

Exploring the genomic and proteomic variations of SARS-CoV-2 spike glycoprotein: a computational biology approach

Syed Mohammad Lokman ^a, Md. Rasheduzzaman ^a, Asma Salauddin ^a, Rocktim Barua ^a,
Afsana Yeasmin Tanzina ^a, Meheadi Hasan Rumi ^a, Md. Imran Hossain ^a, Amam Zonaed
Siddiki ^b, Adnan Mannan ^{a *}, Md. Mahbub Hasan ^a

^a Department of Genetic Engineering & Biotechnology, Faculty of Biological Sciences,
University of Chittagong, Chattogram-4331, Bangladesh

^b Department of Pathology and Parasitology, Chittagong Veterinary and Animal Sciences
University, Chattogram-4202, Bangladesh

Corresponding author: Department of Genetic Engineering & Biotechnology, Faculty of
Biological Sciences, University of Chittagong, Chattogram-4331, Bangladesh

E-mail address: adnan.mannan@cu.ac.bd

Abstract

The newly identified SARS-CoV-2 has now been reported from around 183 countries with more than a million confirmed human cases including more than 68000 deaths. The genomes of SARS-CoV-2 strains isolated from different parts of the world are now available and the unique features of constituent genes and proteins have gotten substantial attention recently. Spike glycoprotein is widely considered as a possible target to be explored because of its role during the entry of coronaviruses into host cells. We analyzed 320 whole-genome sequences and 320 spike protein sequences of SARS-CoV-2 using multiple sequence alignment tools. In this study, 483 unique variations have been identified among the genomes including 25 non-synonymous mutations and one deletion in the spike protein of SARS-CoV-2. Among the 26 variations detected, 12 variations were located at the N-terminal domain and 6 variations at the receptor-binding domain (RBD) which might alter the interaction with receptor molecules. In addition, 22 amino acid insertions were identified in the spike protein of SARS-CoV-2 in comparison with that of SARS-CoV. Phylogenetic analyses of spike protein revealed that Bat coronavirus have a close evolutionary relationship with circulating SARS-CoV-2. The genetic variation analysis data presented in this study can help a better understanding of SARS-CoV-2 pathogenesis. Based on our findings, potential inhibitors can be designed and tested targeting these proposed sites of variation.

Keywords: SARS-CoV-2, Spike protein, Sequence analysis, COVID-19, Genomic variants.

1. Introduction

Coronavirus disease (COVID-19) is a pandemic manifesting respiratory illness and first reported in Wuhan, Hubei province of China in December 2019. The death toll rose to more than 68,000 among 1,250,000 confirmed cases around the Globe (until April 4, 2020) [1]. The virus causing COVID-19 is named as severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). Based on the phylogenetic studies, the SARS-CoV-2 is categorized as a member of the genus Betacoronavirus, the same lineage that includes SARS coronavirus (SARS-CoV) [2] that caused SARS (Severe Acute Respiratory Syndrome) in China during 2002 [3]. Recent studies showed that SARS-CoV-2 has a close relationship with bat SARS-like CoVs [4,5], though the intermediate hosts for zoonotic transmission of SARS-CoV-2 from bats to humans remain undiscovered.

SARS-CoV-2 has been identified as an enveloped, single-stranded positive-sense RNA virus with a genome size of approximately 29.9 kb encoding 27 proteins from 14 ORFs including 15 non-structural, 8 accessory and 4 major structural proteins. Two-thirds of the viral RNA harbours the first ORF (ORF1ab) dedicated for translating polyprotein 1a (pp1a) and polyprotein 1ab (pp1ab), which later undergo proteolytic cleavage to form 15 non-structural proteins. Spike glycoprotein (S), membrane (M), envelope (E) and nucleocapsid (N) are the four major structural proteins of SARS-CoV-2 [[6],[7]]. Interestingly, S glycoprotein is characterized as the critical determinant for viral entry into host cells which consists of two functional subunits namely S1 and S2. The S1 subunit recognizes and binds to the host receptor through the receptor-binding domain (RBD) whereas S2 is responsible for fusion with the host cell membrane [[8],[9],[10]]. MERS-CoV uses dipeptidyl peptidase-4 (DPP4) as entry receptor [11] whereas SARS-CoV and SARS-CoV-2 utilize ACE-2 (angiotensin converting enzyme-2) [12], abundantly available in lung alveolar epithelial cells and enterocytes, suggesting S glycoprotein as a potential drug target to halt the entry of SARS-CoV-2 [13].

According to recent reports, neutralizing antibodies are generated in response to the entry and fusion of surface-exposed S protein (mainly RBD domain) which is predicted to be an

important target for vaccine candidates [[10],[14],[15]]. However, SARS-CoV-2 has emerged with remarkable properties like glutamine-rich 42 aa long exclusive molecular signature (DSQQTVGQQDGSSEDNQTITTIQTIVEVQPQLEMELTPVVQTIE) in position 983-1024 of polyprotein 1ab (pp1ab) [16], diversified receptor-binding domain (RBD), unique furin cleavage site (PRRAR↓SV) at S1/S2 boundary in S glycoprotein which could play roles in viral pathogenesis, diagnosis and treatment [17]. To date, few genomic variations of SARS-CoV-2 are reported [[18],[19]]. There is growing evidence that spike protein, a 1273 amino acid long glycoprotein having multiple domains, possibly plays a major role in SARS-CoV-2 pathogenesis. Viral entry to the host cell is initiated by the receptor-binding domain (RBD) of S1 head. Upon receptor-binding, proteolytic cleavage occurs at S1/S2 cleavage site and two heptad repeats (HR) of S2 stalk form a six-helix bundle structure triggering the release of the fusion peptide. As it comes into close proximity to the transmembrane anchor (TM), the TM domain facilitates membrane destabilization required for fusion between virus-host membranes [[20],[21]]. Insights into the sequence variations of S glycoprotein among available genomes are key to understanding the biology of SARS-CoV-2 infection, developing antiviral treatments and vaccines. In this study, we have analyzed 320 genomic sequences of SARS-CoV-2 to identify mutations between the available genomes followed by the amino acid variations in the glycoprotein S to foresee their impact on the viral entry to host cell from structural biology viewpoint.

2. Methods and Materials

2.1 Dataset

All available sequences (320 whole genome and surface glycoprotein sequences of SARS-CoV-2) related to the COVID-19 pandemic were retrieved from NCBI Virus Variation Resource repository (<https://www.ncbi.nlm.nih.gov/labs/virus/>) [22]. In addition, all 40 S glycoprotein sequences from different coronavirus families were retrieved for phylogenetic

analysis. The NCBI reference sequence of SARS-CoV-2 S glycoprotein, accession number YP_009724390 was used as the canonical sequence for the analyses of spike protein variants.

2.2 Phylogenetic analysis

Variant analyses of SARS-CoV-2 genomes were performed in the Genome Detective Coronavirus Typing Tool Version 1.13 which is specially designed for this virus (<https://www.genomedetective.com/app/typingtool/cov/>) [23]. For multiple sequence alignment (MSA), Genome Detective Coronavirus Typing Tool uses a reference dataset of 431 whole genome sequences (WGS) where 386 WGS were from known nine coronavirus species. The dataset was then aligned with MUSCLE [24]. Entropy ($H(x)$) plot of nucleotide variations in SARS-CoV-2 genome was constructed using BioEdit [25]. MEGA X (version 10.1.7) was used to construct the MSAs and the phylogenetic tree using pairwise alignment and neighbor-joining methods in ClustalW [26,27]. Tree structure was validated by running the analysis on 1000 bootstraps [28] replications dataset and the evolutionary distances were calculated using the Poisson correction method [29].

2.3 Homology modeling of S glycoprotein

Variant sequences of SARS-CoV-2 were modeled in Swiss-Model [30] using the Cryo-EM spike protein structure of SARS-CoV-2 (PDB ID 6VSB) as a template. The overall quality of models was assessed in RAMPAGE server [31] by generating Ramachandran plots (Supplementary Table 1). PyMol and BIOVIA Discovery Studio were used for structure visualization and superpose [32,33].

3. Results

3.1 Genomic variations of SARS-CoV-2

Multiple sequence alignment of the available 320 genomes of SARS-CoV-2 were performed and 483 variations were found throughout the 29,903 bp long SARS-CoV-2 genome with in total 115 variations in UTR region, 130 synonymous variations that cause no amino acid alteration, 228 non-synonymous variations causing change in amino acid residue, 16 INDELs, and 2 variations in non-coding region (Supplementary Table 2). Among the 483 variations, 40 variations (14 synonymous, 25 non-synonymous mutations and one deletion) were observed in the region of ORF S that encodes S glycoprotein which is responsible for viral fusion and entry into the host cell [34]. Notable that, most of the SARS-CoV-2 genome sequences were deposited from the USA (250) and China (50) (Supplementary Fig. 1). Positional variability of the SARS-CoV-2 genome was calculated from the MSA of 320 SARS-CoV-2 whole genomes as a measure of Entropy value ($H(x)$) [35]. Excluding 5' and 3' UTR, ten hotspot of hypervariable position were identified, of which seven were located at ORF1ab (1059C>T, 3037C>T, 8782C>T, 14408C>T, 17747C>T, 17858A>G, 18060C>T) and one at ORF S (23403A>G), ORF3a (25563G>T), and ORF8 (28144T>C) respectively. The variability at position 8782 and 28144 were found to be the highest among the other hotspots (Fig. 1).

3.2 Phylogenetic Analysis of S glycoprotein

The phylogenetic analysis of a total of 66 sequences (26 unique SARS-CoV-2 and 40 different coronavirus S glycoprotein sequences) was performed. The evolutionary distances showed that all the SARS-CoV-2 spike proteins cluster in the same node of the phylogenetic tree confirming the sequences are similar to Refseq YP_009724390 (Fig. 2). Bat coronaviruses has a close evolutionary relationship as different strains were found in the nearest outgroups and clades (Bat coronavirus BM48-31, Bat hp-beta coronavirus, Bat coronavirus HKU9) conferring that

coronavirus has vast geographical spread and bat is the most prevalent host (Fig. 2). In other clades, the clusters were speculated through different hosts which may describe the evolutionary changes of surface glycoprotein due to cross species transmission. Viral hosts reported from different spots at different times is indicative of possible recombination.

3.3 SARS-CoV-2 Spike Protein Variation Analysis

The S glycoprotein sequences of SARS-CoV-2 were retrieved from the NCBI Virus Variation Resource repository and aligned using ClustalW. The position of SARS-CoV-2 spike protein domains was measured by aligning with the SARS-CoV spike protein (Fig. 3) [36,37]. From the sequence identity matrix, 26 unique variants including the canonical sequence of SARS-CoV-2 spike glycoprotein (YP_009724390) were identified to have 25 substitution and a deletion (Fig. 4A and Supplementary Table 3). 215 sequences were found identical with SARS-CoV-2 S protein reference sequence (YP_009724390) while 64 sequences were identical with the same variation of D614G (Supplementary Table 4, 5). Among all the variations, twelve (Y28N, T29I, H49Y, L54F, N74K, E96D, D111N, Y145Del, F157L, G181V, S221W, and S247R) were located at N-terminal domain (NTD). Another 6 variations (A348T, R408I, G476S, V483A, H519Q, A520S) were found at the receptor-binding domain (RBD) while only two variations (A930V, and D936Y) were found at heptad repeat 1 (HR1) domain. Single variations were found in signal peptide (L5F) domain, sub-domain-2 (D614G), sub-domain-3 (A1078V), heptad repeat 2 domain (D1168H), and cytoplasmic tail domain (D1259H) each. Notable that substitution of Cysteine by Phenylalanine was observed at 19 amino acids upstream of the fusion peptide domain (Fig. 4A). The mutation of Aspartic acid to Glycine at position 614 was observed 71 times with entropy value over 0.5 among the available 320 SARS-CoV-2 spike protein sequences (Supplementary Fig. 2).

Alterations of amino acid residual charge from positive to neutral (H49Y, R408I, H519Q), negative to neutral (D111N, D614G, D936Y), negative to positive (D1168H, D1259H), and neutral to positive (N74K, S247R) were seen in variants QHW06059, QHS34546, QIS61422, QIS61338, QIK50427, QIS30615, QIS60978, QIS60582, QIO04367, and QHR84449 respectively due to substitution of amino acid that differ in charge. The remaining 15 variants were mutated with the amino acids that are similar in charge (Fig. 4 A). The SARS-CoV-2 spike protein variants were superposed with the cryo-electron microscopic structure of SARS-CoV-2 spike protein [8]. L5F, N74K, E96D, F157L, G181V, S247R, G476S, V483A, D1168H, and D1259H variants were excluded from superposition due to absence of respective residues in the 3D structure of template (PDB: 6VSB). The superposition showed that most of the residual change were causing incorporation of bulky amino acid residues (T29I, H49Y, L54F, S221W, A348T, H519Q, A520S, A930V, D936Y, and A1078V) in place of smaller size residue except Y28N, D111N, R408I, D614G, and F797C (Fig. 4 B-P).

The sequence comparison of spike glycoprotein between SARS-CoV-2 variants and SARS-CoV (Uniprot ID: P59594) revealed nearly 77.46% similarity and identified the presence of additional 22 amino acids in SARS-CoV-2 spike protein variants resulting from a total of 5 insertions (Supplementary Fig. 3). Among these, the major insertion consisting of 7 amino acids (GTNGTKR) at position 72-78 followed by 4 amino acids (NKSW) at position 149-152 and 6 amino acids (SYLTPG) at 247-252 occurred in N-terminal domain. Insertion of Glycine at 482 was found in receptor binding domain, preceding another insertion of 4 amino acids (NSPR) at position 679-682, just upstream of S1/S2 cleavage site that leads to form a furin-like cleavage site (PRRARS) in the S protein variants of SARS-CoV-2 (Supplementary Fig. 3). The S2 subunit of spike protein, especially the heptad repeat region 2, fusion peptide domain, transmembrane domain, and cytoplasmic tail were found to be highly conserved in the SARS-CoV and the SARS-CoV-2 variants while the S1 subunit was more diverse, specifically the N-terminal domain (NTD) and receptor-binding domain (RBD).

4. Discussion

COVID 19 is one of the most contagious pandemics the world has ever had with 1,250,000 confirmed cases to date (April 4, 2020) and the cases have increased as high as 5 times in less than a month [1]. Phylogenetic analysis showed that the SARS-CoV-2 is a unique coronavirus presumably related to Bat coronavirus (BM48-31, Hp-betacoronavirus). During this study, we investigated the available genomes of SARS-CoV-2 and found variations in 483 positions resulting in 130 synonymous and 228 non-synonymous variants. Out of them, 25 non-synonymous variants were observed in the spike protein of SARS-CoV-2. Viral spike protein is thought to have a crucial role in drug and vaccine development as reported previously in managing the viruses like SARS-CoV and MERS-CoV [[15],[38],[39],[40]]. Likewise, a number of studies targeting SARS-CoV-2 spike protein have been undertaken for the therapeutic measures [41], but the unique structural and functional details of SARS-CoV-2 spike protein are still under scrutiny. We also found a variant (R408I) at receptor binding domain (RBD) that mutated from positively charged Arginine residue to neutral and smaller sized Isoleucine residue (Fig. 4 I). This change might alter the interaction of viral RBD with the host receptor because the R408 residue of SARS-CoV-2 is known to interact with the ACE2 receptor for viral entry [42]. Similarly, alterations of RBD (G476S, V483A, H519Q, and A520S) also could affect the interaction of SARS-CoV-2 spike protein with other molecules which require further investigations. QIA98583 and QIS30615 variants were found to have an alteration of Alanine to Valine (A930V), and Aspartic acid to Tyrosine (D936Y) respectively in the alpha helix of the HR1 domain. Previous reports have indicated that HR1 domain plays a significant role in viral fusion and entry by forming helical bundles with HR2, and mutations including alanine substitution by valine (A1168V) in HR1 region are predominantly responsible for conferring resistance to mouse hepatitis coronaviruses against HR2 derived peptide entry inhibitors [43]. This study hypothesizes the mutation (A930V) found in that of SARS-CoV-2 might also have a role in the emergence of drug-resistance virus strains. Also,

the mutation (D1168H) found in the heptad repeat 2 (HR) SARS-CoV-2 could play a vital role in viral pathogenesis.

The SARS-CoV-2 S protein contains additional furin protease cleavage site, PRRARS, in S1/S2 domain which is conserved among all 320 sequences as revealed during this study (Supplementary Fig. 3). This unique signature is thought to make the SARS-CoV-2 more virulent than SARS-CoV and regarded as novel features of the viral pathogenesis (ref 11). According to previous reports the more the host cell protease can process the coronavirus S can accelerate viral tropism accordingly in influenza virus [[9],[44],[45],[46]]. Apart from that, this could also promote viruses to escape antiviral therapies targeting transmembrane protease TMPRSS2 (ClinicalTrials.gov, NCT04321096) which is well reported protease to cleave at S1/S2 of S glycoprotein [47]. Comparative analyses between SARS-CoV and SARS-CoV-2 spike glycoprotein showed 77% similarity between them where the most diverse region was the N-terminal domain and receptor binding domain. The insertion of 17 additional amino acid residues in the N-terminal domain of SARS-CoV-2 and its high sequence diversity suggests that it may have a role in binding with other cell receptors in humans. This is because the N-terminal domain could function as the receptor-binding domain of various coronaviruses [[34],[36]]. Similar phenomenon has been observed in mouse hepatitis coronavirus (MHV) and porcine transmissible gastroenteritis coronavirus (TGEV) where the N-terminal domain is reported to be attached with the host entry receptor [48,49]. The variation analyses in amino acids indicated the structural features of different domains of the SARS-CoV-2 spike proteins. However, to identify the actual role of involvement of S glycoprotein, a larger dataset regarding genomics and proteomics of SARS-CoV-2 is required as this protein is vital to understand the viral pathogenicity, evolution and development of therapeutics. Further analyses of all the S glycoprotein and SARS-CoV-2 genomes with different epidemiological aspects is warranted to get the pathogenesis of SARS-CoV-2.

