

SARS-CoV-2 Spike Protein Evolution may Cause Difficulties for Vaccine

Jian Zhou

second xiangya hospital

Sun Jingjing

zhejiang university

Gang Lu

shandong first hospital

Wanchun Wang (✉ wanchun.wang@csu.edu.cn)

Second Xiangya Hospital

Lin Wang

Jinan Children Hospital

Research article

Keywords: COVID-19, SARS-CoV-2, S protein, Epitopes, Structure

Posted Date: October 7th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-80010/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: Coronavirus disease 2019 (COVID-19) poses a great threat to human health and life. We performed a bioinformatics analysis to compare the sequence, structure, and epitopes of SARS-CoV-2 spike (S) protein in 10 different countries.

Methods: The amino acid sequences of SARS-CoV-2 S protein were obtained from the NCBI database. We used DNASTAR Lasergene software to analyze the protein's secondary structures. SWISS-MODEL combined with VMD software was used to construct a 3D model of SARS-CoV-2 S protein. DNASTAR Protean and the IEDB database were used to analyze the B cell epitopes and T cell epitopes, respectively.

Results: The results of B cell epitopes analysis indicated that the epitopes of SARS-CoV-2 S protein in Korea and American increased, which suggested that the antigenicity of SARS-CoV-2 in Czech, Korea and American might be enhanced. A small number of B cell epitopes disappeared in the SARS-CoV-2 S protein sequence from Greece, Australia, Sweden and India, which suggested that the antigenicity of SARS-CoV-2 in Greece, Australia, Sweden and India may be weakened. T cell epitope analysis indicated that the antigenicity of SARS-CoV-2 in Czech, Korea and American was enhanced, while antigenicity of SARS-CoV-2 in Greece, Australia, India, Sweden and Thailand may be weakened. The sequence of SARS-CoV-2 S protein has changed as the virus has spread, and the structures and epitopes have changed accordingly.

Conclusion: The mutation leads to a decrease in the antigenicity of SARS-CoV-2, which may be a mechanism for the virus to evade surveillance by the immune system.

Background

In December 2019, coronavirus disease 2019 (COVID-19) was discovered in Wuhan, China, and subsequently spread rapidly across the country and even the world via a clear human-to-human route [1-3]. People with COVID-19 often show respiratory symptoms [4-7]. In severe cases, the infection can cause pneumonia, severe acute respiratory syndrome, kidney failure, and even death [8,9]. Currently, no effective vaccines have been found. Understanding the evolution and outcome of COVID-19 is important for prevention and control of the disease. COVID-19 is caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), which belongs to the genus *Betacoronavirus*. Of the 30 kb SARS-CoV-2 genome, only about 10 kb is the coding region, encoding spike protein (S protein), membrane protein (M protein), envelope protein (E protein), and nucleocapsid protein (Nucleocapsid) [10,11].

SARS-CoV-2 S protein can be cleaved into two subunits, S1 and S2, by the host's protease. The S1 subunit is located at the top of the envelope spike and mediates the virus's binding to the receptor on the host cell surface. The S2 subunit forms a spiked stem, which mediates the fusion of the viral envelope with the host cell membrane and promotes the virus to enter the host cell. Therefore, the S protein of SARS-CoV-2 plays an important role in the process of viral invasion [12,13]. The SARS-CoV-2 S protein is an important target for research into diagnostics, therapeutic antibodies, and vaccine design for COVID-

19. Understanding this protein has the potential to aid SARS-CoV-2 prevention efforts, as disrupting the S protein can block the virus's adsorption and invasion on host cells.

Previous studies have reported the amino acid sequence of SARS-CoV-2 in several countries [14]. However, viruses often mutate during transmission. In regards to viral mutation, there are two possible trends. First, the virus may become more and more antigenic and pathogenic. In this case, either the virus would be destroyed with the host, or people would adopt powerful measures to control transmission, eliminating the virus at the level of human transmission. In a second possibility, mutations could cause the virus to become less and less antigenic and pathogenic, allowing the virus to coexist with the host.

This study intends to compare the sequences, secondary structures, and epitopes of SARS-CoV-2 S protein published by 10 countries, providing a theoretical basis for researching the function and vaccine design of SARS-CoV-2.

Methods

SARS-CoV-2 S protein amino acid sequence

We obtained the amino acid sequence of the SARS-CoV-2 S protein from ten countries from the NCBI database (<https://www.ncbi.nlm.nih.gov/>). The NCBI GenBank accession numbers of these sequences were shown in table 1.

Comparison of SARS-CoV-2 S protein sequence

DNASTAR SeqBuilder was used to construct readable sequences of SARS-CoV-2 S protein. Furthermore, DNASTAR Megalign was used to compare the sequences of SARS-CoV-2 S protein.

Analysis of SARS-CoV-2 S protein secondary structure

DNASTAR Lasergene is a comprehensive sequence tool issued by DNASTar Corporation in the United States. This molecular biology software can be useful for molecular biology, genomics, and structural biology research. It is bioinformatics software for molecular cloning, primer design, next-gen sequencing, multiple sequence alignments, and transcriptome analysis. In this study, we used DNASTAR Lasergene (Madison, WI, USA) to analyze the secondary structures of SARS-CoV-2 S protein via the Garnier-Robson method.

Analysis of SARS-CoV-2 S protein 3D model

SWISS-MODEL is a protein structure online service (<http://swissmodel.expasy.org/>) [15,16]. Visual Molecular Dynamics (VMD) is molecular visualization software for displaying, animating, and analyzing

large biomolecular systems using 3D graphics. VMD software is typically used for reading standard Protein Data Bank (PDB) files and displaying the corresponding structure [17,18]. In this study, we used SWISS-MODEL online service and VMD software to analyze the 3D model of SARS-CoV-2 S protein.

