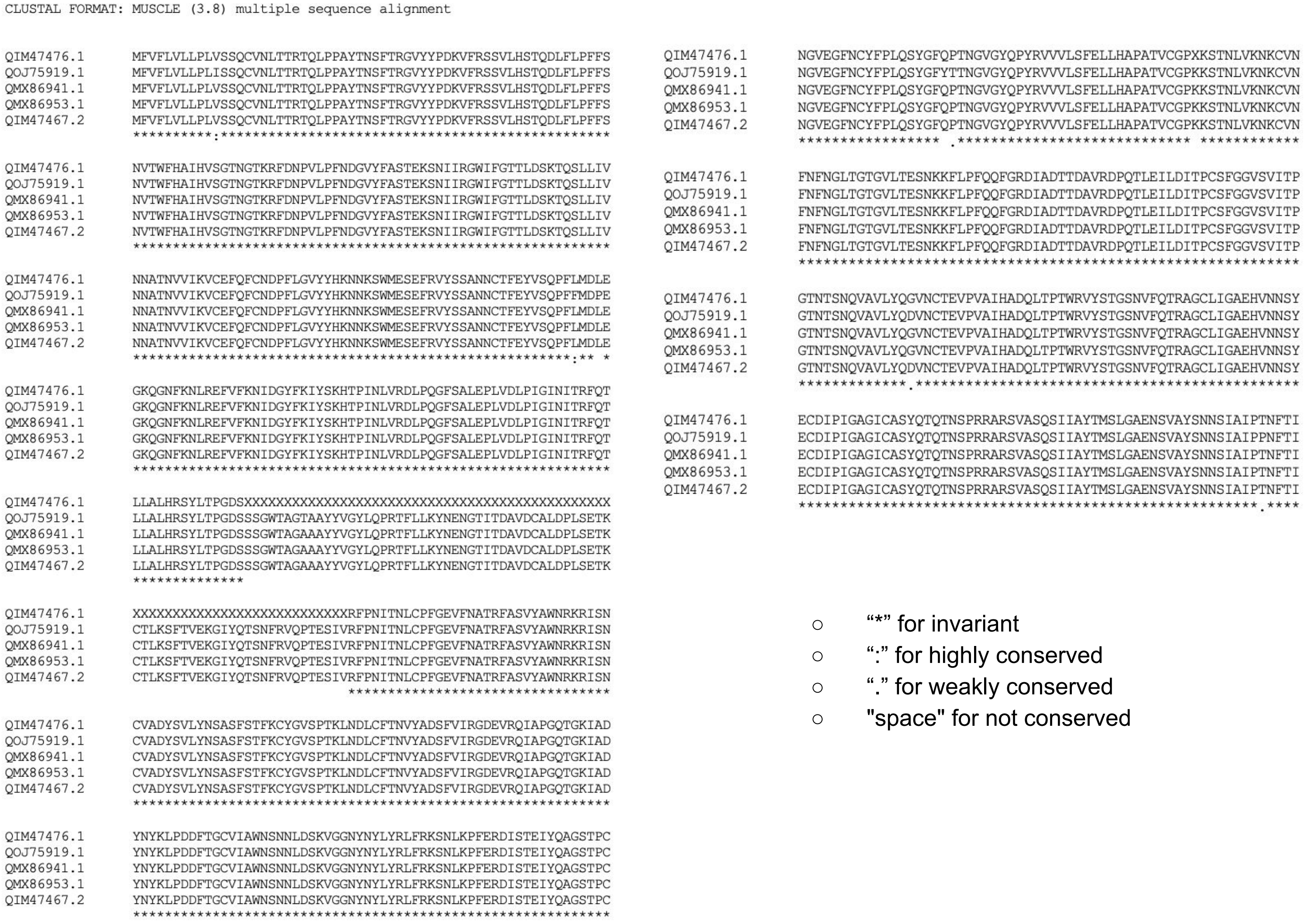


- According to Chunyan Yi et al. , the enhanced binding affinity may be due to greater support provided by residue 501 in SARS-CoV-2 or 487 in SARS-CoV for the stabilization of the overall structure of RBD, or the strengthening of hydrophilic reactions