References

- [1] WHO (World Health Organization), Coronavirus disease (COVID-2019) situation reports, 2020. <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports/> (accessed April 4, 2020).
- [2] A.E. Gorbalenya, Severe acute respiratory syndrome-related coronavirus--The species and its viruses, a statement of the Coronavirus Study Group, *BioRxiv*. (2020).
- [3] T.G. Ksiazek, D. Erdman, C.S. Goldsmith, S.R. Zaki, T. Peret, S. Emery, S. Tong, C. Urbani, J.A. Comer, W. Lim, others, A novel coronavirus associated with severe acute respiratory syndrome, *N. Engl. J. Med.* 348 (2003) 1953–1966.
- [4] W. Li, Z. Shi, M. Yu, W. Ren, C. Smith, J.H. Epstein, H. Wang, G. Crameri, Z. Hu, H. Zhang, others, Bats are natural reservoirs of SARS-like coronaviruses, *Science* (80-.). 310 (2005) 676–679.
- [5] P. Zhou, X.-L. Yang, X.-G. Wang, B. Hu, L. Zhang, W. Zhang, H.-R. Si, Y. Zhu, B. Li, C.-L. Huang, others, A pneumonia outbreak associated with a new coronavirus of probable bat origin, *Nature*. 579 (2020) 270–273.
- [6] F. Wu, S. Zhao, B. Yu, Y.-M. Chen, W. Wang, Z.-G. Song, Y. Hu, Z.-W. Tao, J.-H. Tian, Y.-Y. Pei, others, A new coronavirus associated with human respiratory disease in China, *Nature*. 579 (2020) 265–269.
- [7] A. Wu, Y. Peng, B. Huang, X. Ding, X. Wang, P. Niu, J. Meng, Z. Zhu, Z. Zhang, J. Wang, others, Genome composition and divergence of the novel coronavirus (2019-nCoV) originating in China, *Cell Host Microbe*. (2020).
- [8] D. Wrapp, N. Wang, K.S. Corbett, J.A. Goldsmith, C.-L. Hsieh, O. Abiona, B.S. Graham, J.S. McLellan, Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation, *Science* (80-.). 367 (2020) 1260–1263.

- [9] A.C. Walls, Y.-J. Park, M.A. Tortorici, A. Wall, A.T. McGuire, D. Velesler, Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein, *Cell*. (2020).
- [10] Y. Chen, Y. Guo, Y. Pan, Z.J. Zhao, Structure analysis of the receptor binding of 2019-nCoV, *Biochem. Biophys. Res. Commun.* (2020).
- [11] V.S. Raj, H. Mou, S.L. Smits, D.H.W. Dekkers, M.A. Müller, R. Dijkman, D. Muth, J.A.A. Demmers, A. Zaki, R.A.M. Fouchier, others, Dipeptidyl peptidase 4 is a functional receptor for the emerging human coronavirus-EMC, *Nature*. 495 (2013) 251–254.
- [12] W. Li, M.J. Moore, N. Vasilieva, J. Sui, S.K. Wong, M.A. Berne, M. Somasundaran, J.L. Sullivan, K. Luzuriaga, T.C. Greenough, others, Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus, *Nature*. 426 (2003) 450–454.
- [13] M. Letko, A. Marzi, V. Munster, Functional assessment of cell entry and receptor usage for SARS-CoV-2 and other lineage B betacoronaviruses, *Nat. Microbiol.* (2020) 1–8.
- [14] M. Hoffmann, H. Kleine-Weber, S. Schroeder, N. Krüger, T. Herrler, S. Erichsen, T.S. Schiergens, G. Herrler, N.-H. Wu, A. Nitsche, others, SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor, *Cell*. (2020).
- [15] X. Tian, C. Li, A. Huang, S. Xia, S. Lu, Z. Shi, L. Lu, S. Jiang, Z. Yang, Y. Wu, others, Potent binding of 2019 novel coronavirus spike protein by a SARS coronavirus-specific human monoclonal antibody, *Emerg. Microbes Infect.* 9 (2020) 382–385.
- [16] Y. Cárdenas-Conejo, A. Liñan-Rico, D.A. García-Rodríguez, S. Centeno-Leija, H. Serrano-Posada, An exclusive 42 amino acid signature in pp1ab protein provides

- p>insights into the evolutive history of the 2019 novel human-pathogenic coronavirus (SARS-CoV2),
- J. Med. Virol.*
- (2020).
- [17] B. Coutard, C. Valle, X. de Lamballerie, B. Canard, N.G. Seidah, E. Decroly, The spike glycoprotein of the new coronavirus 2019-nCoV contains a furin-like cleavage site absent in CoV of the same clade, *Antiviral Res.* 176 (2020) 104742.
- [18] C. Ceraolo, F.M. Giorgi, Genomic variance of the 2019-nCoV coronavirus, *J. Med. Virol.* (2020).
- [19] R. Lu, X. Zhao, J. Li, P. Niu, B. Yang, H. Wu, W. Wang, H. Song, B. Huang, N. Zhu, others, Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding, *Lancet.* 395 (2020) 565–574.
- [20] B.J. Bosch, R. van der Zee, C.A.M. de Haan, P.J.M. Rottier, The coronavirus spike protein is a class I virus fusion protein: structural and functional characterization of the fusion core complex, *J. Virol.* 77 (2003) 8801–8811.
- [21] S. Liu, G. Xiao, Y. Chen, Y. He, J. Niu, C.R. Escalante, H. Xiong, J. Farmer, A.K. Debnath, P. Tien, others, Interaction between heptad repeat 1 and 2 regions in spike protein of SARS-associated coronavirus: implications for virus fusogenic mechanism and identification of fusion inhibitors, *Lancet.* 363 (2004) 938–947.
- [22] E.L. Hatcher, S.A. Zhdanov, Y. Bao, O. Blinkova, E.P. Nawrocki, Y. Ostapchuk, A.A. Schäffer, J.R. Brister, Virus Variation Resource--improved response to emergent viral outbreaks, *Nucleic Acids Res.* 45 (2017) D482--D490.
- [23] S. Cleemput, W. Dumon, V. Fonseca, W.A. Karim, M. Giovanetti, L.C. Alcantara, K. Deforche, T. de Oliveira, others, Genome Detective Coronavirus Typing Tool for rapid identification and characterization of novel coronavirus genomes, (2020).
- [24] R.C. Edgar, MUSCLE: multiple sequence alignment with improved accuracy and

- speed, in: Proceedings. 2004 IEEE Comput. Syst. Bioinforma. Conf. 2004. CSB 2004., 2004: pp. 728–729.
- [25] T.A. Hall, others, BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT, in: Nucleic Acids Symp. Ser., 1999: pp. 95–98.
- [26] S. Kumar, G. Stecher, M. Li, C. Knyaz, K. Tamura, MEGA X: molecular evolutionary genetics analysis across computing platforms, Mol. Biol. Evol. 35 (2018) 1547–1549.
- [27] N. Saitou, M. Nei, The neighbor-joining method: a new method for reconstructing phylogenetic trees., Mol. Biol. Evol. 4 (1987) 406–425.
- [28] B. Efron, E. Halloran, S. Holmes, Bootstrap confidence levels for phylogenetic trees, Proc. Natl. Acad. Sci. 93 (1996) 13429.
- [29] E. Zuckerkandl, L. Pauling, Evolutionary divergence and convergence in proteins, in: Evol. Genes Proteins, Elsevier, 1965: pp. 97–166.
- [30] W. Andrew, B. Martino, B. Stefan, S. Gabriel, T. Gerardo, G. Rafal, Heer Florian T, de Beer Tjaart A P, Rempfer Christine, Bordoli Lorenza, Lepore Rosalba, Schwede Torsten. SWISS-MODEL: homology modelling of protein structures and complexes, Nucleic Acids Res. 46 (2018) W296--W303.
- [31] M.G. Prisant, J.S. Richardson, D.C. Richardson, Structure validation by Calpha geometry: Phi, psi and Cbeta deviation, Proteins. 50 (2003) 437–450.
- [32] W.L. DeLano, others, Pymol: An open-source molecular graphics tool, CCP4 Newsl. Protein Crystallogr. 40 (2002) 82–92.
- [33] D.S. BIOVIA, San Diego: Dassault Systèmes, (2015).
- [34] Receptor recognition mechanisms of coronaviruses: a decade of structural studies,

(n.d.).

- [35] E. Manaresi, I. Conti, G. Bua, F. Bonvicini, G. Gallinella, A Parvovirus B19 synthetic genome: Sequence features and functional competence, *Virology*. 508 (2017) 54–62.
- [36] Y. Yuan, D. Cao, Y. Zhang, J. Ma, J. Qi, Q. Wang, G. Lu, Y. Wu, J. Yan, Y. Shi, others, Cryo-EM structures of MERS-CoV and SARS-CoV spike glycoproteins reveal the dynamic receptor binding domains, *Nat. Commun.* 8 (2017) 15092.
- [37] M. Gui, W. Song, H. Zhou, J. Xu, S. Chen, Y. Xiang, X. Wang, Cryo-electron microscopy structures of the SARS-CoV spike glycoprotein reveal a prerequisite conformational state for receptor binding, *Cell Res.* 27 (2017) 119–129.
- [38] S.U. Kapadia, J.K. Rose, E. Lamirande, L. Vogel, K. Subbarao, A. Roberts, Long-term protection from SARS coronavirus infection conferred by a single immunization with an attenuated VSV-based vaccine, *Virology*. 340 (2005) 174–182.
- [39] H.A. Elshabrawy, M.M. Coughlin, S.C. Baker, B.S. Prabhakar, Human monoclonal antibodies against highly conserved HR1 and HR2 domains of the SARS-CoV spike protein are more broadly neutralizing, *PLoS One*. 7 (2012).
- [40] L. Du, Z. Kou, C. Ma, X. Tao, L. Wang, G. Zhao, Y. Chen, F. Yu, C.T.K. Tseng, Y. Zhou, S. Jiang, A truncated receptor-binding domain of MERS-CoV spike protein potently inhibits MERS-CoV infection and induces strong neutralizing antibody responses: Implication for developing therapeutics and vaccines, *PLoS One*. 8 (2013) 2–10. <https://doi.org/10.1371/journal.pone.0081587>.
- [41] S. Xia, Y. Zhu, M. Liu, Q. Lan, W. Xu, Y. Wu, T. Ying, S. Liu, Z. Shi, S. Jiang, others, Fusion mechanism of 2019-nCoV and fusion inhibitors targeting HR1 domain in spike protein, *Cell. Mol. Immunol.* (2020) 1–3.
- [42] J.T. Ortega, M.L. Serrano, F.H. Pujol, H.R. Rangel, Role of changes in SARS-CoV-2

- spike protein in the interaction with the human ACE2 receptor: An in silico analysis, EXCLI J. 19 (2020) 410.
- [43] B.J. Bosch, J.W.A. Rossen, W. Bartelink, S.J. Zuurveen, C.A.M. de Haan, S. Duquerroy, C.A.B. Boucher, P.J.M. Rottier, Coronavirus escape from heptad repeat 2 (HR2)-derived peptide entry inhibition as a result of mutations in the HR1 domain of the spike fusion protein, J. Virol. 82 (2008) 2580–2585.
- [44] H.-D. Klenk, W. Garten, Host cell proteases controlling virus pathogenicity, Trends Microbiol. 2 (1994) 39–43.
- [45] D.A. Steinhauer, Role of hemagglutinin cleavage for the pathogenicity of influenza virus, Virology. 258 (1999) 1–20.
- [46] J.K. Millet, G.R. Whittaker, Host cell proteases: Critical determinants of coronavirus tropism and pathogenesis, Virus Res. 202 (2015) 120–134.
- [47] A.R. Fehr, S. Perlman, Coronaviruses: an overview of their replication and pathogenesis, in: Coronaviruses, Springer, 2015: pp. 1–23.
- [48] R.K. Williams, G.-S. Jiang, K. V Holmes, Receptor for mouse hepatitis virus is a member of the carcinoembryonic antigen family of glycoproteins., Proc. Natl. Acad. Sci. 88 (1991) 5533–5536.
- [49] B. Delmas, J. Gelfi, R. L’Haridon, H. Sjöström, H. Laude, others, Aminopeptidase N is a major receptor for the enteropathogenic coronavirus TGEV, Nature. 357 (1992) 417–420.

Figure legends

Fig 1: Nucleotide sequence variation among 320 SARS-CoV-2 whole genomes. A.

Positional organization of major structural protein-encoding genes in orange color (S = Spike protein, E = Envelope protein, M = Membrane protein, N = Nucleocapsid protein) and accessory protein ORFs in blue colors. B. Variability within 320 SARS-CoV-2 genomic sequences represented by entropy ($H(x)$) value across genomic location. Two highest frequency of alterations were found at position 8785 of ORF1a and 28144 of ORF8. C. The respective alignment view of each highly variable regions.

Fig 2: Sequence phylogeny of SARS-CoV-2 spike glycoprotein variants and other coronavirus spike proteins based on amino acid sequences, retrieved from NCBI database using neighbor-joining methods in ClustalW and tree structure was validated by running the analysis on 1000 bootstraps. The branch length is indicated in the scale bar. The accession number YP_009724390 represents identical sequences out of SARS-CoV-2 spike proteins.

Fig. 3: Overall architecture of the SARS-CoV-2 S glycoprotein. A. Schematic diagram of the SARS-CoV S glycoprotein showing domain organization (Reconstructed from Y. Yuan et al., 2017 and M. Gui et al., 2017). B. Schematic domain organization diagram of the SARS-CoV-2 S glycoprotein constructed by aligning with SARS-CoV S protein domain. C. Homology model of SARS-CoV-2 S protein reference sequence YP_009724390 with PDB:6VSB. S protein trimer with two protomers surface shadowed (left). Ribbon diagram of SARS-CoV-2 S glycoprotein monomer from B. Here, NTD: N-terminal domain; RBD: receptor-binding domain; SD: subdomain; CR: connecting region; HR: heptad repeat; CH: central helix; BH: b-hairpin; FP: fusion peptide; TM: transmembrane domain; CT: cytoplasmic tail.

Fig 4: Variability within 320 SARS-CoV-2 S protein sequences. A. Schematic representations of mutations across the spike protein domain organization. Blue, red, and

black color represented charge of amino acid residue as positive, negative, and neutral respectively. B-N, Superposed structure of SARS-CoV-2 spike protein variants with the Cryo-EM structure of SARS-CoV-2 Spike Protein (PDB: 6VSB). Template residues are indicated by green color and variants' residues are indicated as red color. Here, B: Y28N, C: T29I, D: H49Y, E: L54F, F: D111N, G: S221W, H: A348T, I: R408I, J: H519Q, K: A520S, L: D614G, M: F797C, N: A930V, O: D936Y, and P: A1078V.

Fig. 1.