Analysis of SARS-CoV-2 S protein B cell epitopes

Epitopes determine antigen specificity of a protein, which is the foundation of protein antigenicity [19]. There are many antigen indexes that can be used to assess the antigenicity of a protein, such as charge distribution secondary structure, hydrophilicity, accessibility, and flexibility [20-22]. Currently, there is no perfect method to analyze antigenic epitopes, but there are several rules that can be followed to analyze which fragments of a protein will be antigenic [23]. First, antigenic epitopes should be located in solvent-accessible sections, which contain hydrophobic and hydrophilic residues [24,25]. Second, the peptides with long loops connecting secondary structure motifs will be selected, preferably. In accordance with the rules outlined above, B cell epitopes, hydrophilicity, antigenic index, and surface probability of SARS-CoV-2 S protein were analyzed using DNASTAR Protean software.

Analysis of SARS-CoV-2 S protein T cell epitopes

The Immune Epitope Database (http://tools.immuneepitope.org/analyze/html/mhc_II_binding.html) online service was used to analyze the half-maximal inhibitory concentration (IC50) values of peptides that bind to the major histocompatibility complex (MHC) class II molecules of SARS-CoV-2 S protein.

Results

Search results

We retrieved a total of 2,000 sequences. After excluding sequences that did not contain mutated amino acids, we retained 116 sequences with mutated amino acids for analysis. After further excluding 96 sequences that contain repeated mutated amino acids (453 mutated amino acids: 1 repeated mutant sequence; 614 mutated amino acids: 93 repeated mutant sequence; 829 mutated amino acids: 2 repeated mutant sequence), 20 sequences from 10 different countries containing mutated amino acids were used for analysis. SARS-CoV-2 S protein contains 1273 amino acids.

Comparison of amino acid sequences of SARS-CoV-2 S protein

The comparison of amino acid sequences of SARS-CoV-2 S protein between 20 sequences was analyzed using DNASTAR Megalign software. As shown in Figure 1 & Table 1, 18 sequences contained different mutant amino acid and 2 sequences (No. 7 and No. 9) from American contained 2 mutant amino acid.

Table 1
Characteristics and epitopes change of sequence.

No.	Accession numbers	Country	Mutant amino acid position	Original amino acid	Mutant amino acid	B cell epitopes	T cell epitopes
1/A	MT019533.1	China	-	-	-	-	-
2/B	MT450891.1	American	5	L	F	-	-
3/C	MT371568.1	Czech	115	Q	R	-	↑
4/D	MT328034.1	Greece	197	I	V	↓	↓
5/E	MT039890.1	korea	221	S	W	↑	↑
6/F	MT007544.1	Australia	247	S	R	↓	↓
7/G	MT457395.1	American	261/453	G/Y	D/F	↓/-	-/↑
8/H	MT358637.1	India	271	Q	R	-	↓
9/I	MT370991.1	American	323/1248	T/C	I/F	↑/↑	↑/↑
10/J	MT457401.1	American	367	V	F	↑	↑
11/K	MT457391.1	American	501	N	T	-	-
12/L	MT447170.1	Thailand	614	D	G	-	-
13/M	MT371019.1	American	631	P	S	-	-
14/N	MT327745.1	Turkey	771	V	I	-	-
15/O	QIC53204.1	Sweden	797	F	C	↓	↓
16/P	MT447176.1	Thailand	829	A	T	-	↓
17/Q	MT050493.1	India	930	A	V	↓	↓
18/R	MT444634.1	American	939	S	F	↑	↑
19/S	MT451876.1	India	1181	K	R	-	↑
20/T	MT396242.1	India	1250	C	F	↓	-

These 20 sequences had more than 99% similarity (Figure 2). The comparison of SARS-CoV-2 S protein sequence between different countries indicated that SARS-CoV-2 S protein may have mutated as it has spread.

Comparison of secondary structures of SARS-CoV-2 S protein

A protein's secondary structure poses a significant impact on the virus's antigenic epitopes. Usually, α -helices and β -sheets have a higher chemical bond energy. They are solid to maintain the protein's structure, leading to difficulty in interacting with antibodies. The α -helices and β -sheets exist in the internal structure of a protein and generally do not act as antigenic epitopes. Meanwhile, the turn and coil, which are exposed on the surface of a protein, are usually looser and prone to distortion. Potential epitopes are usually on the protein's turn and coil. In order to further understand the structural differences of SARS-CoV-2 S protein between different countries, we used DNASTAR Lasergene to analyze the secondary structures. As shown in Figure 3, an additional turn region appeared in the mutation area of the sequence from Korea (No. 5/E) and American (No. 9/I), which suggested that more antigenic epitopes may have appeared here, and the antigenicity of SARS-CoV-2 in Korea and American may be enhanced. For the sequence from Australia (No. 6/F), an additional α -helix region appeared in the sequence's mutation area, while the coil region and turn region disappeared in the mutation area of the sequence from Sweden (No. 15/O) and India (No. 17/Q), respectively. This suggests that the antigenicity of SARS-CoV-2 in Sweden and India may be weakened.

Comparison of 3D models of SARS-CoV-2 S protein

The spatial structure of proteins can indicate their biological function. Further analysis of the tertiary structures of proteins is the ultimate goal of predicting a protein's structure. Understanding protein structures is of great significance to understand the relationship between structure and function. In the present study, we used SWISS-MODEL online service combined with VMD software to compare the 3D structure of SARS-CoV-2 S protein in five countries. As shown in Figure 4, most of the mutated amino acids were located on the outside of the SARS-CoV-2 S protein, including turn and coil regions, which suggests that the exterior of the protein structure is exposed to the external environment and seems to be prone to mutations.