We Wanted to Find Strains that Have Evolved at Position 501

- We decided to see if we could find a SARS-CoV-2 spike protein sequence that would have a mutation at position 501
- We visited the NCBI Virus SARS-CoV-2 Data Hub to collect two sequences each from different geographic regions
  - United Kingdom, Spain, Iran, Saudi Arabia, Italy, Venezuela, Peru, and Brazil
- All of the sequences were downloaded in FASTA format
  - We used google character count to see what the nucleotide for each sequence was at position 501
- We then created a clustal alignment by visiting phylogeny.fr
  - We scrolled down the page to the section called ‘Phylogeny analysis’, and clicked on the text ‘One Click’.
  - Uploaded the set of sequences in FASTA format into the large text box, and clicked submit
  - Clicked on ‘3. Alignment’, then under ‘outputs’ clicked on ‘Alignment in Clustal format’
  - Results are shown on next few slides



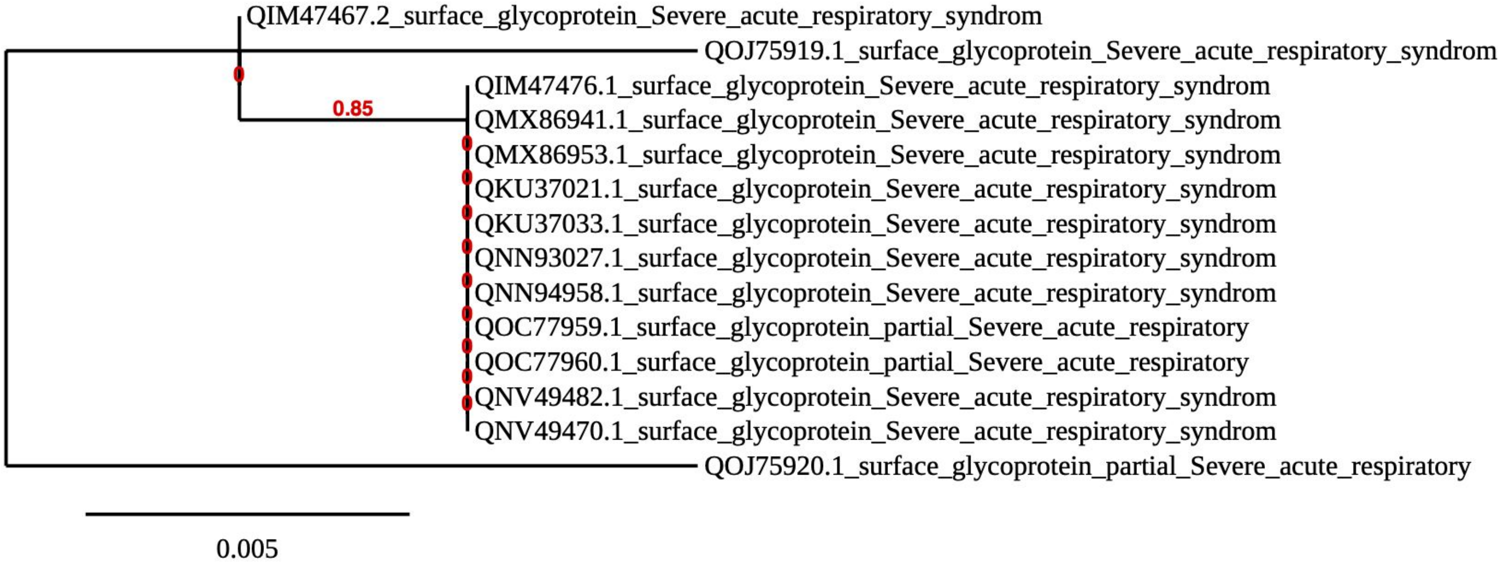
```
CLUSTAL FORMAT: MUSCLE (3.8) multiple sequence alignment


QIM47476.1      MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFLPFFS
QOJ75919.1      MFVFLVLLPLISSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFLPFFS
QMX86941.1      MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFLPFFS
QMX86953.1      MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFLPFFS
QIM47467.2      MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFLPFFS
                **********:*************************************************

QIM47476.1      NVTWFHAIHVSGTNGTKRFDNPVLPFNDGVYFASTEKSNIIRGWIFGTTLDSKTQSLLIV
QOJ75919.1      NVTWFHAIHVSGTNGTKRFDNPVLPFNDGVYFASTEKSNIIRGWIFGTTLDSKTQSLLIV
QMX86941.1      NVTWFHAIHVSGTNGTKRFDNPVLPFNDGVYFASTEKSNIIRGWIFGTTLDSKTQSLLIV
QMX86953.1      NVTWFHAIHVSGTNGTKRFDNPVLPFNDGVYFASTEKSNIIRGWIFGTTLDSKTQSLLIV
QIM47467.2      NVTWFHAIHVSGTNGTKRFDNPVLPFNDGVYFASTEKSNIIRGWIFGTTLDSKTQSLLIV
                ************************************************************

QIM47476.1      NNATNVVIKVCEFQFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVSQPFLMDLE
QOJ75919.1      NNATNVVIKVCEFQFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVSQPFFMDPE
QMX86941.1      NNATNVVIKVCEFQFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVSQPFLMDLE
QMX86953.1      NNATNVVIKVCEFQFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVSQPFLMDLE
QIM47467.2      NNATNVVIKVCEFQFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVSQPFLMDLE
                *******************************************************:** *

QIM47476.1      GKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRDLPQGFSALEPLVDLPIGINITRFQT
QOJ75919.1      GKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRDLPQGFSALEPLVDLPIGINITRFQT
QMX86941.1      GKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRDLPQGFSALEPLVDLPIGINITRFQT
QMX86953.1      GKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRDLPQGFSALEPLVDLPIGINITRFQT
QIM47467.2      GKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRDLPQGFSALEPLVDLPIGINITRFQT
                ************************************************************

QIM47476.1      LLALHRSYLTPGDSXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
QOJ75919.1      LLALHRSYLTPGDSSSGWTAGTAAYYVGYLQPRTFLLKYNENGTITDAVDCALDPLSETK
QMX86941.1      LLALHRSYLTPGDSSSGWTAGAAAYYVGYLQPRTFLLKYNENGTITDAVDCALDPLSETK
QMX86953.1      LLALHRSYLTPGDSSSGWTAGAAAYYVGYLQPRTFLLKYNENGTITDAVDCALDPLSETK
QIM47467.2      LLALHRSYLTPGDSSSGWTAGAAAYYVGYLQPRTFLLKYNENGTITDAVDCALDPLSETK