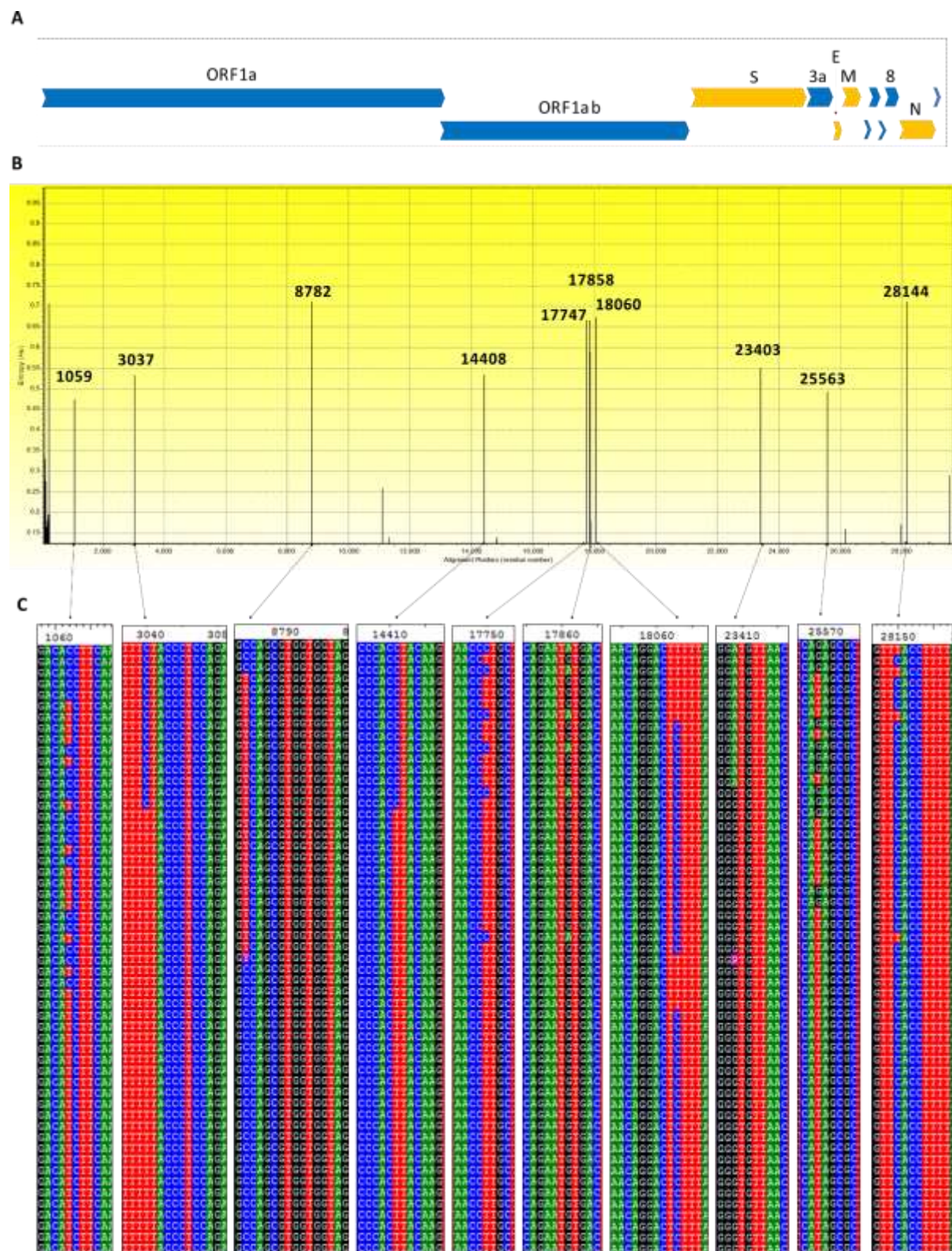


Fig. 2.

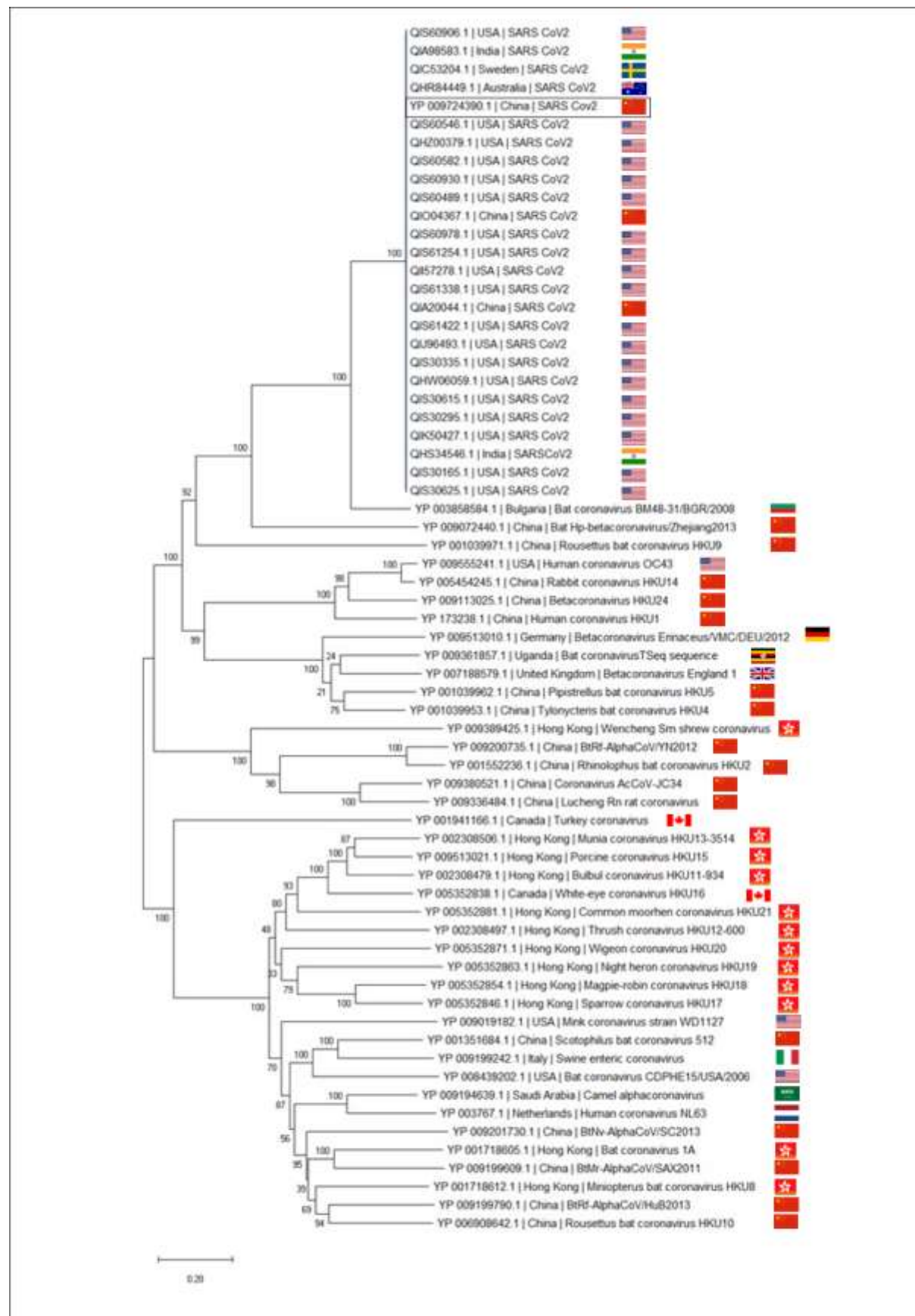


Fig. 3.

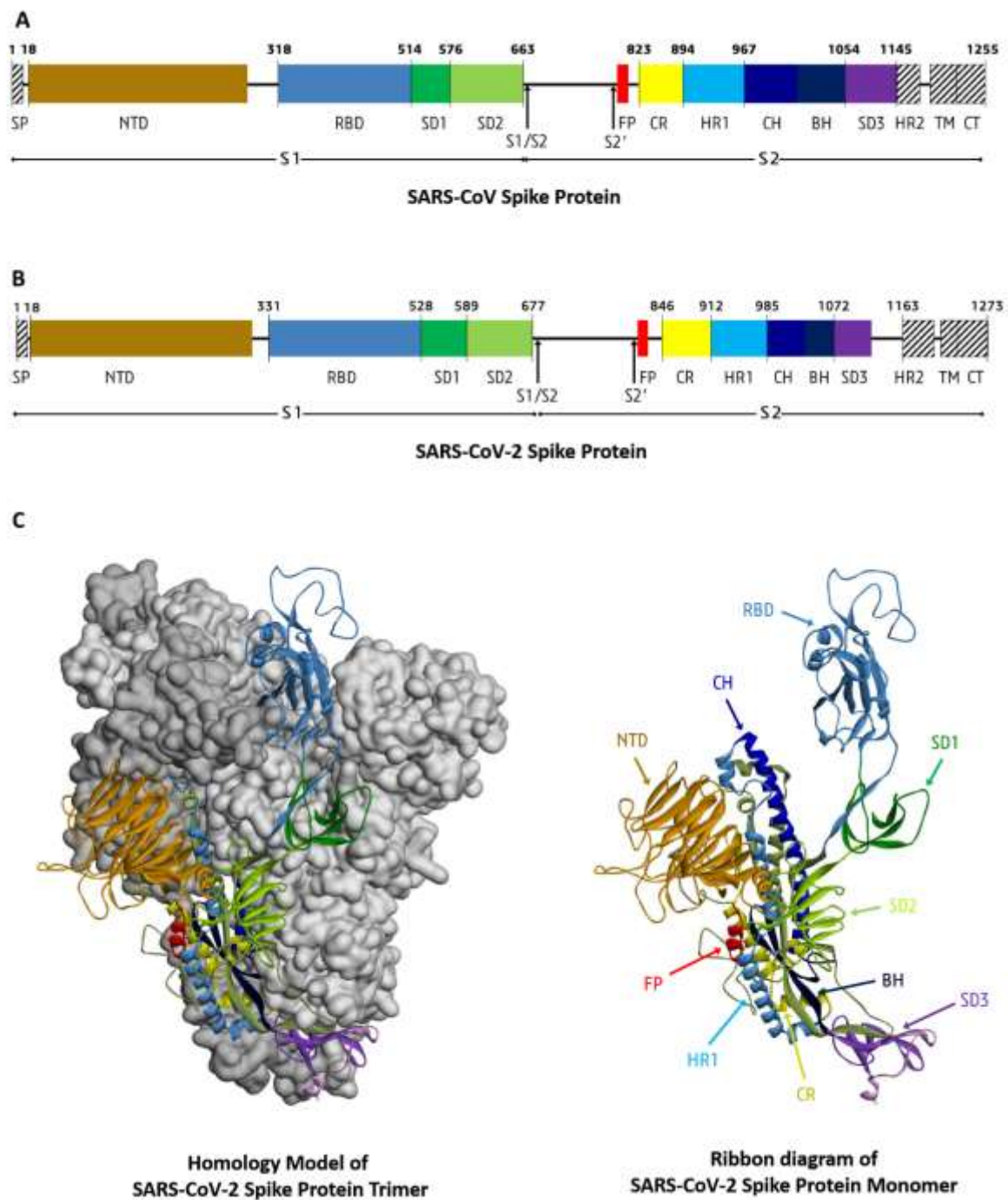
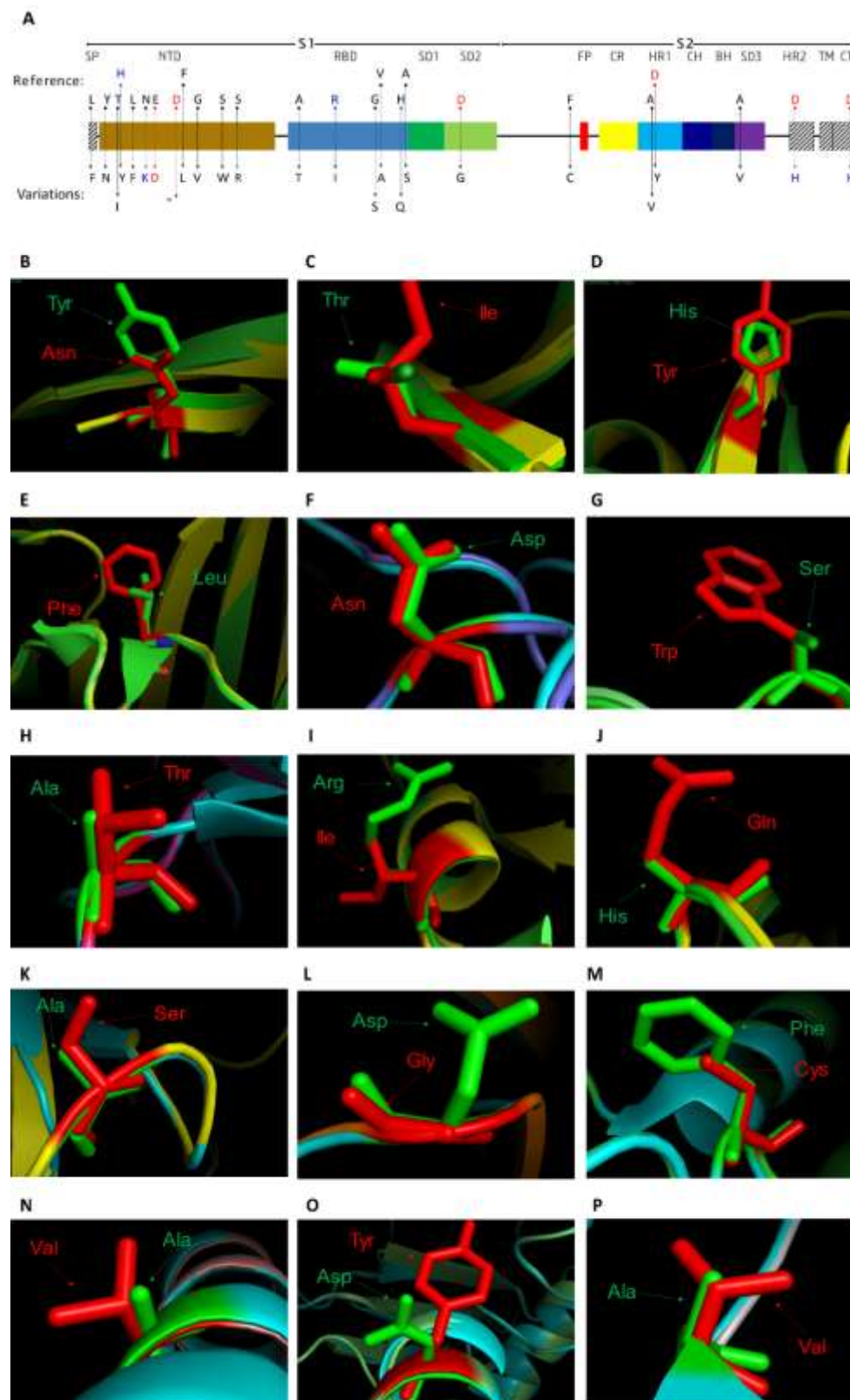
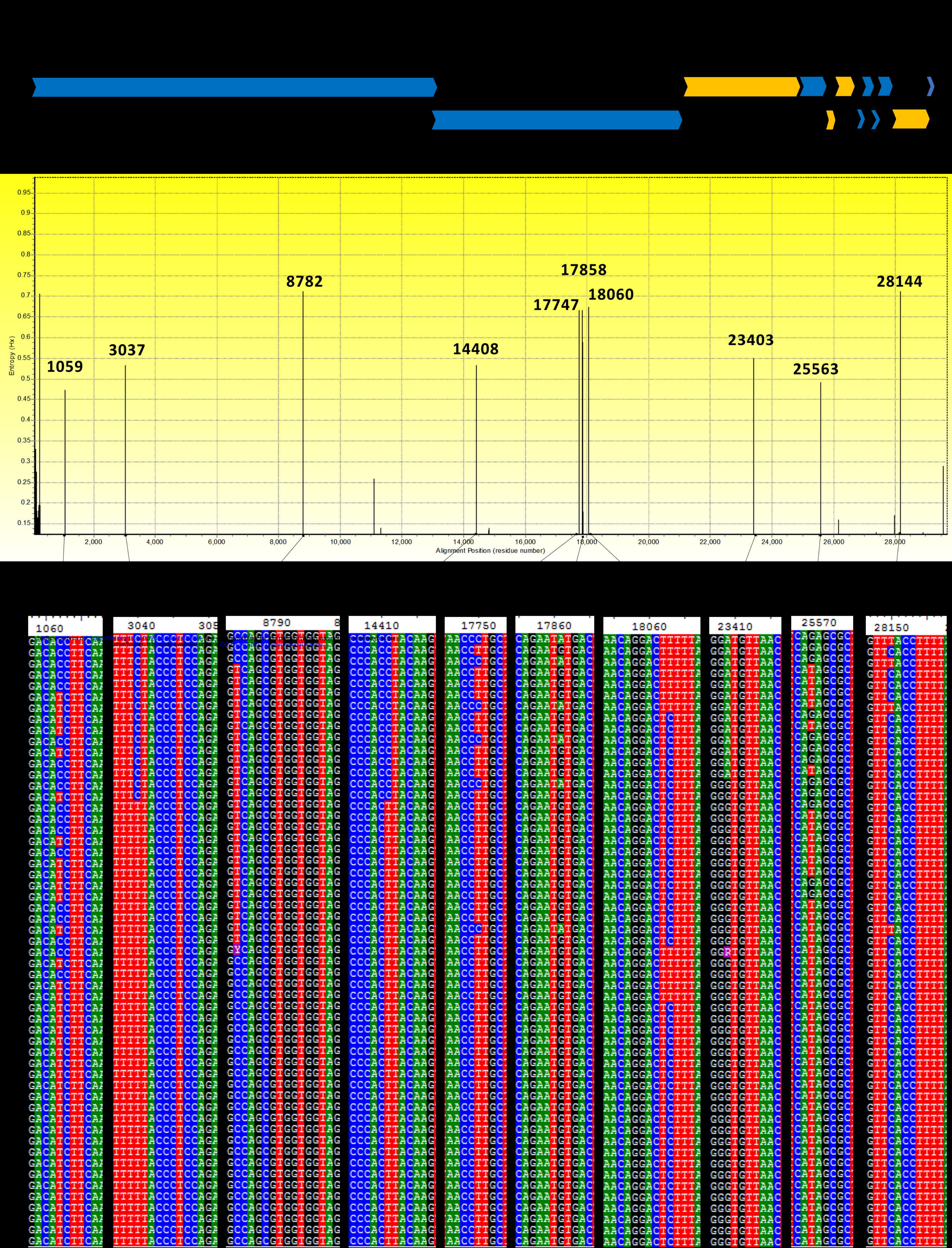
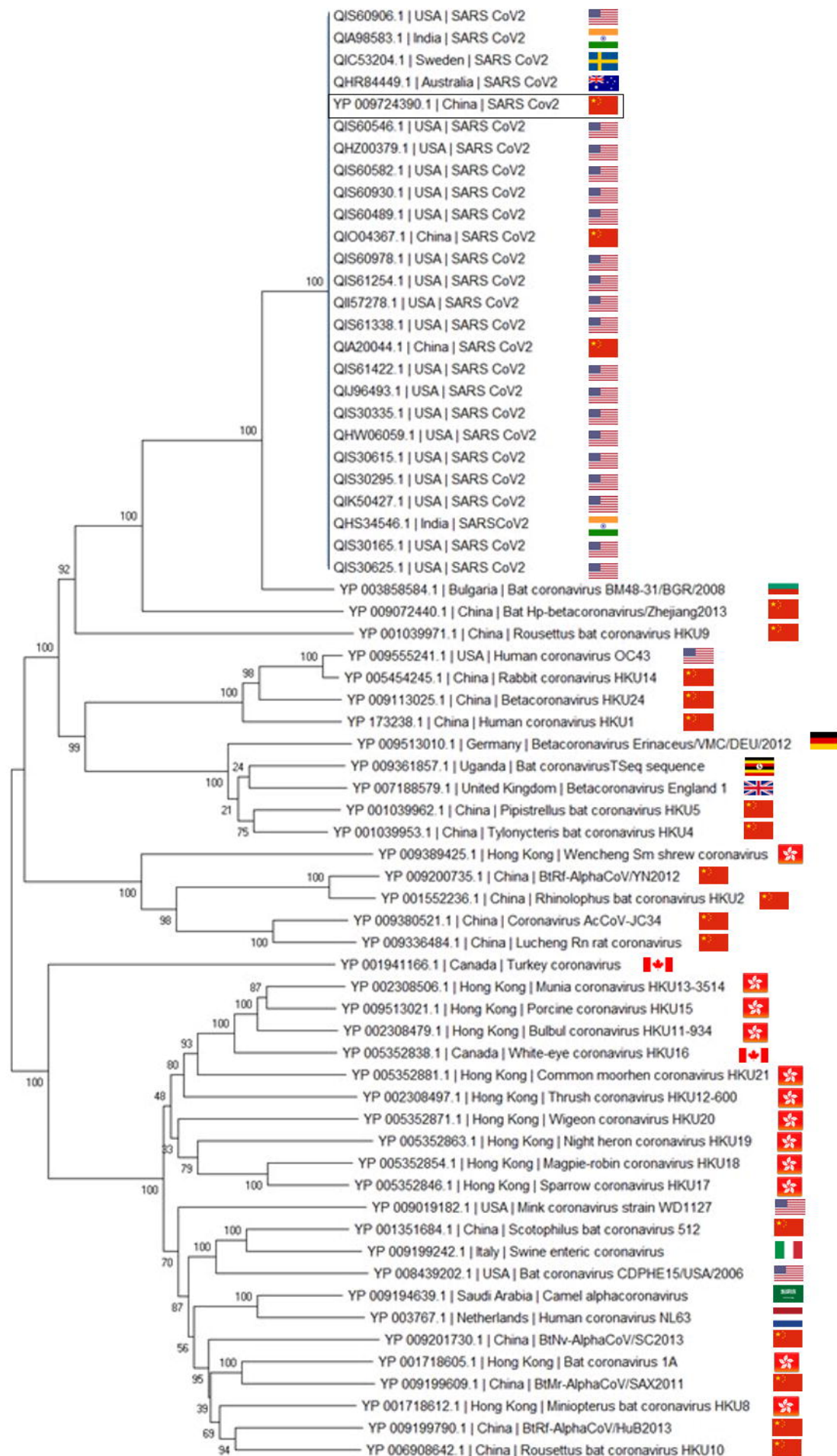