Comparison of B cell epitopes of SARS-CoV-2 S protein

Epitopes are a special chemical group in an antigen molecule that determines antigen specificity, and include B-cell epitopes and T-cell epitopes. B-cell epitopes are composed of hydrophilic amino acids on the surface of the antigen. They are easily accessible and recognized by B-cell receptors and antibody molecules, and generally consist of 3–5 amino acid residues. The term "T cell epitope" refers to the epitope recognized by the T cell receptor, which is mostly present in the antigen molecule. It needs to be processed by antigen-presenting cells to combine with MHC molecules to form a complex before it can be recognized by TCR. In the present study, the B cell epitopes of SARS-CoV-2 S protein in 10 countries were predicted by DNASTAR Protean software. The result showed that the epitopes of SARS-CoV-2 S protein in

Korea and American increased, while a small number of B cell epitopes disappeared in the SARS-CoV-2 S protein sequence from Greece, Australia, Sweden and India, which suggested that the antigenicity of SARS-CoV-2 in Greece, Australia, Sweden and India may be weakened.

Comparison of T cell epitopes of SARS-CoV-2 S protein

In order to analyze the T cell epitopes of SARS-CoV-2 S protein from five countries, we used the Immune Epitope Database (http://tools.immuneepitope.org/analyze/html/mhc_II_binding.html) (IEDB) online service for analysis. The IC50 values of peptides binding to the MHC class II molecules of SARS-CoV-2 S protein in 10 countries were predicted with IEDB. As shown in Appendix tables 1, the minimum percentile ranks of MHC II alleles on the S protein mutation area in each of the 20 sequences were listed. According to the results of IEDB analysis, the percentile rank score of the sequence in Czech, Korea and American was lower than in control (Appendix table 1), while the percentile rank scores of the sequences from Greece, Australia, India, Sweden and Thailand were higher than the control. This suggested that the antigenicity of SARS-CoV-2 in Czech, Korea and American may be enhanced, while antigenicity of SARS-CoV-2 in Greece, Australia, India, Sweden and Thailand may be weakened. The results of T cell epitope analysis were consistent with the results of secondary structure analysis.

Discussion

S protein is the largest structural protein of SARS-CoV-2 [26]. Previous studies have reported that SARS-CoV-2 S protein can induce the immune system to produce neutralizing antibodies [27]. S protein is the most antigenic protein of SARS-CoV-2, and it is also the main target protein for the development of vaccines and clinical diagnosis.

DNA has a double-helix structure that is relatively stable and does not often make mistakes during replication and propagation. RNA, however, has a single-stranded structure, which lacks this stability, making it more likely than DNA to mutate during replication and propagation. SARS-CoV-2, an RNA virus, is likely to mutate during transmission. In order to understand if the sequence, structure, and epitopes of SARS-CoV-2 have changed, we performed this bioinformatics analysis to compare these properties of SARS-CoV-2 in the sequences published by China, USA, Czech, Korea, Greece, Australia, Thailand, Turkey, Sweden, and India.

According to the results, the 5th, 115th, 197th, 221st, 247th, 261st, 271st, 323rd, 367th, 453rd, 501st, 614th, 631st, 771st, 797th, 829th, 930th, 939th, 1181st, 1248th, and 1250th amino acids have changed in the sequences published by USA, Czech, Korea, Greece, Australia, Thailand, Turkey, Sweden, and India. However, the similarity of the sequence was still more than 99%, indicating that the sequence did not change greatly as the virus spread. After analyzing the secondary structures of SARS-CoV-2 S protein, we found that the secondary structures had changed as the amino acid mutated, creating an additional turn region in the mutation area of the sequence in Korea, while a coil region and a turn region disappeared in

the mutation amino acids of Sweden and India, respectively. The turn and coil are mostly located on the surface of the protein, and structures are prominent. There are possible epitope regions in the turn and coil. Therefore, the antigenicity of SARS-CoV-2 in Korea may be enhanced, while it may be weakened in Sweden and India. We constructed a 3D model of SARS-CoV-2 S protein to show the location of mutated amino acids and found that almost all of the mutated amino acids were on the outside of the SARS-CoV-2 S protein. This suggests that the exterior of the protein structure is exposed to the external environment and seems to be prone to mutation. In terms of epitope analysis, B cell epitopes analysis suggested that antigenicity of SARS-CoV-2 from Greece, Australia, Sweden and India (MT050493.1) were weakened while the antigenicity of SARS-CoV-2 in Korea, USA (MT370991.1, MT457401.1, MT457395.1, MT444634.1), India (MT396242.1) may be enhanced. Australia, Sweden, and India (MT050493.1) may be weakened. These results of B cell epitope analysis were almost consistent with the results of T cell epitope analysis. Interestingly, the antigenicity of SARS-CoV-2 from USA were enhanced after mutation. Previous research indicated that USA may be the birthplace of SARS-CoV-2 (Forster et al., 2020) (Appendix figure 1), suggesting SARS-CoV-2 of China may come from the United States. Therefore, the mutations may lead to a decrease in the antigenicity of the virus in the process of spreading. The antigen mutations bring difficulties to herd immunity, and there may be double virus infections, which may bring opportunities for resistance to mutant viruses.

Conclusion

In this study, we found that the sequence of SARS-CoV-2 S protein has changed as it spread to different countries, and as such, the structures and epitopes changed accordingly. In general, we found that the mutation leads to a decrease in the antigenicity of SARS-CoV-2 in the process of spreading. This may be a mechanism for the virus to evade surveillance by the immune system, thereby protecting itself, or it may be a new evolutionary direction. In this case, the prevalence of the virus may become more widespread. Therefore, more research concerning the structure and properties, clinical treatment, control strategy, and a potential vaccine for SARS-CoV-2 is urgently needed.

Abbreviations

COVID-19: Coronavirus Disease 2019

SARS-CoV-2: Severe Acute Respiratory Syndrome Coronavirus 2

S protein: spike protein

M protein: membrane Protein

E protein: envelope protein

N protein: nucleocapsid protein

PDB: Protein Data Bank

VMD: Visual Molecular Dynamics

IEDB: Immune Epitope Database

Declarations

Ethics approval and consent to participate

Not applicable.

Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author upon reasonable request.

Competing interests

The authors declare that they do not have any competing interests.

Funding

This work was supported by Mittal Innovation Project of Central South University (Grant No. GCX20190879Y) and the Fundamental Research Funds for the Central Universities of Central South University (Grant No. 2019zzts816 and 2018zzts930). The study funders/sponsors had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Author contributions

Concept and design: GL and JZ. Acquisition, analysis, or interpretation of data: JZ, LW and GL. Drafting of the manuscript: JZ and JS. Critical revision of the manuscript for important intellectual content: GL. Statistical analysis: JZ.

Acknowledgments

The authors would like to thank all of the co-investigators and colleagues who made this study possible. We thank LetPub (www.letpub.com) for its linguistic assistance during the preparation of this

References

1. Wang D, Hu B, Hu C et al. Clinical Characteristics of 138 Hospitalized Patients With 2019 Novel Coronavirus-Infected Pneumonia in Wuhan, China. *JAMA*. 2020
2. Chang, Lin M, Wei L et al. Epidemiologic and Clinical Characteristics of Novel Coronavirus Infections Involving 13 Patients Outside Wuhan, China. *JAMA*. 2020
3. Li Q, Guan X, Wu P et al. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia. *N Engl J Med*. 2020
4. Wan S, Xiang Y, Fang W et al. Clinical Features and Treatment of COVID-19 Patients in Northeast Chongqing. *J Med Virol*. 2020
5. Huang C, Wang Y, Li X et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet*. 2020
6. Chen N, Zhou M, Dong X et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. *Lancet*. 2020
7. Chan JF, Yuan S, Kok KH et al. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. *Lancet*. 2020
8. Zhu N, Zhang D, Wang W et al. A Novel Coronavirus from Patients with Pneumonia in China, 2019. *N Engl J Med*. 2020
9. Holshue ML, DeBolt C, Lindquist S et al. First Case of 2019 Novel Coronavirus in the United States. *N Engl J Med*. 2020
10. Hulswit RJ, de Haan CA, Bosch BJ. Coronavirus Spike Protein and Tropism Changes. *Adv Virus Res*. 2016 96:29-57
11. Xu X, Chen P, Wang J et al. Evolution of the novel coronavirus from the ongoing Wuhan outbreak and modeling of its spike protein for risk of human transmission. *Sci China Life Sci*. 2020 63(3):457-460
12. Hoffmann M, Kleine-Weber H, Schroeder S et al. SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell*. 2020
13. Yan R, Zhang Y, Li Y et al. Structural basis for the recognition of the SARS-CoV-2 by full-length human ACE2. *Science*. 2020
14. Sah R, Rodriguez-Morales AJ, Jha R et al. Complete Genome Sequence of a 2019 Novel Coronavirus (SARS-CoV-2) Strain Isolated in Nepal. *Microbiol Resour Announc*. 2020 9(11)
15. Guex N, Peitsch MC. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis*. 1997 18(15):2714-23
16. Guex N, Peitsch MC, Schwede T. Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: a historical perspective. *Electrophoresis*. 2009 30 Suppl 1:S162-73

17. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. *J Mol Graph.* 1996 14(1):33-8, 27-8
18. Schwede T, Kopp J, Guex N et al. SWISS-MODEL: An automated protein homology-modeling server. *Nucleic Acids Res.* 2003 31(13):3381-5
19. Van Regenmortel MH. What is a B-cell epitope? *Methods Mol Biol.* 2009 524:3-20
20. Kyte J, Doolittle RF. A simple method for displaying the hydropathic character of a protein. *J Mol Biol.* 1982 157(1):105-32
21. Carter JM, Loomis-Price L. B cell epitope mapping using synthetic peptides. *Curr Protoc Immunol.* 2004 Chapter 9:Unit 9.4
22. Tong JC, Tammi MT. Prediction of protein allergenicity using local description of amino acid sequence. *Front Biosci.* 2008 13:6072-8
23. Welling GW, Weijer WJ, van der Zee R et al. Prediction of sequential antigenic regions in proteins. *Febs Lett.* 1985 188(2):215-8
24. Gershoni JM, Stern B, Denisova G. Combinatorial libraries, epitope structure and the prediction of protein conformations. *Immunol Today.* 1997 18(3):108-10
25. Subramani A, Floudas CA. Structure prediction of loops with fixed and flexible stems. *J Phys Chem B.* 2012 116(23):6670-82
26. Wrapp D, Wang N, Corbett KS et al. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science.* 2020 367(6483):1260-1263
27. Tian X, Li C, Huang A et al. Potent binding of 2019 novel coronavirus spike protein by a SARS coronavirus-specific human monoclonal antibody. *Emerg Microbes Infect.* 2020 9(1):382-385