                **************                                              

QIM47476.1      XXXXXXXXXXXXXXXXXXXXXXXXXXXRFPNITNLCPFGEVFNATRFASVYAWNRKRISN
QOJ75919.1      CTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISN
QMX86941.1      CTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISN
QMX86953.1      CTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISN
QIM47467.2      CTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISN
                                           *********************************

QIM47476.1      CVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQTGKIAD
QOJ75919.1      CVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQTGKIAD
QMX86941.1      CVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQTGKIAD
QMX86953.1      CVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQTGKIAD
QIM47467.2      CVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQTGKIAD
                ************************************************************

QIM47476.1      YNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPC
QOJ75919.1      YNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPC
QMX86941.1      YNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPC
QMX86953.1      YNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPC
QIM47467.2      YNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPC
                ************************************************************

QIM47476.1      NGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPATVCGPXKSTNLVKNKCVN
QOJ75919.1      NGVEGFNCYFPLQSYGFYTTNGVGYQPYRVVVLSFELLHAPATVCGPKKSTNLVKNKCVN
QMX86941.1      NGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPATVCGPKKSTNLVKNKCVN
QMX86953.1      NGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPATVCGPKKSTNLVKNKCVN
QIM47467.2      NGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPATVCGPKKSTNLVKNKCVN
                ***************** .**************************** ************

QIM47476.1      FNFNGLTGTGVLTESNKKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVITP
QOJ75919.1      FNFNGLTGTGVLTESNKKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVITP
QMX86941.1      FNFNGLTGTGVLTESNKKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVITP
QMX86953.1      FNFNGLTGTGVLTESNKKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVITP
QIM47467.2      FNFNGLTGTGVLTESNKKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVITP
                ************************************************************

QIM47476.1      GTNTSNQVAVLYQGVNCTEVPVAIHADQLTPTWRVYSTGSNVFQTRAGCLIGAEHVNNSY
QOJ75919.1      GTNTSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGSNVFQTRAGCLIGAEHVNNSY
QMX86941.1      GTNTSNQVAVLYQGVNCTEVPVAIHADQLTPTWRVYSTGSNVFQTRAGCLIGAEHVNNSY
QMX86953.1      GTNTSNQVAVLYQGVNCTEVPVAIHADQLTPTWRVYSTGSNVFQTRAGCLIGAEHVNNSY
QIM47467.2      GTNTSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGSNVFQTRAGCLIGAEHVNNSY
                *************.**********************************************

QIM47476.1      ECDIPIGAGICASYQTQTNSPRRARSVASQSIIAYTMSLGAENSVAYSNNSIAIPTNFTI
QOJ75919.1      ECDIPIGAGICASYQTQTNSPRRARSVASQSIIAYTMSLGAENSVAYSNNSIAIPPNFTI
QMX86941.1      ECDIPIGAGICASYQTQTNSPRRARSVASQSIIAYTMSLGAENSVAYSNNSIAIPTNFTI
QMX86953.1      ECDIPIGAGICASYQTQTNSPRRARSVASQSIIAYTMSLGAENSVAYSNNSIAIPTNFTI
QIM47467.2      ECDIPIGAGICASYQTQTNSPRRARSVASQSIIAYTMSLGAENSVAYSNNSIAIPTNFTI
                *******************************************************.****
```