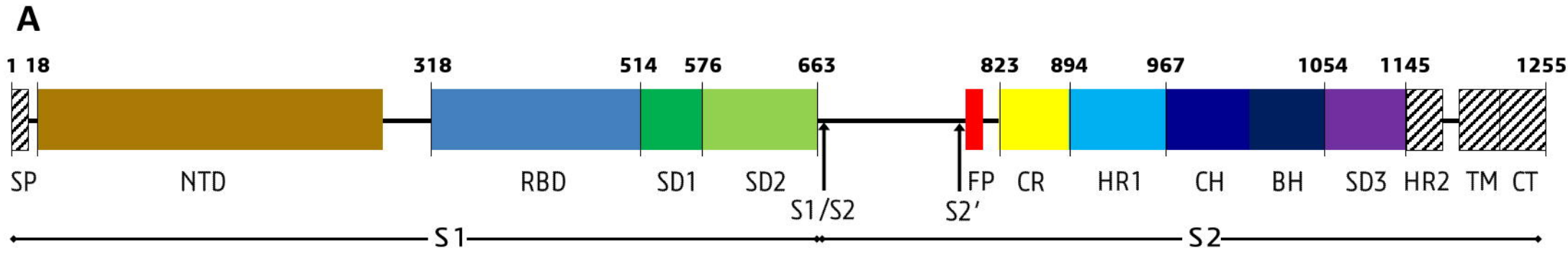
Fig. 4





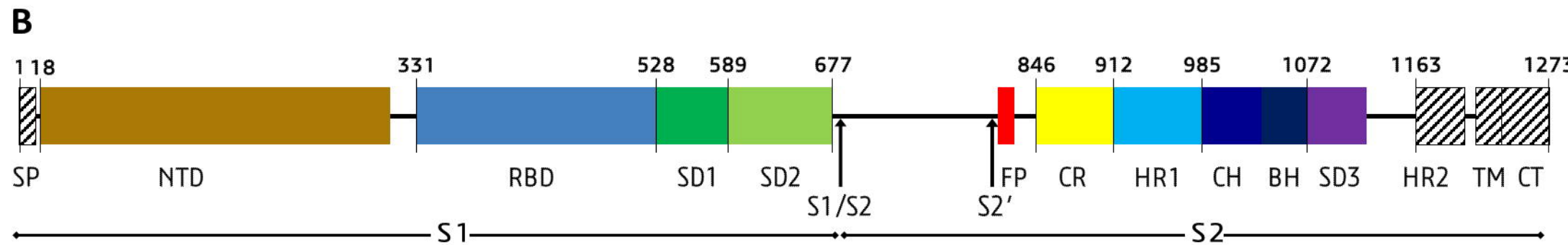


0.20

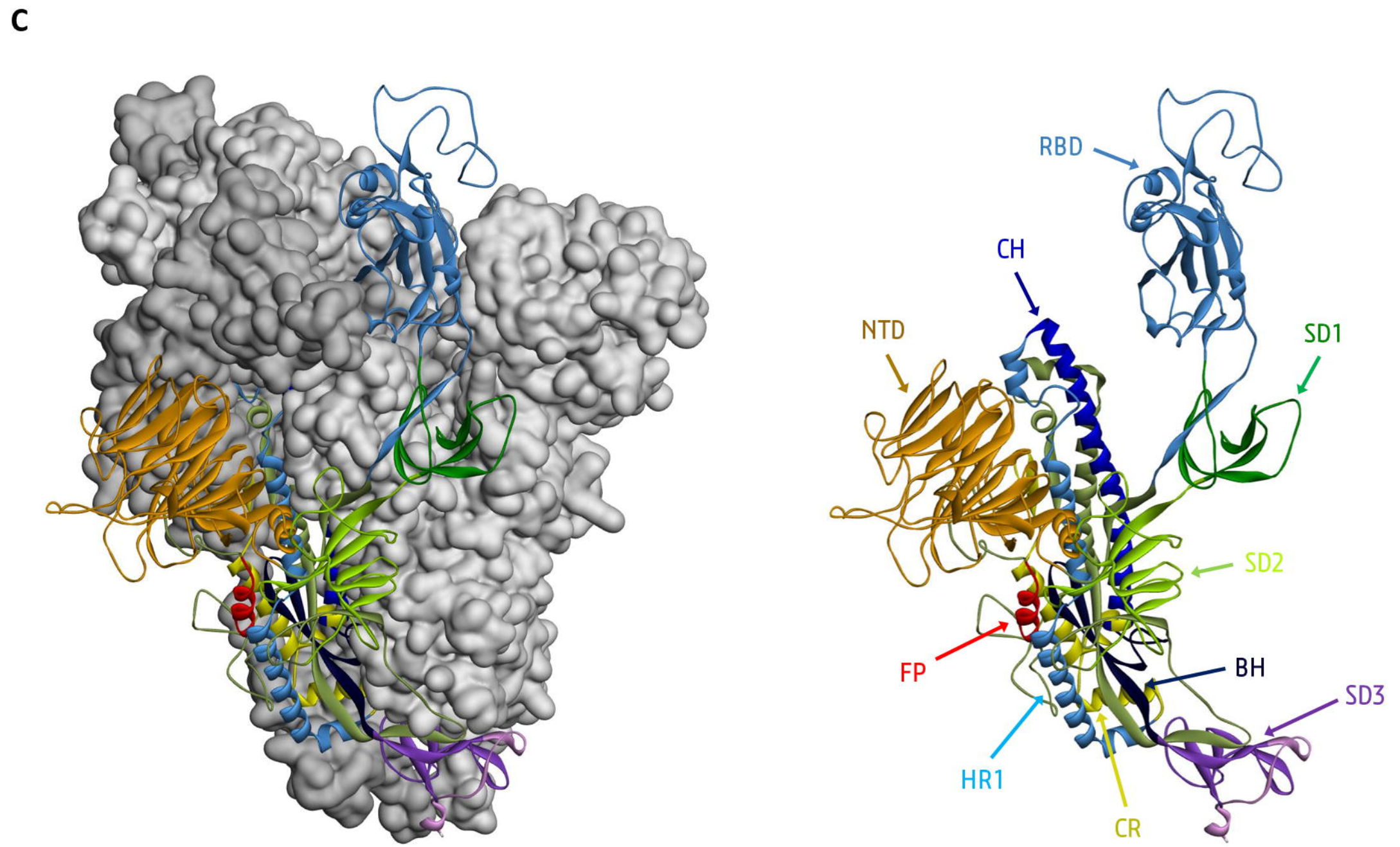


SARS-CoV Spike Protein

bioRxiv preprint doi: <https://doi.org/10.1101/2020.04.07.030924>; this version posted April 7, 2020. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder. It is made available under aCC-BY-NC-ND 4.0 International license.