Figures

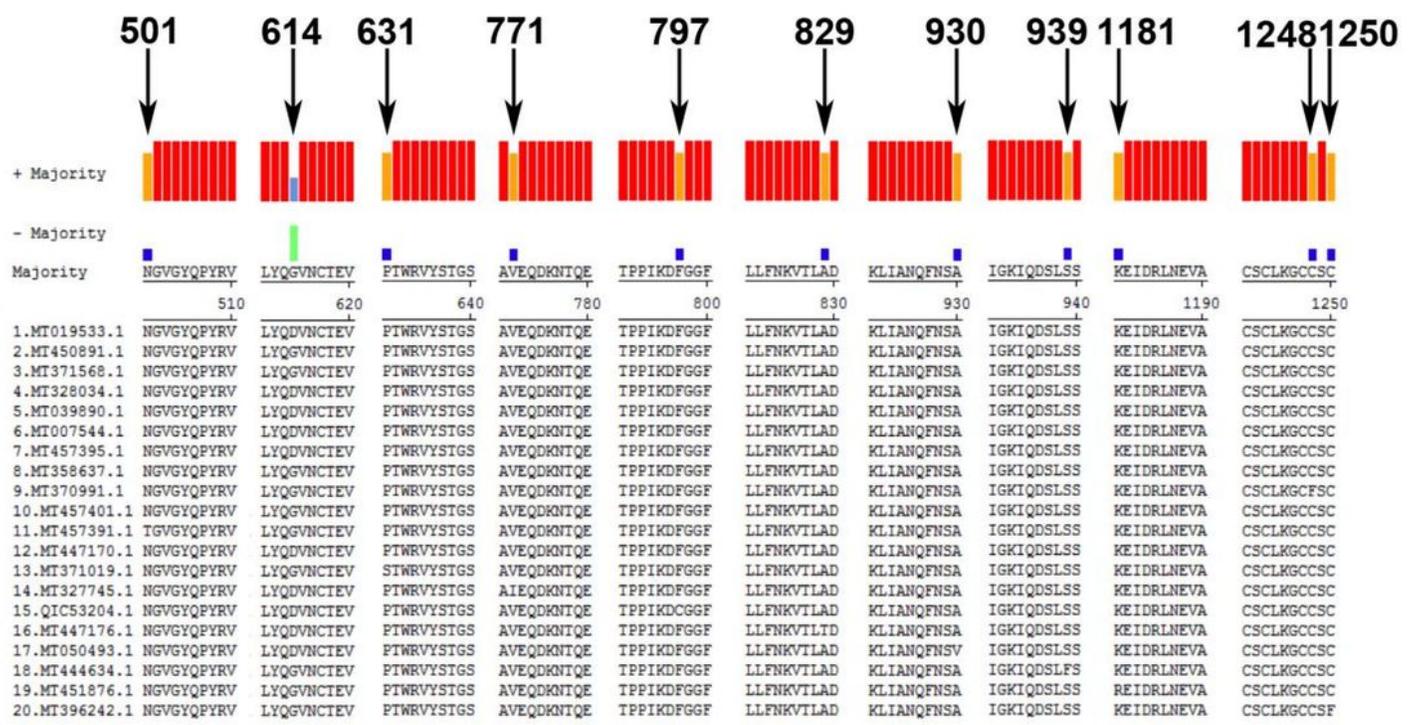
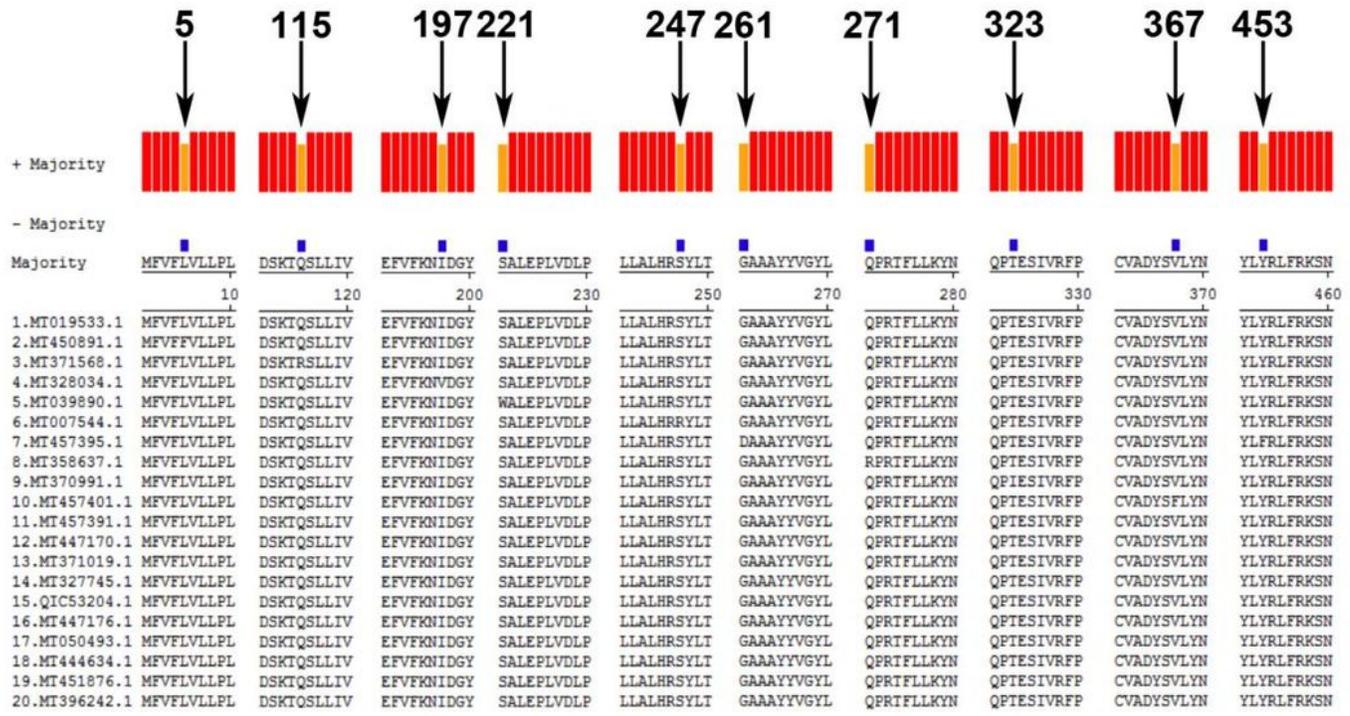


Figure 1

Comparison of amino acid sequences of SARS-CoV-2 S protein. Yellow bars indicate differential amino acid positions. No. 1: China, (No. 2, No. 7, No. 9, No. 10, No. 11, No. 13 and No. 18): American, No. 3: Czech, No. 4: Greece, No. 5: Korea, No. 6: Australia, (No. 8, No. 17, No. 19 and No. 20): India, No. 12 and No. 16: Thailand, No. 14: Turkey, No. 15: Sweden.