- “*” for invariant
- “:” for highly conserved
- “.” for weakly conserved
- "space" for not conserved

```
QIM47476.1      SVTTEILPVSMTKTSVDCTMYICGDSTECSNLLLQYGSFCTQLNRALTGIAVEQDKNTQE
QOJ75919.1      SVTTEFLPVSMTKTSVDCTMYICGDSTECSNLLLQYGSFCTQLNRALTGIAVEQDKNTQE
QMX86941.1      SVTTEILPVSMTKTSVDCTMYICGDSTECSNLLLQYGSFCTQLNRALTGIAVEQDKNTQE
QMX86953.1      SVTTEILPVSMTKTSVDCTMYICGDSTECSNLLLQYGSFCTQLNRALTGIAVEQDKNTQE
QIM47467.2      SVTTEILPVSMTKTSVDCTMYICGDSTECSNLLLQYGSFCTQLNRALTGIAVEQDKNTQE
                *****:******************************************************

QIM47476.1      VFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRSFIEDLLFNKVTLADAGFIKQYGDC
QOJ75919.1      VFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRSFIEDLLFNKVTLADAGFIKQYGDC
QMX86941.1      VFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRSFIEDLLFNKVTLADAGFIKQYGDC
QMX86953.1      VFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRSFIEDLLFNKVTLADAGFIKQYGDC
QIM47467.2      VFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRSFIEDLLFNKVTLADAGFIKQYGDC
                ************************************************************

QIM47476.1      LGDIAARDLICAQKFNGLTVLPPLLTDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAM
QOJ75919.1      LGDIAARDLICAQKFNGLTVLPPLLTDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAM
QMX86941.1      LGDIAARDLICAQKFNGLTVLPPLLTDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAM
QMX86953.1      LGDIAARDLICAQKFNGLTVLPPLLTDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAM
QIM47467.2      LGDIAARDLICAQKFNGLTVLPPLLTDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAM
                ************************************************************

QIM47476.1      QMAYRFNGIGVTQNVLYENQKLIANQFNSAIGKIQDSLSSTASALGKLQDVVNQNAQALN
QOJ75919.1      QMAYRFNGIGVTQNVLYENQKLIANQFNSAIGKIQDSLSSTASALGKLQDVVNQNAQALN
QMX86941.1      QMAYRFNGIGVTQNVLYENQKLIANQFNSAIGKIQDSLSSTASALGKLQDVVNQNAQALN
QMX86953.1      QMAYRFNGIGVTQNVLYENQKLIANQFNSAIGKIQDSLSSTASALGKLQDVVNQNAQALN
QIM47467.2      QMAYRFNGIGVTQNVLYENQKLIANQFNSAIGKIQDSLSSTASALGKLQDVVNQNAQALN
                ************************************************************

QIM47476.1      TLVKQLSSNFGAISSVLNDILSRLDKVEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRA
QOJ75919.1      TLVKQLSSNFGAISSVLNDILSRLDKVEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRA
QMX86941.1      TLVKQLSSNFGAISSVLNDILSRLDKVEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRA
QMX86953.1      TLVKQLSSNFGAISSVLNDILSRLDKVEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRA
QIM47467.2      TLVKQLSSNFGAISSVLNDILSRLDKVEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRA
                ************************************************************

QIM47476.1      SANLAATKMSECVLGQSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFTTAPA
QOJ75919.1      SANLAATKMSECVLGQSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFTTAPA
QMX86941.1      SANLAATKMSECVLGQSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFTTAPA
QMX86953.1      SANLAATKMSECVLGQSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFTTAPA
QIM47467.2      SANLAATKMSECVLGQSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFTTAPA
                ************************************************************

QIM47476.1      ICHDGKAHFPREGVFVSNGTHWFVTQRNFYEPQIITTDNTFVSGNCDVVIGIVNNTVYDP
QOJ75919.1      ICHDGKAHFPREGVFVSNGTHWFVTQRNFYEPQIITTDNTFVSGNCDVVIGIVNNTVYDP
QMX86941.1      ICHDGKAHFPREGVFVSNGTHWFVTQRNFYEPQIITTDNTFVSGNCDVVIGIVNNTVYDP
QMX86953.1      ICHDGKAHFPREGVFVSNGTHWFVTQRNFYEPQIITTDNTFVSGNCDVVIGIVNNTVYDP
QIM47467.2      ICHDGKAHFPREGVFVSNGTHWFVTQRNFYEPQIITTDNTFVSGNCDVVIGIVNNTVYDP
                ************************************************************

QIM47476.1      LQPELDSFKEELDKYFKNHTSPDVDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDL
QOJ75919.1      LQPELDSFKEELDKYFKNHTSPDVDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDL
QMX86941.1      LQPELDSFKEELDKYFKNHTSPDVDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDL
QMX86953.1      LQPELDSFKEELDKYFKNHTSPDVDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDL
QIM47467.2      LQPELDSFKEELDKYFKNHTSPDVDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDL
                ************************************************************

QIM47476.1      QELGKYEQYIKWPWYIWLGFIAGLIAIVMVTIMLCCMTSCCSCLKGCCSCGSCCKFDEDD
QOJ75919.1      QELGKYEQYIKWPWYIWLGFIAGLIAIVMVTIMLCCMTSCCSCLKGCCSCGSCCKFDEDD
QMX86941.1      QELGKYEQYIKWPWYIWLGFIAGLIAIVMVTIMLCCMTSCCSCLKGCCSCGSCCKFDEDD
QMX86953.1      QELGKYEQYIKWPWYIWLGFIAGLIAIVMVTIMLCCMTSCCSCLKGCCSCGSCCKFDEDD
QIM47467.2      QELGKYEQYIKWPWYIWLGFIAGLIAIVMVTIMLCCMTSCCSCLKGCCSCGSCCKFDEDD
                ************************************************************

QIM47476.1      SEPVLKGVKLHYT
QOJ75919.1      SEPVLKGVKLHYT
QMX86941.1      SEPVLKGVKLHYT
QMX86953.1      SEPVLKGVKLHYT
QIM47467.2      SEPVLKGVKLHYT
                *************
```