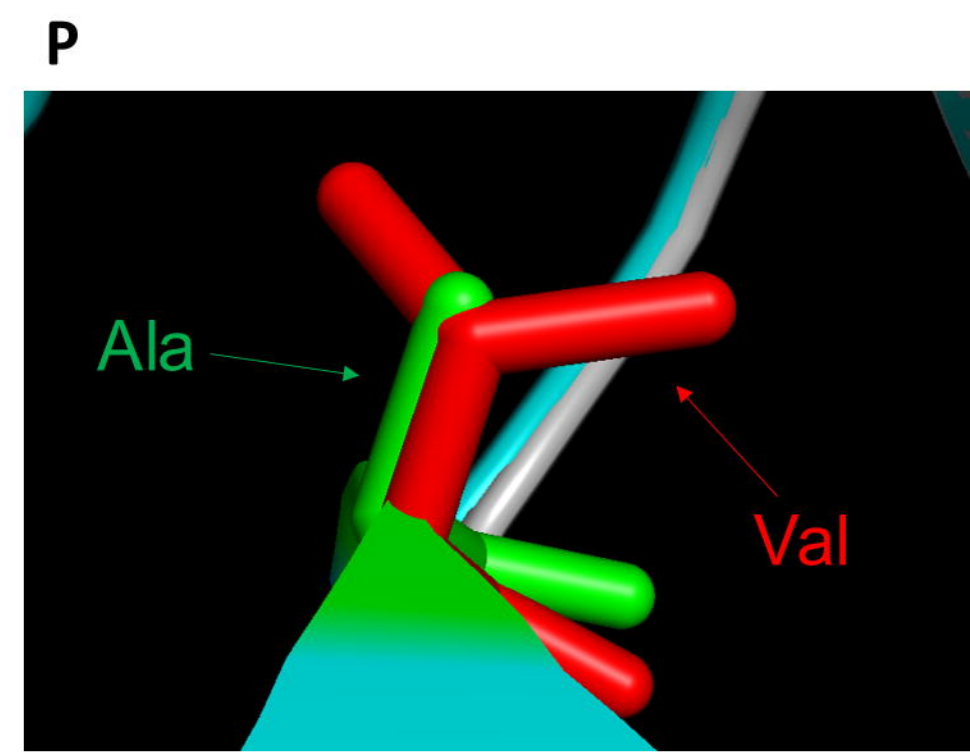
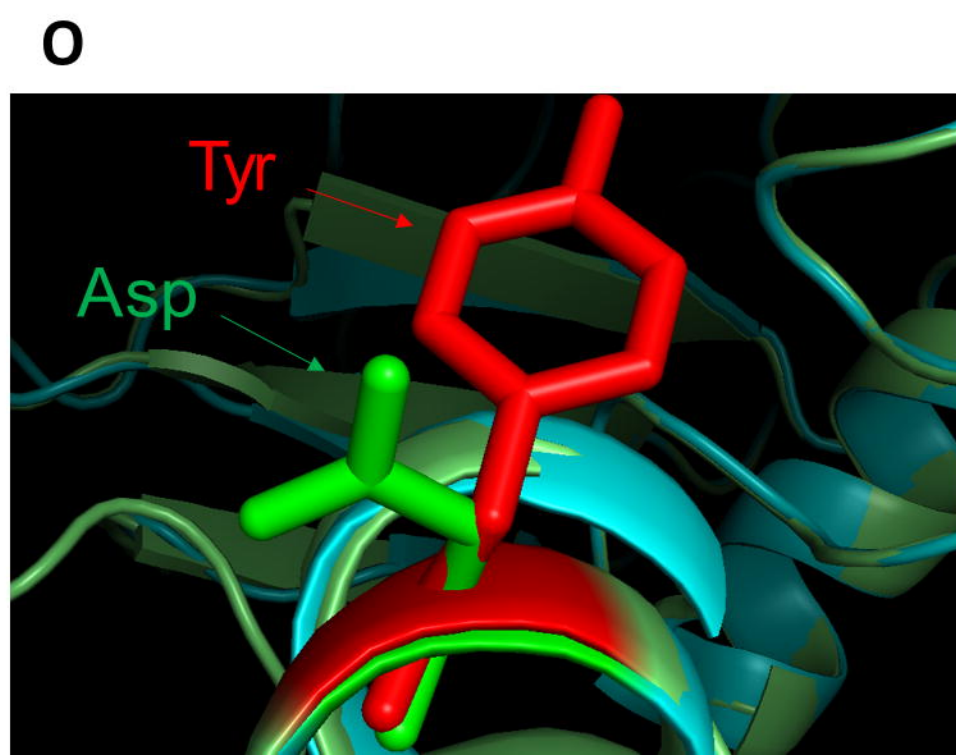
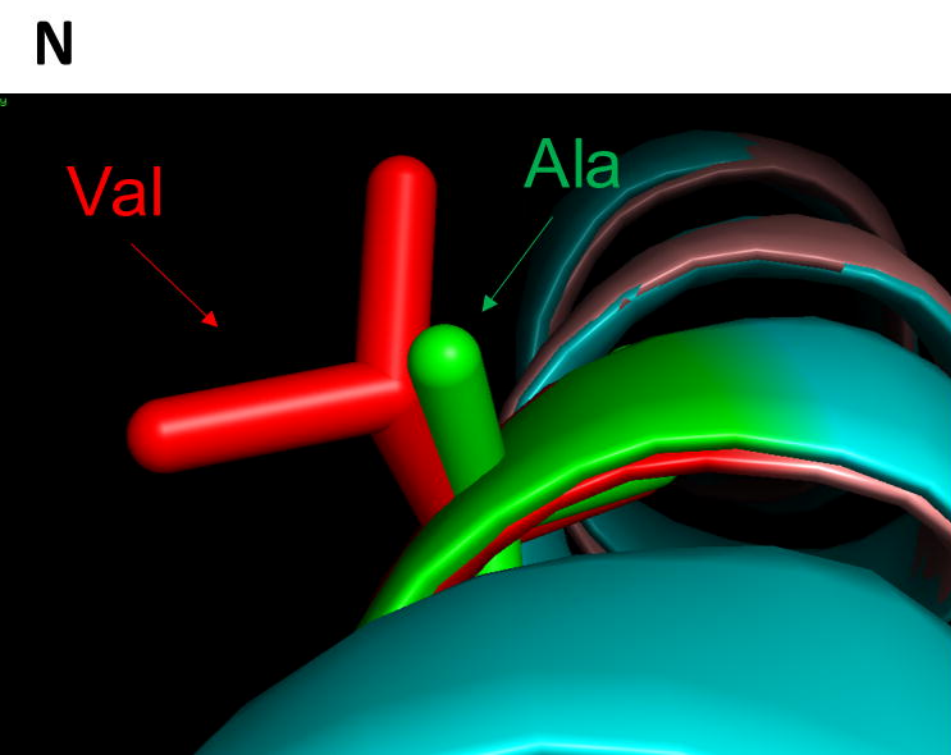
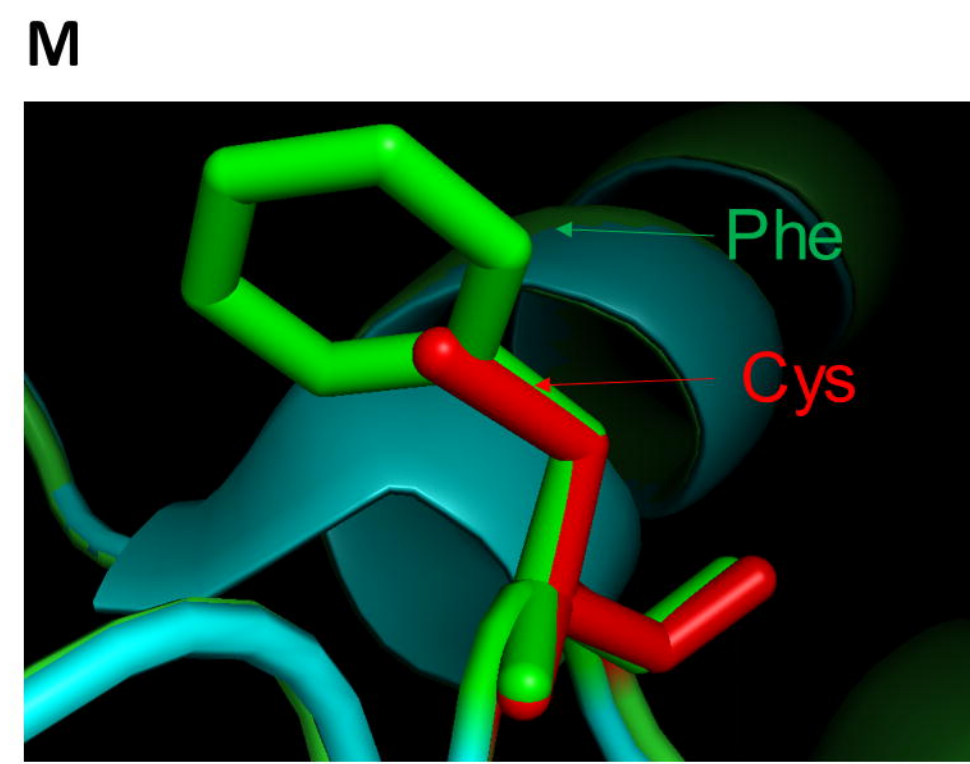
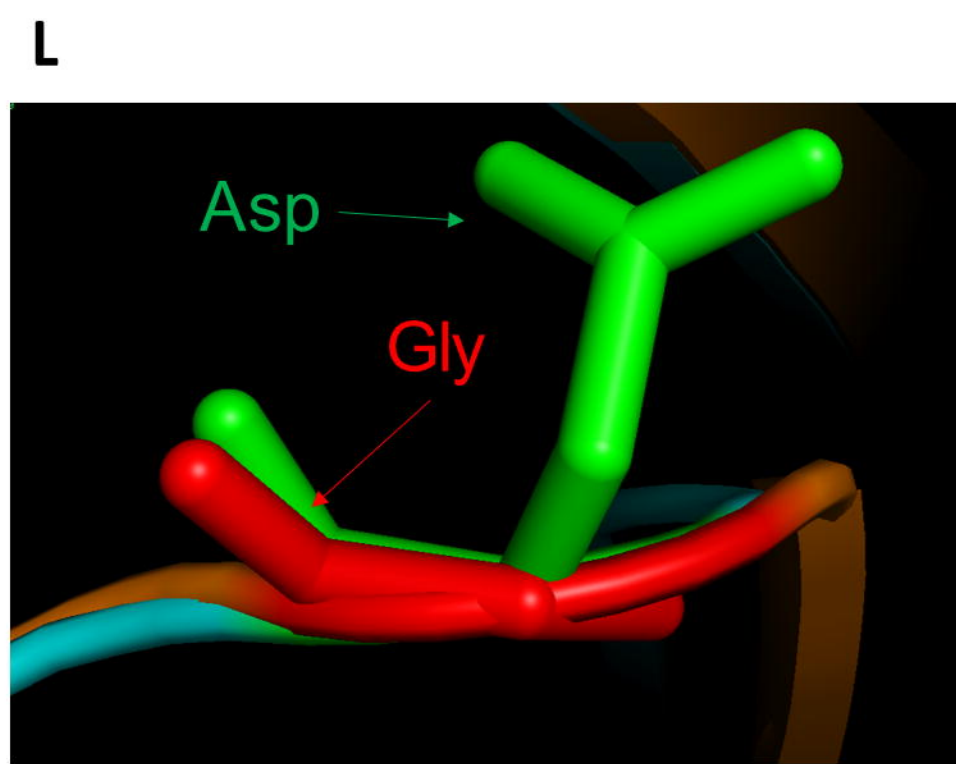
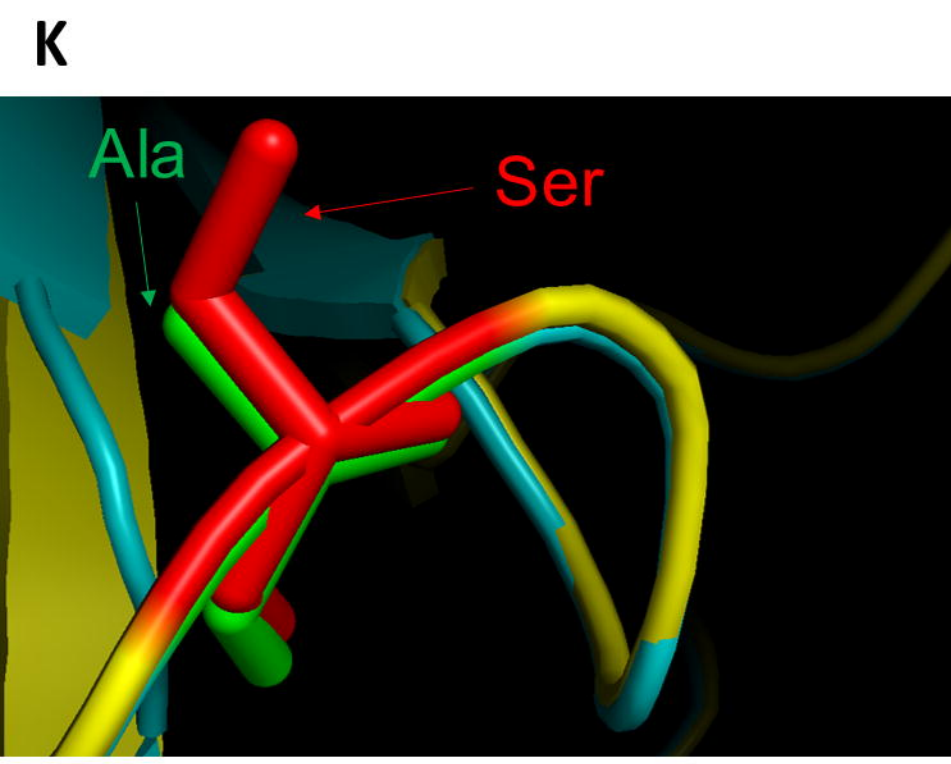
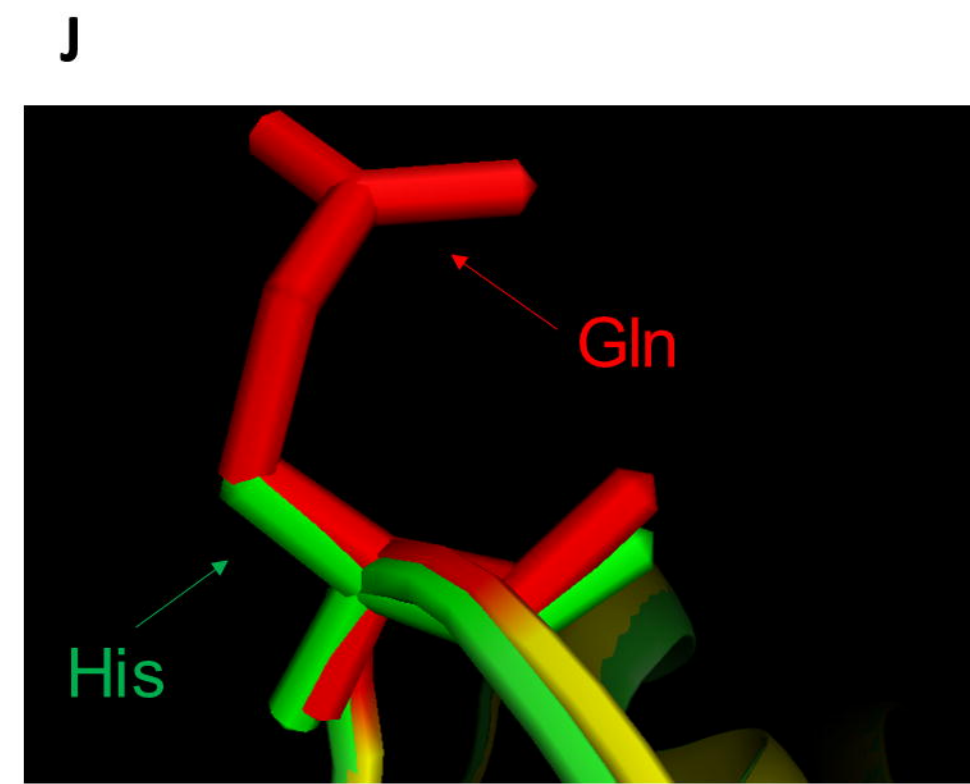
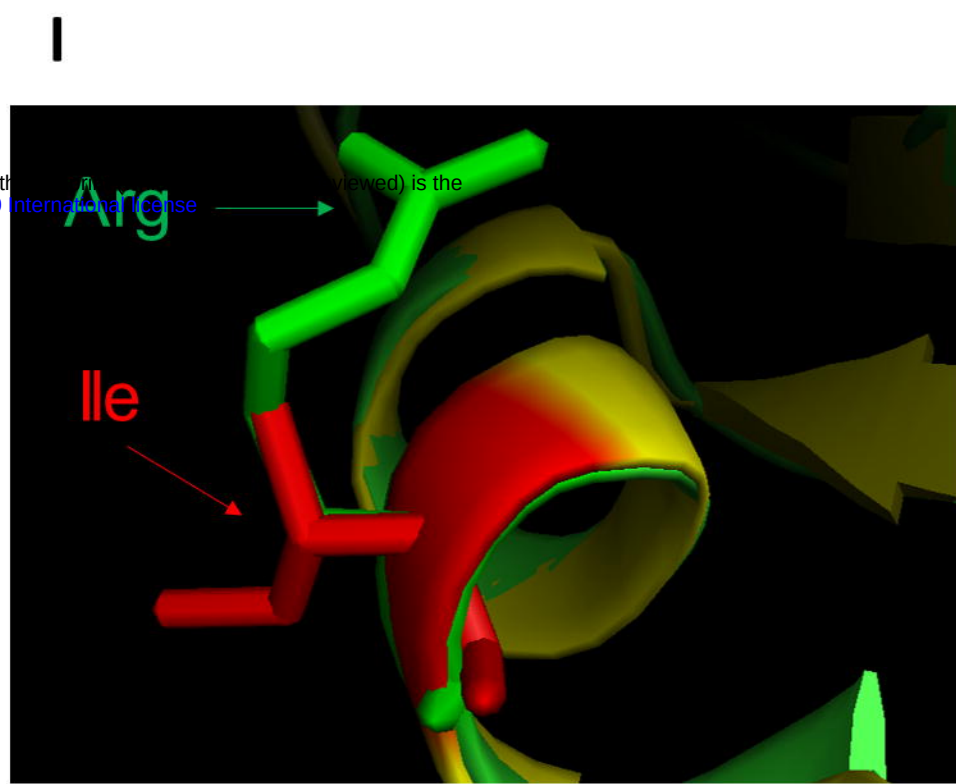
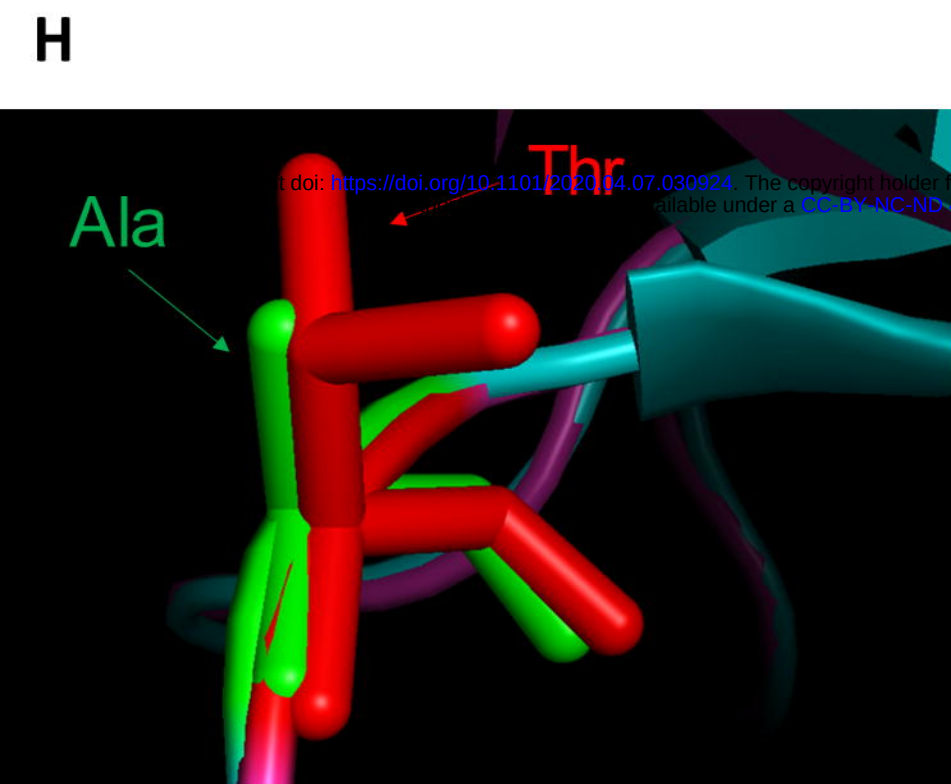
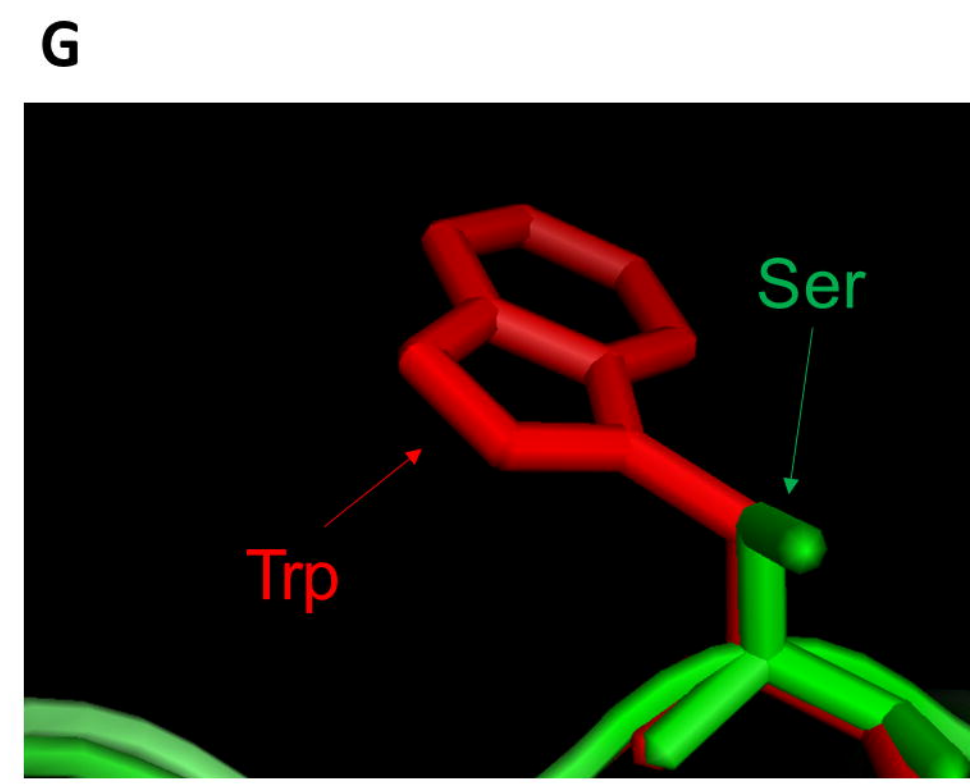
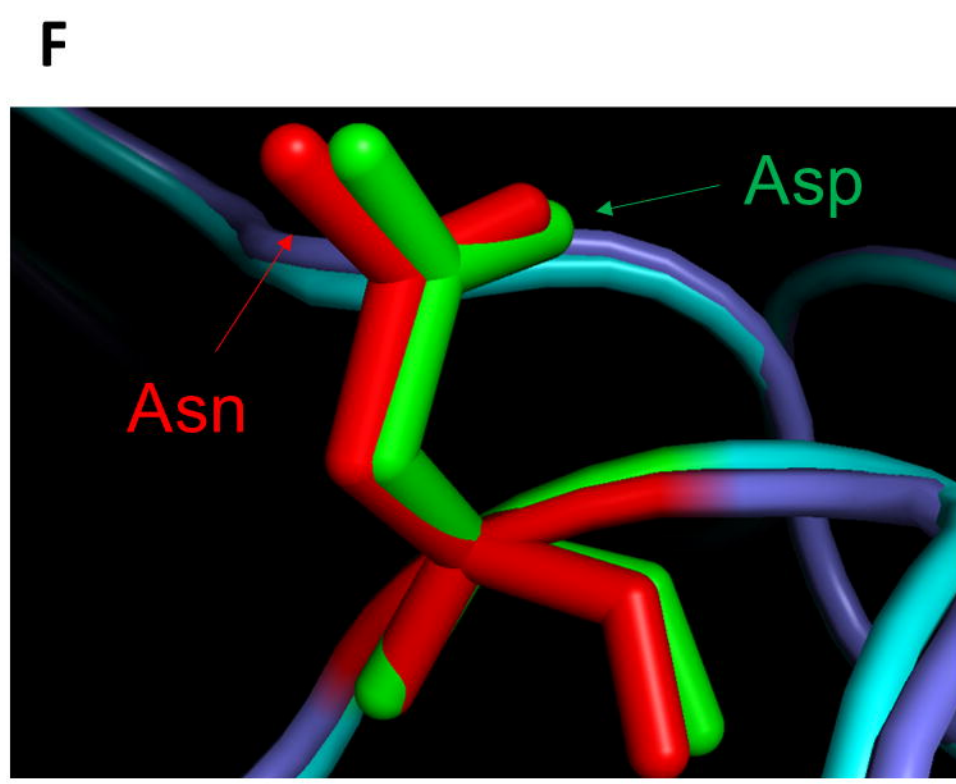
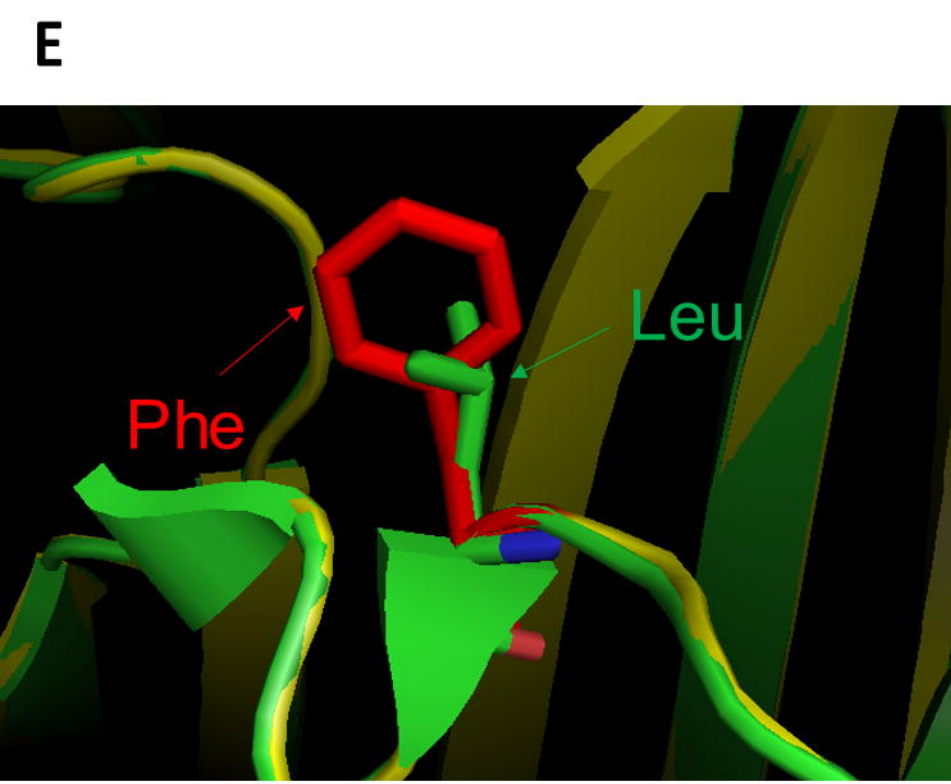
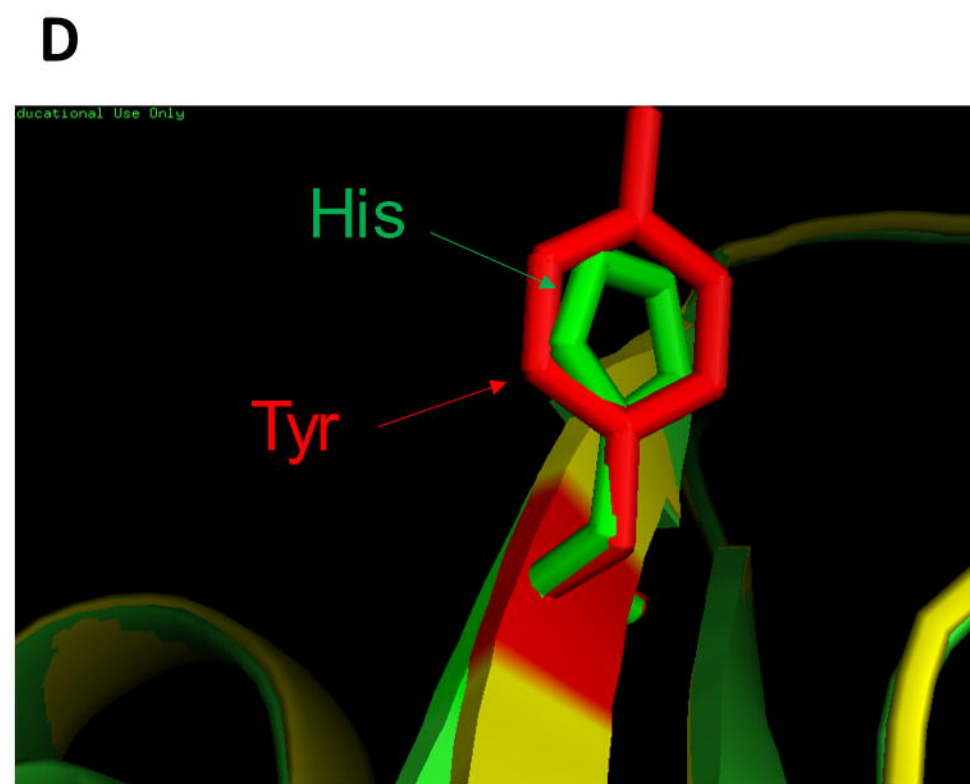
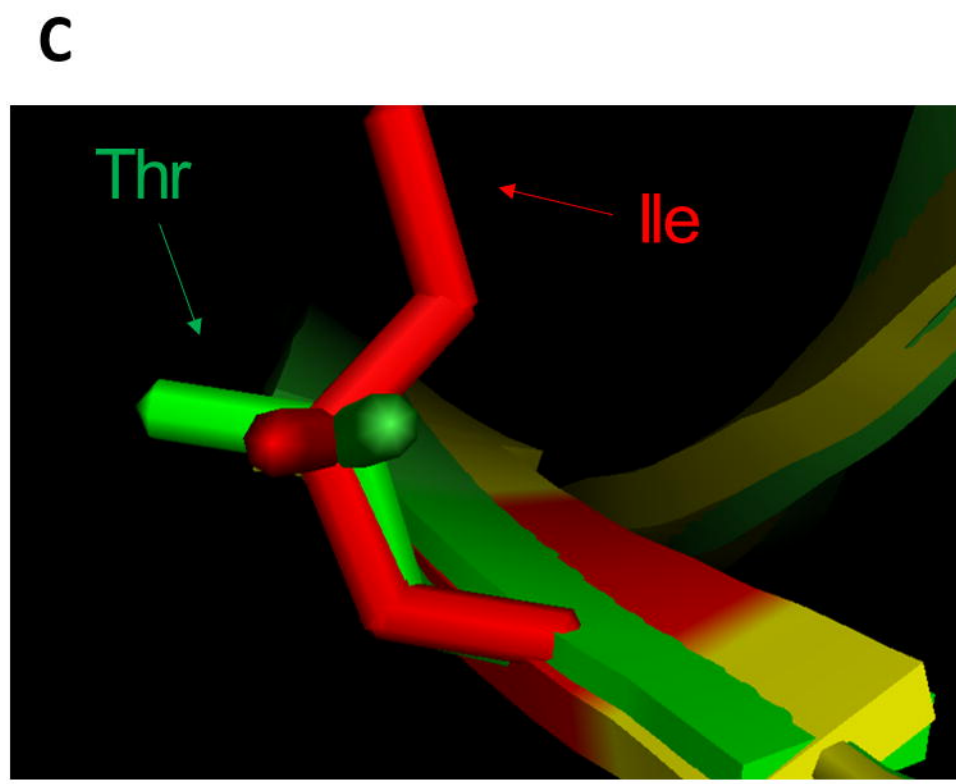
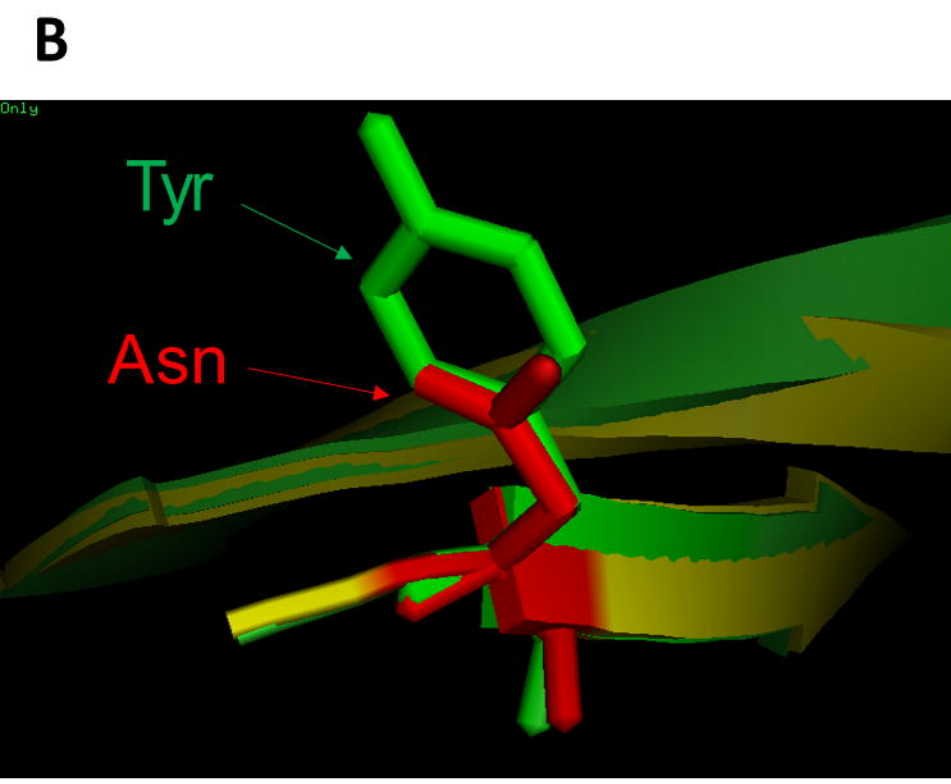
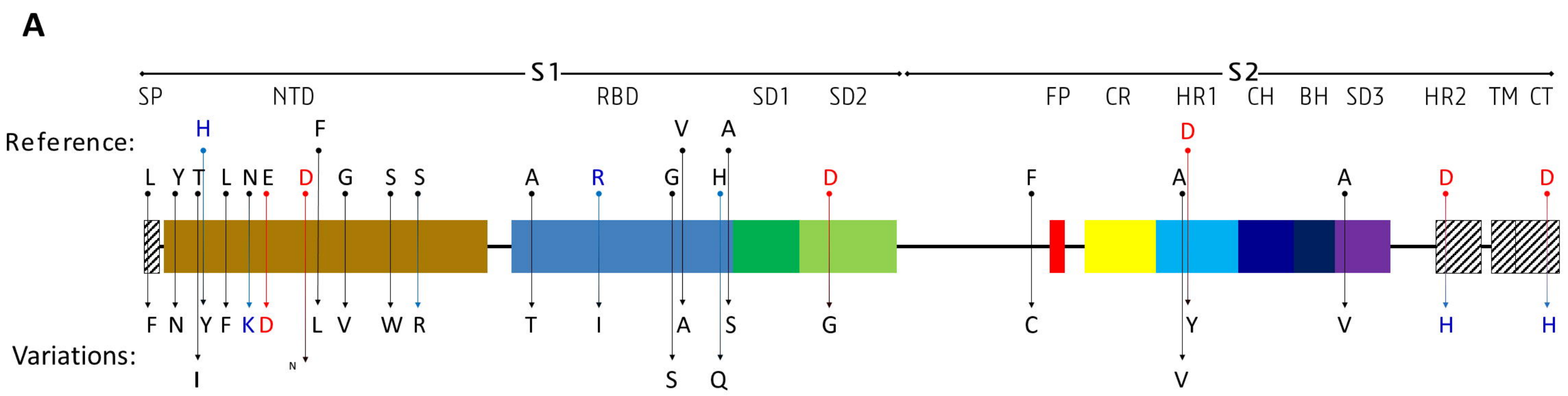