A

	C	S	T	P	A	G	N	D	E	Q	H	R	K	M	I	L	V	F	Y	W	
C	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0
S	0	0	2	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	1	0
T	0	0	0	2	0	2	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
P	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
G	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0
N	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
D	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
E	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Q	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0
H	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
R	0	1	0	0	0	0	0	2	0	2	0	0	0	0	0	0	0	0	0	0	0
K	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0
M	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
I	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0
L	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
V	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
F	4	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0
Y	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
W	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

B

		Percent identity																				
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
Divergence	1	0	99.8	99.8	99.9	99.9	99.9	99.8	99.8	99.8	99.8	99.8	100.0	99.8	99.9	99.9	99.9	99.9	99.8	99.8	99.8	1
	2	0.2	0	99.8	99.8	99.8	99.8	99.7	99.8	99.8	99.8	99.8	99.8	99.8	99.8	99.8	99.8	99.8	99.8	99.8	99.8	2
	3	0.2	0.2	0	99.8	99.8	99.8	99.7	99.8	99.8	99.8	99.8	99.8	99.8	99.8	99.8	99.8	99.8	99.8	99.8	99.8	3
	4	0.1	0.2	0.2	0	99.8	99.8	99.8	99.8	99.7	99.8	99.8	99.9	99.8	99.8	99.8	99.8	99.8	99.8	99.8	99.8	4
	5	0.1	0.2	0.2	0.2	0	99.8	99.8	99.8	99.7	99.8	99.8	99.9	99.8	99.8	99.8	99.8	99.8	99.8	99.8	99.8	5
	6	0.1	0.2	0.2	0.2	0.2	0	99.8	99.8	99.7	99.8	99.8	99.9	99.8	99.8	99.8	99.8	99.8	99.8	99.8	99.8	6
	7	0.2	0.3	0.3	0.2	0.2	0.2	0	99.7	99.6	99.7	99.7	99.8	99.7	99.8	99.8	99.8	99.7	99.7	99.7	99.7	7
	8	0.2	0.2	0.2	0.2	0.2	0.2	0.3	0	99.8	99.8	99.8	99.8	99.8	99.8	99.8	99.8	99.8	99.8	99.8	99.8	8
	9	0.2	0.2	0.2	0.3	0.3	0.3	0.4	0.2	0	99.8	99.8	99.8	99.8	99.7	99.7	99.7	99.7	99.7	99.7	99.7	9
	10	0.2	0.2	0.2	0.2	0.2	0.2	0.3	0.2	0.2	0	99.8	99.8	99.8	99.8	99.8	99.8	99.8	99.8	99.8	99.8	10
	11	0.2	0.2	0.2	0.2	0.2	0.2	0.3	0.2	0.2	0.2	0	99.8	99.8	99.8	99.8	99.8	99.8	99.8	99.8	99.8	11
	12	0.0	0.2	0.2	0.1	0.1	0.1	0.2	0.2	0.2	0.2	0.2	0	99.8	99.9	99.9	99.9	99.9	99.8	99.8	99.8	12
	13	0.2	0.2	0.2	0.2	0.2	0.2	0.3	0.2	0.2	0.2	0.2	0.2	0	99.8	99.8	99.8	99.8	99.8	99.8	99.8	13
	14	0.1	0.2	0.2	0.2	0.2	0.2	0.2	0.3	0.2	0.2	0.1	0.2	0	99.8	99.8	99.8	99.8	99.8	99.8	99.8	14
	15	0.1	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.3	0.2	0.2	0.1	0.2	0.2	0	99.8	99.8	99.8	99.8	99.8	15
	16	0.1	0.2	0.2	0.2	0.2	0.2	0.2	0.3	0.2	0.2	0.1	0.2	0.2	0.2	0.2	0	99.8	99.8	99.8	99.8	16
	17	0.1	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.3	0.2	0.2	0.1	0.2	0.2	0.2	0.2	0	99.8	99.8	99.8	17
	18	0.2	0.2	0.2	0.2	0.2	0.2	0.3	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0	99.8	99.8	18
	19	0.2	0.2	0.2	0.2	0.2	0.2	0.3	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0	99.8	19
	20	0.2	0.2	0.2	0.2	0.2	0.2	0.3	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0	20

Figure 2

Similarity of amino acid sequences of SARS-CoV-2 S protein. No. 1: China, (No. 2, No. 7, No. 9, No. 10, No. 11, No. 13 and No. 18): American, No. 3: Czech, No. 4: Greece, No. 5: Korea, No. 6: Australia, (No. 8, No. 17, No. 19 and No. 20): India, No. 12 and No. 16: Thailand, No. 14: Turkey, No. 15: Sweden.