Created a Phylogenetic Tree

- We then created a phylogenetic tree by visiting phylogeny.fr
    - We scrolled down the page to the section called ‘Phylogeny analysis’, and clicked on the text ‘One Click’.
    - Uploaded the set of sequences in FASTA format into the large text box, and clicked submit
    - Once completed, we clicked on ‘tree rendering’ to get the phylogenetic tree shown in the next slide


QIM47467.2_surface_glycoprotein_Severe_acute_respiratory_syndrom
QOJ75919.1_surface_glycoprotein_Severe_acute_respiratory_syndrom
0
QIM47476.1_surface_glycoprotein_Severe_acute_respiratory_syndrom
0.85
QMX86941.1_surface_glycoprotein_Severe_acute_respiratory_syndrom
0
QMX86953.1_surface_glycoprotein_Severe_acute_respiratory_syndrom
0
QKU37021.1_surface_glycoprotein_Severe_acute_respiratory_syndrom
0
QKU37033.1_surface_glycoprotein_Severe_acute_respiratory_syndrom
0
QNN93027.1_surface_glycoprotein_Severe_acute_respiratory_syndrom
0
QNN94958.1_surface_glycoprotein_Severe_acute_respiratory_syndrom
0
QOC77959.1_surface_glycoprotein_partial_Severe_acute_respiratory
0
QOC77960.1_surface_glycoprotein_partial_Severe_acute_respiratory
0
QNV49482.1_surface_glycoprotein_Severe_acute_respiratory_syndrom
0
QNV49470.1_surface_glycoprotein_Severe_acute_respiratory_syndrom
QOJ75920.1_surface_glycoprotein_partial_Severe_acute_respiratory
0.005

Sequences only showed Asn at Position 501

- Our results show most of the sequences having Asparagine at position 501
- The results obtained do not support our research question, due to not finding a mutation at the 501 position of any of the sequences downloaded
- Research has shown that a change at 501 position would have adversarial effects
  - Promotes hydrophilic network
  - Stabilizes structure of RBD
  - Asparagine has a high propensity to hydrogen bond

Larger Number of Sequences May be Analyzed to Find Mutation at Position 501

- We look forward to researching a larger number of sequences
- Our focus would be on Sequences that have a mutation of Asparagine (n) to Threonine (t) at position 501 as previous literature showed that this kind of mutation results in a higher binding affinity to human ACE2 receptors (Wan et al, 2020)
- We want to explore the effects of other mutations at position 501 on spike protein affinity to human ACE2 receptores and if there's a an amino mutation that would cause less/ no affinity to ACE2.