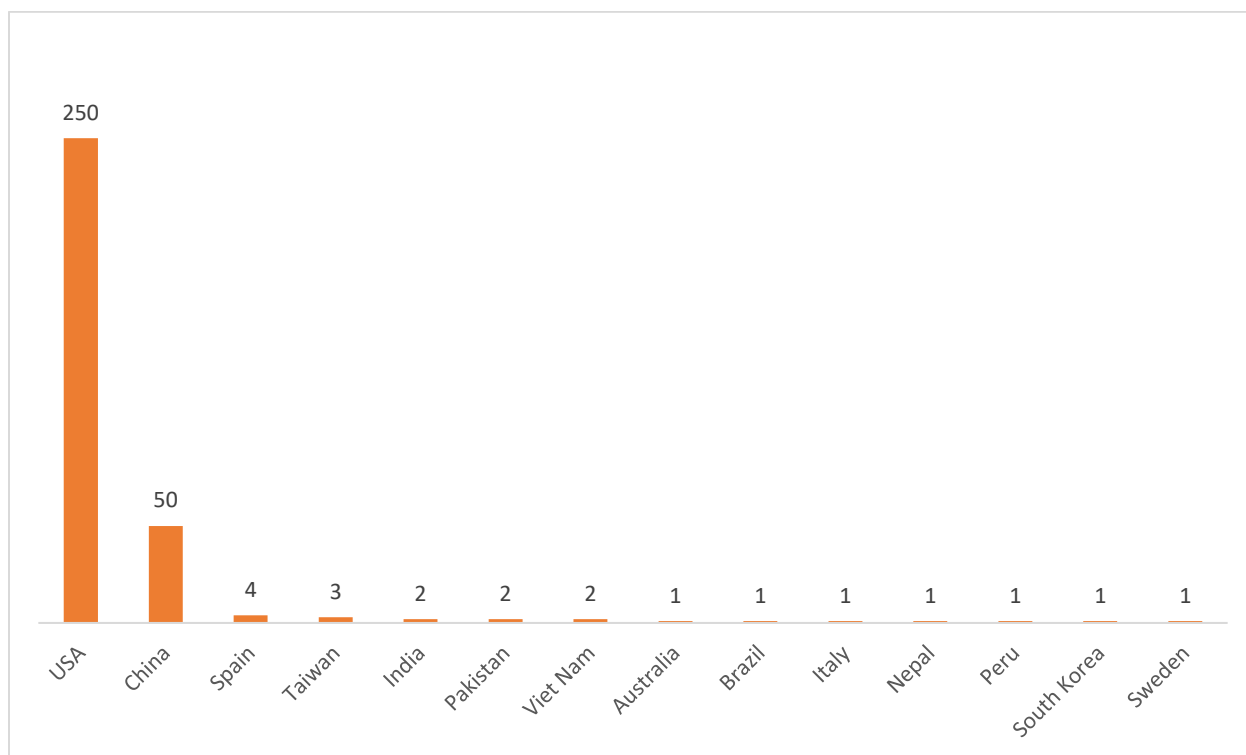
SARS-CoV-2 Spike Protein



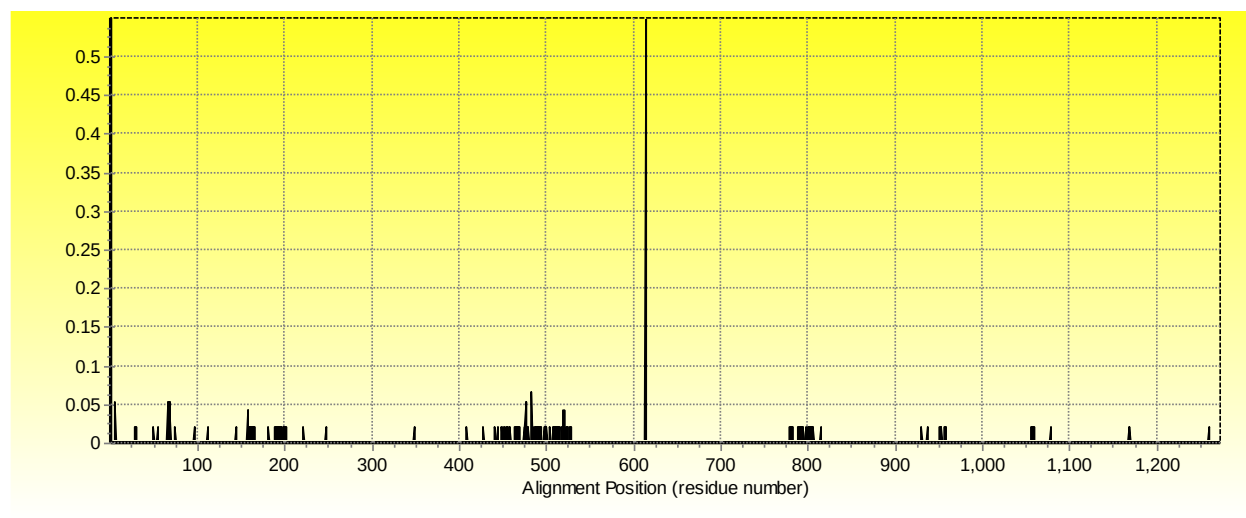
**Homology Model of
SARS-CoV-2 Spike Protein Trimer**