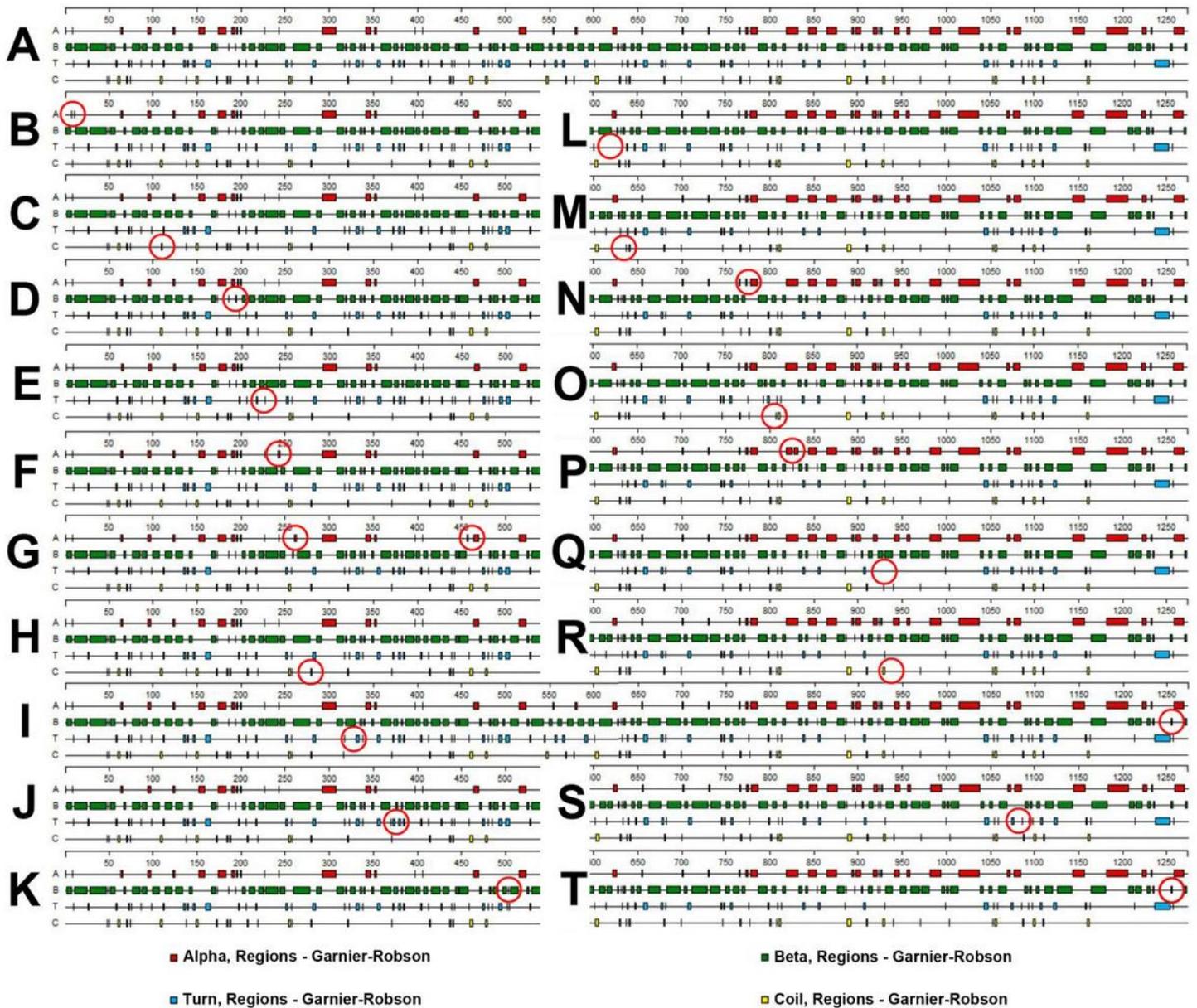


Figure 3

Comparison of secondary structures of SARS-CoV-2 S protein. Changed regions are marked in red circles. A: China, (B, G, I, J, K, M and R): American, C: Czech, D: Greece, E: Korea, F: Australia, (H, Q, S and T): India, L and P: Thailand, N: Turkey, O: Sweden.

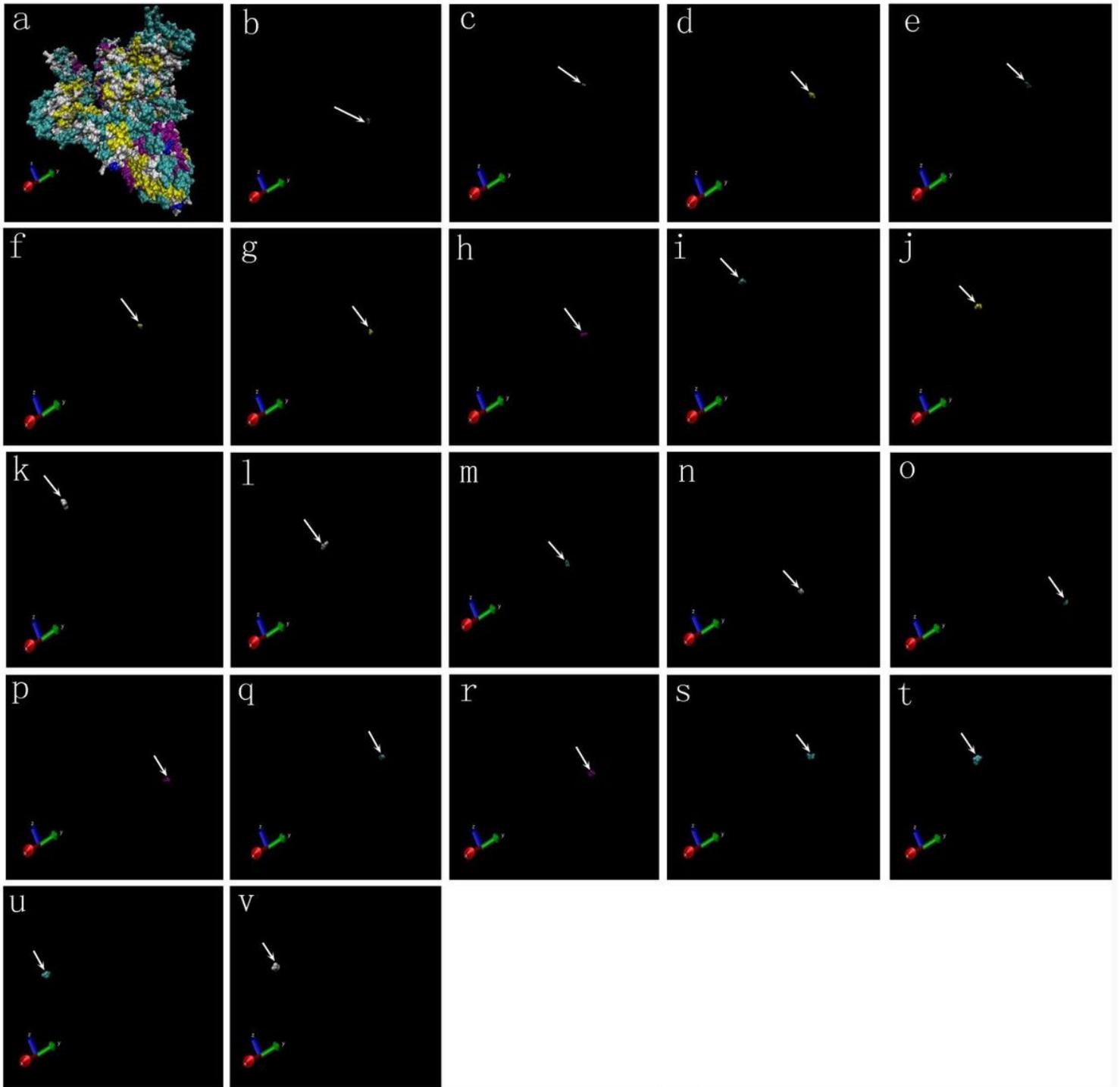


Figure 4

Comparison of 3D models of SARS-CoV-2 S protein. White: coil; cyan: turn; tan: bridge beta; yellow: extended beta; purple: alpha helix; blue: 3-10 helix.

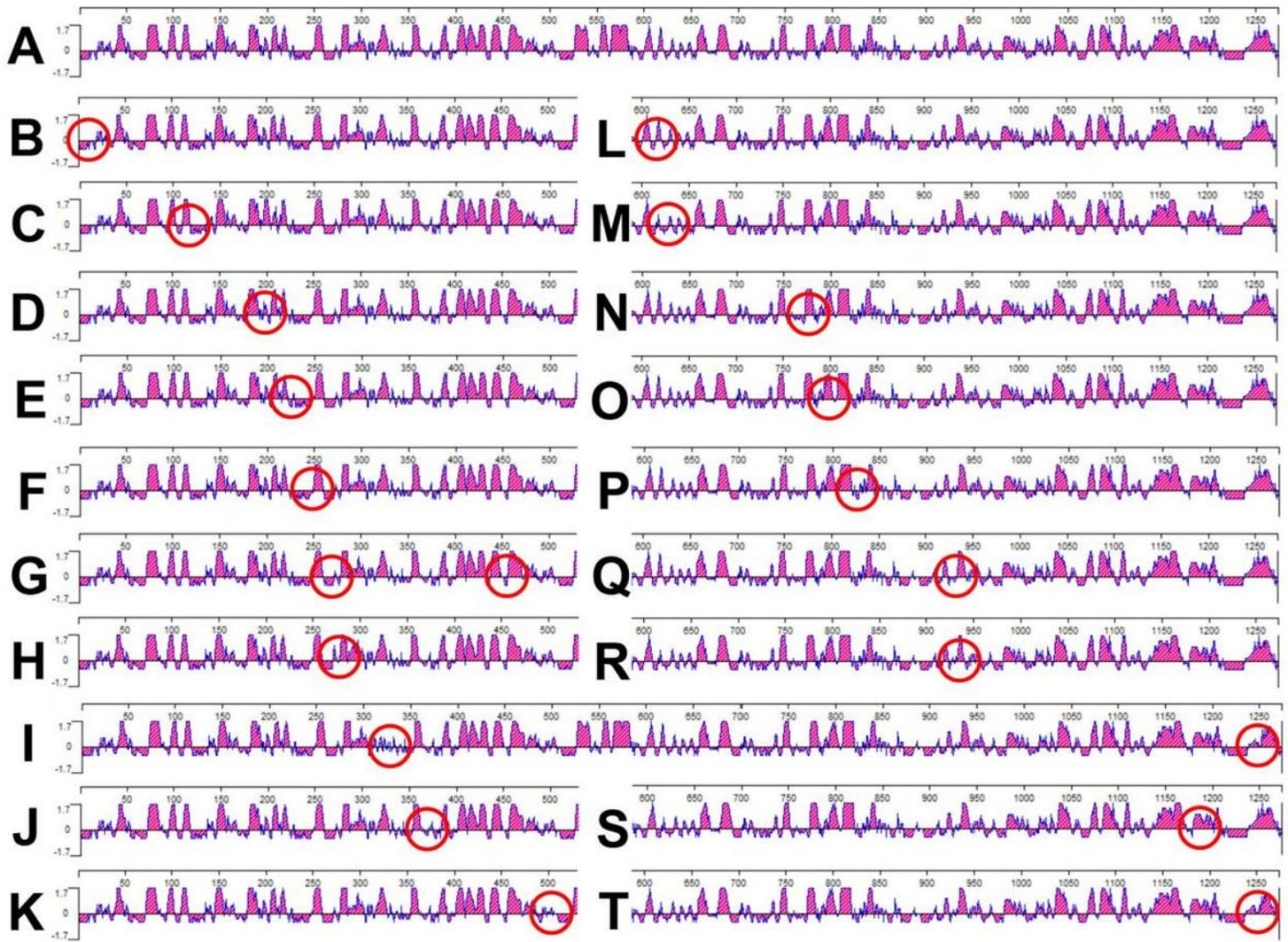


Figure 5

Comparison of B cell epitopes of SARS-CoV-2 S protein. Corresponding B cell epitopes of differential amino acid regions are marked in red circles. A: China, (B, G, I, J, K, M and R): American, C: Czech, D: Greece, E: Korea, F: Australia, (H, Q, S and T): India, L and P: Thailand, N: Turkey, O: Sweden.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Appendixfigureandtable.doc](#)