**Ribbon diagram of
SARS-CoV-2 Spike Protein Monomer**





Supplementary Fig. 1. **Genomic data sources based on geo location of 320 SARS-CoV-2 whole genome sequences.**



Supplementary Fig. 2. **Entropy plot measuring positional variability of SARS-CoV-2 spike protein sequences.**

YP_009724390	1	MFVFLVLLPL	VSSQCVN--L	TTRTQLPPAY	TN--SFTRGV	YYPDKVFRSS	VLHSTQDLFL
QHS34546.1	1	MFVFLVLLPL	VSSQCVN--L	TTRTQLPPAY	TN--SFTRGV	YYPDKVFRSS	VLHSTQDLFL
QIS60906.1	1	MFVFLVLLPL	VSSQCVN--L	TTRTQLPPAY	TN--SFTRGV	YYPDKVFRSS	VLHSTQDLFL
QIS60930.1	1	MFVFLVLLPL	VSSQCVN--L	TTRTQLPPAY	TN--SFTRGV	YYPDKVFRSS	VLHSTQDLFL
QIS61338.1	1	MFVFLVLLPL	VSSQCVN--L	TTRTQLPPAY	TN--SFTRGV	YYPDKVFRSS	VLHSTQDLFL
QIS60489.1	1	MFVFLVLLPL	VSSQCVN--L	TTRTQLPPAY	TN--SFTRGV	YYPDKVFRSS	VLHSTQDLFL
QIS30625.1	1	MFVFLVLLPL	VSSQCVN--L	TTRTQLPPAY	TN--SFTRGV	YYPDKVFRSS	VLHSTQDLFL
QIS30335.1	1	MFVFLVLLPL	VSSQCVN--L	TTRTQLPPAY	TN--SFTRGV	YYPDKVFRSS	VLHSTQDLFL
QIC53204.1	1	MFVFLVLLPL	VSSQCVN--L	TTRTQLPPAY	TN--SFTRGV	YYPDKVFRSS	VLHSTQDLFL
QHW06059.1	1	MFVFLVLLPL	VSSQCVN--L	TTRTQLPPAY	TN--SFTRGV	YYPDKVFRSS	VLHSTQDLFL
QHZ00379.1	1	MFVFLVLLPL	VSSQCVN--L	TTRTQLPPAY	TN--SFTRGV	YYPDKVFRSS	VLHSTQDLFL
QHR84449.1	1	MFVFLVLLPL	VSSQCVN--L	TTRTQLPPAY	TN--SFTRGV	YYPDKVFRSS	VLHSTQDLFL
QIJ96493.1	1	MFVFLVLLPL	VSSQCVN--L	TTRTQLPPAY	TN--SFTRGV	YYPDKVFRSS	VLHSTQDLFL
QIA98583.1	1	MFVFLVLLPL	VSSQCVN--L	TTRTQLPPAY	TN--SFTRGV	YYPDKVFRSS	VLHSTQDLFL
QII57278.1	1	MFVFLVLLPL	VSSQCVN--L	TTRTQLPPAY	TN--SFTRGV	YYPDKVFRSS	VLHSTQDLFL
QIA20044.1	1	MFVFLVLLPL	VSSQCVN--L	TTRTQLPPAY	TN--SFTRGV	YYPDKVFRSS	VLHSTQDLFL
QIS60546.1	1	MFVFLVLLPL	VSSQCVN--L	TTRTQLPPAY	TN--SFTRGV	YYPDKVFRSS	VLHSTQDLFL
QIS30625.1	1	MFVFLVLLPL	VSSQCVN--L	TTRTQLPPAY	TN--SFTRGV	YYPDKVFRSS	VLHSTQDLFL
QIS61254.1	1	MFVFLVLLPL	VSSQCVN--L	TTRTQLPPAY	TN--SFTRGV	YYPDKVFRSS	VLHSTQDLFL
QIS61422.1	1	MFVFLVLLPL	VSSQCVN--L	TTRTQLPPAY	TN--SFTRGV	YYPDKVFRSS	VLHSTQDLFL
QIS60978.1	1	MFVFLVLLPL	VSSQCVN--L	TTRTQLPPAY	TN--SFTRGV	YYPDKVFRSS	VLHSTQDLFL
QIO04367.1	1	MFVFLVLLPL	VSSQCVN--L	TTRTQLPPAY	TN--SFTRGV	YYPDKVFRSS	VLHSTQDLFL
QIS30295.1	1	MFVFLVLLPL	VSSQCVN--L	TTRTQLPPAY	TN--SFTRGV	YYPDKVFRSS	VLHSTQDLFL
QIS30615.1	1	MFVFLVLLPL	VSSQCVN--L	TTRTQLPPAY	TN--SFTRGV	YYPDKVFRSS	VLHSTQDLFL
QIK50427.1	1	MFVFLVLLPL	VSSQCVN--L	TTRTQLPPAY	TN--SFTRGV	YYPDKVFRSS	VLHSTQDLFL
QIS30165.1	1	MFVFLVLLPL	VSSQCVN--L	TTRTQLPPAY	TN--SFTRGV	YYPDKVFRSS	VLHSTQDLFL
SARS-CoV	1	MFVFLVLLPL	VSSQCVN--L	TTRTQLPPAY	TN--SFTRGV	YYPDKVFRSS	VLHSTQDLFL

SP

NTD

YP_009724390	57	PFFSNVTWFH	AIHVS GTNGT	KRFDNPVLPF	NDGVYFASTE	KSNIIRGWIF	GTTLD SKTQS
QHS34546.1	57	PFFSNVTWFH	AIHVS GTNGT	KRFDNPVLPF	NDGVYFASTE	KSNIIRGWIF	GTTLD SKTQS
QIS60906.1	57	PFFSNVTWFH	AIHVS GTNGT	KRFDNPVLPF	NDGVYFASTE	KSNIIRGWIF	GTTLD SKTQS
QIS60930.1	57	PFFSNVTWFH	AIHVS GTNGT	KRFDNPVLPF	NDGVYFASTE	KSNIIRGWIF	GTTLD SKTQS
QIS61338.1	57	PFFSNVTWFH	AIHVS GTNGT	KRFDNPVLPF	NDGVYFASTE	KSNIIRGWIF	GTTLD SKTQS
QIS60489.1	57	PFFSNVTWFH	AIHVS GTNGT	KRFDNPVLPF	NDGVYFASTE	KSNIIRGWIF	GTTLD SKTQS
QIS30625.1	57	PFFSNVTWFH	AIHVS GTNGT	KRFDNPVLPF	NDGVYFASTE	KSNIIRGWIF	GTTLD SKTQS
QIS30335.1	57	PFFSNVTWFH	AIHVS GTNGT	KRFDNPVLPF	NDGVYFASTE	KSNIIRGWIF	GTTLD SKTQS
QIC53204.1	57	PFFSNVTWFH	AIHVS GTNGT	KRFDNPVLPF	NDGVYFASTE	KSNIIRGWIF	GTTLD SKTQS
QHW06059.1	57	PFFSNVTWFH	AIHVS GTNGT	KRFDNPVLPF	NDGVYFASTE	KSNIIRGWIF	GTTLD SKTQS
QHZ00379.1	57	PFFSNVTWFH	AIHVS GTNGT	KRFDNPVLPF	NDGVYFASTE	KSNIIRGWIF	GTTLD SKTQS
QHR84449.1	57	PFFSNVTWFH	AIHVS GTNGT	KRFDNPVLPF	NDGVYFASTE	KSNIIRGWIF	GTTLD SKTQS
QIJ96493.1	57	PFFSNVTWFH	AIHVS GTNGT	KRFDNPVLPF	NDGVYFASTE	KSNIIRGWIF	GTTLD SKTQS
QIA98583.1	57	PFFSNVTWFH	AIHVS GTNGT	KRFDNPVLPF	NDGVYFASTE	KSNIIRGWIF	GTTLD SKTQS
QII57278.1	57	PFFSNVTWFH	AIHVS GTNGT	KRFDNPVLPF	NDGVYFASTE	KSNIIRGWIF	GTTLD SKTQS
QIA20044.1	57	PFFSNVTWFH	AIHVS GTNGT	KRFDNPVLPF	NDGVYFASTE	KSNIIRGWIF	GTTLD SKTQS
QIS60546.1	57	PFFSNVTWFH	AIHVS GTNGT	KRFDNPVLPF	NDGVYFASTE	KSNIIRGWIF	GTTLD SKTQS
QIS60582.1	57	PFFSNVTWFH	AIHVS GTNGT	KRFDNPVLPF	NDGVYFASTE	KSNIIRGWIF	GTTLD SKTQS
QIS61254.1	57	PFFSNVTWFH	AIHVS GTNGT	KRFDNPVLPF	NDGVYFASTE	KSNIIRGWIF	GTTLD SKTQS
QIS61422.1	57	PFFSNVTWFH	AIHVS GTNGT	KRFDNPVLPF	NDGVYFASTE	KSNIIRGWIF	GTTLD SKTQS
QIS60978.1	57	PFFSNVTWFH	AIHVS GTNGT	KRFDNPVLPF	NDGVYFASTE	KSNIIRGWIF	GTTLD SKTQS
QIO04367.1	57	PFFSNVTWFH	AIHVS GTNGT	KRFDNPVLPF	NDGVYFASTE	KSNIIRGWIF	GTTLD SKTQS
QIS30295.1	57	PFFSNVTWFH	AIHVS GTNGT	KRFDNPVLPF	NDGVYFASTE	KSNIIRGWIF	GTTLD SKTQS
QIS30615.1	57	PFFSNVTWFH	AIHVS GTNGT	KRFDNPVLPF	NDGVYFASTE	KSNIIRGWIF	GTTLD SKTQS
QIK50427.1	57	PFFSNVTWFH	AIHVS GTNGT	KRFDNPVLPF	NDGVYFASTE	KSNIIRGWIF	GTTLD SKTQS
QIS30165.1	57	PFFSNVTWFH	AIHVS GTNGT	KRFDNPVLPF	NDGVYFASTE	KSNIIRGWIF	GTTLD SKTQS
SARS-CoV	61	PFFSNVTWFH	AIHVS GTNGT	KRFDNPVLPF	NDGVYFASTE	KSNIIRGWIF	GTTLD SKTQS

NTD

YP_009724390	117	LLIVNNATNV	VIKVEFQFC	NDPFLGVYYH	KNNKSWMESE	FRVYSSANNC	TFEYVSQPFL
QHS34546.1	117	LLIVNNATNV	VIKVEFQFC	NDPFLGV-YH	KNNKSWMESE	FRVYSSANNC	TFEYVSQPFL
QIS60906.1	117	LLIVNNATNV	VIKVEFQFC	NDPFLGVYYH	KNNKSWMESE	FRVYSSANNC	TFEYVSQPFL
QIS60930.1	117	LLIVNNATNV	VIKVEFQFC	NDPFLGVYYH	KNNKSWMESE	FRVYSSANNC	TFEYVSQPFL
QIS61338.1	117	LLIVNNATNV	VIKVEFQFC	NDPFLGVYYH	KNNKSWMESE	FRVYSSANNC	TFEYVSQPFL
QIS60489.1	117	LLIVNNATNV	VIKVEFQFC	NDPFLGVYYH	KNNKSWMESE	FRVYSSANNC	TFEYVSQPFL
QIS30625.1	117	LLIVNNATNV	VIKVEFQFC	NDPFLGVYYH	KNNKSWMESE	FRVYSSANNC	TFEYVSQPFL
QIS30335.1	117	LLIVNNATNV	VIKVEFQFC	NDPFLGVYYH	KNNKSWMESE	FRVYSSANNC	TFEYVSQPFL
QIC53204.1	117	LLIVNNATNV	VIKVEFQFC	NDPFLGVYYH	KNNKSWMESE	FRVYSSANNC	TFEYVSQPFL
QHW06059.1	117	LLIVNNATNV	VIKVEFQFC	NDPFLGVYYH	KNNKSWMESE	FRVYSSANNC	TFEYVSQPFL
QHZ00379.1	117	LLIVNNATNV	VIKVEFQFC	NDPFLGVYYH	KNNKSWMESE	FRVYSSANNC	TFEYVSQPFL
QHR84449.1	117	LLIVNNATNV	VIKVEFQFC	NDPFLGVYYH	KNNKSWMESE	FRVYSSANNC	TFEYVSQPFL
QIJ96493.1	117	LLIVNNATNV	VIKVEFQFC	NDPFLGVYYH	KNNKSWMESE	FRVYSSANNC	TFEYVSQPFL
QIA98583.1	117	LLIVNNATNV	VIKVEFQFC	NDPFLGVYYH	KNNKSWMESE	FRVYSSANNC	TFEYVSQPFL
QII57278.1	117	LLIVNNATNV	VIKVEFQFC	NDPFLGVYYH	KNNKSWMESE	FRVYSSANNC	TFEYVSQPFL
QIA20044.1	117	LLIVNNATNV	VIKVEFQFC	NDPFLGVYYH	KNNKSWMESE	FRVYSSANNC	TFEYVSQPFL
QIS60546.1	117	LLIVNNATNV	VIKVEFQFC	NDPFLGVYYH	KNNKSWMESE	FRVYSSANNC	TFEYVSQPFL
QIS60582.1	117	LLIVNNATNV	VIKVEFQFC	NDPFLGVYYH	KNNKSWMESE	FRVYSSANNC	TFEYVSQPFL
QIS61254.1	117	LLIVNNATNV	VIKVEFQFC	NDPFLGVYYH	KNNKSWMESE	FRVYSSANNC	TFEYVSQPFL
QIS61422.1	117	LLIVNNATNV	VIKVEFQFC	NDPFLGVYYH	KNNKSWMESE	FRVYSSANNC	TFEYVSQPFL
QIS60978.1	117	LLIVNNATNV	VIKVEFQFC	NDPFLGVYYH	KNNKSWMESE	FRVYSSANNC	TFEYVSQPFL
QIO04367.1	117	LLIVNNATNV	VIKVEFQFC	NDPFLGVYYH	KNNKSWMESE	FRVYSSANNC	TFEYVSQPFL
QIS30295.1	117	LLIVNNATNV	VIKVEFQFC	NDPFLGVYYH	KNNKSWMESE	FRVYSSANNC	TFEYVSQPFL
QIS30615.1	117	LLIVNNATNV	VIKVEFQFC	NDPFLGVYYH	KNNKSWMESE	FRVYSSANNC	TFEYVSQPFL
QIK50427.1	117	LLIVNNATNV	VIKVEFQFC	NDPFLGVYYH	KNNKSWMESE	FRVYSSANNC	TFEYVSQPFL
QIS30165.1	117	LLIVNNATNV	VIKVEFQFC	NDPFLGVYYH	KNNKSWMESE	FRVYSSANNC	TFEYVSQPFL
SARS-CoV	114	VIIINNSTNV	VIRACNEELC	NDPFLGVSKP	MG----TQTH	TMIFDNAFNC	TFEYISDAFS

NTD

YP_009724390	177	MDLEGKQGNF	KNLREFVFKN	IDGYFKIYSK	HTPINLVRDL	PQGFSALEPL	VDLPIGINIT
QHS34546.1	176	MDLEGKQGNF	KNLREFVFKN	IDGYFKIYSK	HTPINLVRDL	PQGFSALEPL	VDLPIGINIT
QIS60906.1	177	MDLEGKQGNF	KNLREFVFKN	IDGYFKIYSK	HTPINLVRDL	PQGFSALEPL	VDLPIGINIT
QIS60930.1	177	MDLEGKQGNF	KNLREFVFKN	IDGYFKIYSK	HTPINLVRDL	PQGFSALEPL	VDLPIGINIT
QIS61338.1	177	MDLEGKQGNF	KNLREFVFKN	IDGYFKIYSK	HTPINLVRDL	PQGFSALEPL	VDLPIGINIT
QIS60489.1	177	MDLEGKQGNF	KNLREFVFKN	IDGYFKIYSK	HTPINLVRDL	PQGFSALEPL	VDLPIGINIT
QIS30625.1	177	MDLEGKQGNF	KNLREFVFKN	IDGYFKIYSK	HTPINLVRDL	PQGFSALEPL	VDLPIGINIT
QIS30335.1	177	MDLEGKQGNF	KNLREFVFKN	IDGYFKIYSK	HTPINLVRDL	PQGFSALEPL	VDLPIGINIT
QIC53204.1	177	MDLEGKQGNF	KNLREFVFKN	IDGYFKIYSK	HTPINLVRDL	PQGFSALEPL	VDLPIGINIT
QHW06059.1	177	MDLEGKQGNF	KNLREFVFKN	IDGYFKIYSK	HTPINLVRDL	PQGFSALEPL	VDLPIGINIT
QHZ00379.1	177	MDLEGKQGNF	KNLREFVFKN	IDGYFKIYSK	HTPINLVRDL	PQGFVALEPL	VDLPIGINIT
QHR84449.1	177	MDLEGKQGNF	KNLREFVFKN	IDGYFKIYSK	HTPINLVRDL	PQGFSALEPL	VDLPIGINIT
QIJ96493.1	177	MDLEVKQGNF	KNLREFVFKN	IDGYFKIYSK	HTPINLVRDL	PQGFSALEPL	VDLPIGINIT
QIA98583.1	177	MDLEGKQGNF	KNLREFVFKN	IDGYFKIYSK	HTPINLVRDL	PQGFSALEPL	VDLPIGINIT
QII57278.1	177	MDLEGKQGNF	KNLREFVFKN	IDGYFKIYSK	HTPINLVRDL	PQGFSALEPL	VDLPIGINIT
QIA20044.1	177	MDLEGKQGNF	KNLREFVFKN	IDGYFKIYSK	HTPINLVRDL	PQGFSALEPL	VDLPIGINIT
QIS60546.1	177	MDLEGKQGNF	KNLREFVFKN	IDGYFKIYSK	HTPINLVRDL	PQGFSALEPL	VDLPIGINIT
QIS60582.1	177	MDLEGKQGNF	KNLREFVFKN	IDGYFKIYSK	HTPINLVRDL	PQGFSALEPL	VDLPIGINIT
QIS61254.1	177	MDLEGKQGNF	KNLREFVFKN	IDGYFKIYSK	HTPINLVRDL	PQGFSALEPL	VDLPIGINIT
QIS61422.1	177	MDLEGKQGNF	KNLREFVFKN	IDGYFKIYSK	HTPINLVRDL	PQGFSALEPL	VDLPIGINIT
QIS60978.1	177	MDLEGKQGNF	KNLREFVFKN	IDGYFKIYSK	HTPINLVRDL	PQGFSALEPL	VDLPIGINIT
QIO04367.1	177	MDLEGKQGNF	KNLREFVFKN	IDGYFKIYSK	HTPINLVRDL	PQGFSALEPL	VDLPIGINIT
QIS30295.1	177	MDLEGKQGNF	KNLREFVFKN	IDGYFKIYSK	HTPINLVRDL	PQGFSALEPL	VDLPIGINIT
QIS30615.1	177	MDLEGKQGNF	KNLREFVFKN	IDGYFKIYSK	HTPINLVRDL	PQGFSALEPL	VDLPIGINIT
QIK50427.1	177	MDLEGKQGNF	KNLREFVFKN	IDGYFKIYSK	HTPINLVRDL	PQGFSALEPL	VDLPIGINIT
QIS30165.1	177	MDLEGKQGNF	KNLREFVFKN	IDGYFKIYSK	HTPINLVRDL	PQGFSALEPL	VDLPIGINIT
SARS-CoV	170	LDVSEKSGNF	KHLREFVFKN	KDGLYVYKG	YQPIDVVRDL	PSGENTLKPI	FKLPLGINIT

NTD

YP_009724390	237	RFQTLALHR	SYLTPGDSSS	GWTAGAAAYY	VGYLQPRTFI	LKYNENGTIT	DAVDCALDPL
QHS34546.1	236	RFQTLALHR	SYLTPGDSSS	GWTAGAAAYY	VGYLQPRTFI	LKYNENGTIT	DAVDCALDPL
QIS60906.1	237	RFQTLALHR	SYLTPGDSSS	GWTAGAAAYY	VGYLQPRTFI	LKYNENGTIT	DAVDCALDPL
QIS60930.1	237	RFQTLALHR	SYLTPGDSSS	GWTAGAAAYY	VGYLQPRTFI	LKYNENGTIT	DAVDCALDPL
QIS61338.1	237	RFQTLALHR	SYLTPGDSSS	GWTAGAAAYY	VGYLQPRTFI	LKYNENGTIT	DAVDCALDPL
QIS60489.1	237	RFQTLALHR	SYLTPGDSSS	GWTAGAAAYY	VGYLQPRTFI	LKYNENGTIT	DAVDCALDPL
QIS30625.1	237	RFQTLALHR	SYLTPGDSSS	GWTAGAAAYY	VGYLQPRTFI	LKYNENGTIT	DAVDCALDPL
QIS30335.1	237	RFQTLALHR	SYLTPGDSSS	GWTAGAAAYY	VGYLQPRTFI	LKYNENGTIT	DAVDCALDPL
QIC53204.1	237	RFQTLALHR	SYLTPGDSSS	GWTAGAAAYY	VGYLQPRTFI	LKYNENGTIT	DAVDCALDPL
QHW06059.1	237	RFQTLALHR	SYLTPGDSSS	GWTAGAAAYY	VGYLQPRTFI	LKYNENGTIT	DAVDCALDPL
QHZ00379.1	237	RFQTLALHR	SYLTPGDSSS	GWTAGAAAYY	VGYLQPRTFI	LKYNENGTIT	DAVDCALDPL
QHR84449.1	237	RFQTLALHR	SYLTPGDSSS	GWTAGAAAYY	VGYLQPRTFI	LKYNENGTIT	DAVDCALDPL
QIJ96493.1	237	RFQTLALHR	SYLTPGDSSS	GWTAGAAAYY	VGYLQPRTFI	LKYNENGTIT	DAVDCALDPL
QIA98583.1	237	RFQTLALHR	SYLTPGDSSS	GWTAGAAAYY	VGYLQPRTFI	LKYNENGTIT	DAVDCALDPL
QII57278.1	237	RFQTLALHR	SYLTPGDSSS	GWTAGAAAYY	VGYLQPRTFI	LKYNENGTIT	DAVDCALDPL
QIA20044.1	237	RFQTLALHR	SYLTPGDSSS	GWTAGAAAYY	VGYLQPRTFI	LKYNENGTIT	DAVDCALDPL
QIS60546.1	237	RFQTLALHR	SYLTPGDSSS	GWTAGAAAYY	VGYLQPRTFI	LKYNENGTIT	DAVDCALDPL
QIS60582.1	237	RFQTLALHR	SYLTPGDSSS	GWTAGAAAYY	VGYLQPRTFI	LKYNENGTIT	DAVDCALDPL
QIS61254.1	237	RFQTLALHR	SYLTPGDSSS	GWTAGAAAYY	VGYLQPRTFI	LKYNENGTIT	DAVDCALDPL
QIS61422.1	237	RFQTLALHR	SYLTPGDSSS	GWTAGAAAYY	VGYLQPRTFI	LKYNENGTIT	DAVDCALDPL
QIS60978.1	237	RFQTLALHR	SYLTPGDSSS	GWTAGAAAYY	VGYLQPRTFI	LKYNENGTIT	DAVDCALDPL
QIO04367.1	237	RFQTLALHR	SYLTPGDSSS	GWTAGAAAYY	VGYLQPRTFI	LKYNENGTIT	DAVDCALDPL
QIS30295.1	237	RFQTLALHR	SYLTPGDSSS	GWTAGAAAYY	VGYLQPRTFI	LKYNENGTIT	DAVDCALDPL
QIS30615.1	237	RFQTLALHR	SYLTPGDSSS	GWTAGAAAYY	VGYLQPRTFI	LKYNENGTIT	DAVDCALDPL
QIK50427.1	237	RFQTLALHR	SYLTPGDSSS	GWTAGAAAYY	VGYLQPRTFI	LKYNENGTIT	DAVDCALDPL
QIS30165.1	237	RFQTLALHR	SYLTPGDSSS	GWTAGAAAYY	VGYLQPRTFI	LKYNENGTIT	DAVDCALDPL
SARS-CoV	230	NFRALITAFS	-----PAQD	IWGTSAAAYF	VGYLQPRTFI	LKYNENGTIT	DAVDCSQNPL

NTD

YP_009724390	297	SETKCTLKSF	TVEKGIYQTS	NFRVQPTESI	VRFPNITNLC	PFGEVFNATR	FASVYAWNRR
QHS34546.1	296	SETKCTLKSF	TVEKGIYQTS	NFRVQPTESI	VRFPNITNLC	PFGEVFNATR	FASVYAWNRR
QIS60906.1	297	SETKCTLKSF	TVEKGIYQTS	NFRVQPTESI	VRFPNITNLC	PFGEVFNATR	FASVYAWNRR
QIS60930.1	297	SETKCTLKSF	TVEKGIYQTS	NFRVQPTESI	VRFPNITNLC	PFGEVFNATR	FASVYAWNRR
QIS61338.1	297	SETKCTLKSF	TVEKGIYQTS	NFRVQPTESI	VRFPNITNLC	PFGEVFNATR	FASVYAWNRR
QIS60489.1	297	SETKCTLKSF	TVEKGIYQTS	NFRVQPTESI	VRFPNITNLC	PFGEVFNATR	FASVYAWNRR
QIS30625.1	297	SETKCTLKSF	TVEKGIYQTS	NFRVQPTESI	VRFPNITNLC	PFGEVFNATR	FASVYAWNRR
QIS30335.1	297	SETKCTLKSF	TVEKGIYQTS	NFRVQPTESI	VRFPNITNLC	PFGEVFNATR	FASVYAWNRR
QIC53204.1	297	SETKCTLKSF	TVEKGIYQTS	NFRVQPTESI	VRFPNITNLC	PFGEVFNATR	FASVYAWNRR
QHW06059.1	297	SETKCTLKSF	TVEKGIYQTS	NFRVQPTESI	VRFPNITNLC	PFGEVFNATR	FASVYAWNRR
QHZ00379.1	297	SETKCTLKSF	TVEKGIYQTS	NFRVQPTESI	VRFPNITNLC	PFGEVFNATR	FASVYAWNRR
QHR84449.1	297	SETKCTLKSF	TVEKGIYQTS	NFRVQPTESI	VRFPNITNLC	PFGEVFNATR	FASVYAWNRR
QIJ96493.1	297	SETKCTLKSF	TVEKGIYQTS	NFRVQPTESI	VRFPNITNLC	PFGEVFNATR	FASVYAWNRR
QIA98583.1	297	SETKCTLKSF	TVEKGIYQTS	NFRVQPTESI	VRFPNITNLC	PFGEVFNATR	FASVYAWNRR
QII57278.1	297	SETKCTLKSF	TVEKGIYQTS	NFRVQPTESI	VRFPNITNLC	PFGEVFNATR	FASVYAWNRR
QIA20044.1	297	SETKCTLKSF	TVEKGIYQTS	NFRVQPTESI	VRFPNITNLC	PFGEVFNATR	FASVYAWNRR
QIS60546.1	297	SETKCTLKSF	TVEKGIYQTS	NFRVQPTESI	VRFPNITNLC	PFGEVFNATR	FASVYAWNRR
QIS60582.1	297	SETKCTLKSF	TVEKGIYQTS	NFRVQPTESI	VRFPNITNLC	PFGEVFNATR	FASVYAWNRR
QIS61254.1	297	SETKCTLKSF	TVEKGIYQTS	NFRVQPTESI	VRFPNITNLC	PFGEVFNATR	FASVYAWNRR
QIS61422.1	297	SETKCTLKSF	TVEKGIYQTS	NFRVQPTESI	VRFPNITNLC	PFGEVFNATR	FASVYAWNRR
QIS60978.1	297	SETKCTLKSF	TVEKGIYQTS	NFRVQPTESI	VRFPNITNLC	PFGEVFNATR	FASVYAWNRR
QIO04367.1	297	SETKCTLKSF	TVEKGIYQTS	NFRVQPTESI	VRFPNITNLC	PFGEVFNATR	FASVYAWNRR
QIS30295.1	297	SETKCTLKSF	TVEKGIYQTS	NFRVQPTESI	VRFPNITNLC	PFGEVFNATR	FASVYAWNRR
QIS30615.1	297	SETKCTLKSF	TVEKGIYQTS	NFRVQPTESI	VRFPNITNLC	PFGEVFNATR	FASVYAWNRR
QIK50427.1	297	SETKCTLKSF	TVEKGIYQTS	NFRVQPTESI	VRFPNITNLC	PFGEVFNATR	FASVYAWNRR
QIS30165.1	297	SETKCTLKSF	TVEKGIYQTS	NFRVQPTESI	VRFPNITNLC	PFGEVFNATR	FASVYAWNRR
SARS-CoV	284	AELKCSVKSF	EIDKGIYQTS	NFRVBSGDV	VRFPNITNLC	PFGEVFNATR	FASVYAWERK

NTD

RBD

YP_009724390	357	RISNCVADYS	VLNSASFST	FKCYGVSPTK	LNDLCFTNVY	ADSFVIRGDE	VRQIAPGQTG
QHS34546.1	356	RISNCVADYS	VLNSASFST	FKCYGVSPTK	LNDLCFTNVY	ADSFVIRGDE	VRQIAPGQTG
QIS60906.1	357	RISNCVADYS	VLNSASFST	FKCYGVSPTK	LNDLCFTNVY	ADSFVIRGDE	VRQIAPGQTG
QIS60930.1	357	RISNCVADYS	VLNSASFST	FKCYGVSPTK	LNDLCFTNVY	ADSFVIRGDE	VRQIAPGQTG
QIS61338.1	357	RISNCVADYS	VLNSASFST	FKCYGVSPTK	LNDLCFTNVY	ADSFVIRGDE	VRQIAPGQTG
QIS60489.1	357	RISNCVADYS	VLNSASFST	FKCYGVSPTK	LNDLCFTNVY	ADSFVIRGDE	VRQIAPGQTG
QIS30625.1	357	RISNCVADYS	VLNSASFST	FKCYGVSPTK	LNDLCFTNVY	ADSFVIRGDE	VRQIAPGQTG
QIS30335.1	357	RISNCVADYS	VLNSASFST	FKCYGVSPTK	LNDLCFTNVY	ADSFVIRGDE	VRQIAPGQTG
QIC53204.1	357	RISNCVADYS	VLNSASFST	FKCYGVSPTK	LNDLCFTNVY	ADSFVIRGDE	VRQIAPGQTG
QHW06059.1	357	RISNCVADYS	VLNSASFST	FKCYGVSPTK	LNDLCFTNVY	ADSFVIRGDE	VRQIAPGQTG
QHZ00379.1	357	RISNCVADYS	VLNSASFST	FKCYGVSPTK	LNDLCFTNVY	ADSFVIRGDE	VRQIAPGQTG
QHR84449.1	357	RISNCVADYS	VLNSASFST	FKCYGVSPTK	LNDLCFTNVY	ADSFVIRGDE	VRQIAPGQTG
QIJ96493.1	357	RISNCVADYS	VLNSASFST	FKCYGVSPTK	LNDLCFTNVY	ADSFVIRGDE	VRQIAPGQTG
QIA98583.1	357	RISNCVADYS	VLNSASFST	FKCYGVSPTK	LNDLCFTNVY	ADSFVIRGDE	VRQIAPGQTG
QII57278.1	357	RISNCVADYS	VLNSASFST	FKCYGVSPTK	LNDLCFTNVY	ADSFVIRGDE	VRQIAPGQTG
QIA20044.1	357	RISNCVADYS	VLNSASFST	FKCYGVSPTK	LNDLCFTNVY	ADSFVIRGDE	VRQIAPGQTG
QIS60546.1	357	RISNCVADYS	VLNSASFST	FKCYGVSPTK	LNDLCFTNVY	ADSFVIRGDE	VRQIAPGQTG
QIS60582.1	357	RISNCVADYS	VLNSASFST	FKCYGVSPTK	LNDLCFTNVY	ADSFVIRGDE	VRQIAPGQTG
QIS61254.1	357	RISNCVADYS	VLNSASFST	FKCYGVSPTK	LNDLCFTNVY	ADSFVIRGDE	VRQIAPGQTG
QIS61422.1	357	RISNCVADYS	VLNSASFST	FKCYGVSPTK	LNDLCFTNVY	ADSFVIRGDE	VRQIAPGQTG
QIS60978.1	357	RISNCVADYS	VLNSASFST	FKCYGVSPTK	LNDLCFTNVY	ADSFVIRGDE	VRQIAPGQTG
QIO04367.1	357	RISNCVADYS	VLNSASFST	FKCYGVSPTK	LNDLCFTNVY	ADSFVIRGDE	VRQIAPGQTG
QIS30295.1	357	RISNCVADYS	VLNSASFST	FKCYGVSPTK	LNDLCFTNVY	ADSFVIRGDE	VRQIAPGQTG
QIS30615.1	357	RISNCVADYS	VLNSASFST	FKCYGVSPTK	LNDLCFTNVY	ADSFVIRGDE	VRQIAPGQTG
QIK50427.1	357	RISNCVADYS	VLNSASFST	FKCYGVSPTK	LNDLCFTNVY	ADSFVIRGDE	VRQIAPGQTG
QIS30165.1	357	RISNCVADYS	VLNSASFST	FKCYGVSPTK	LNDLCFTNVY	ADSFVIRGDE	VRQIAPGQTG
SARS-CoV	344	RISNCVADYS	VLNSASFST	FKCYGVSPTK	LNDLCFTNVY	ADSFVIRGDE	VRQIAPGQTG

RBD

YP_009724390	417	KIADYNYKLP	DDFTGCVIAW	NSNNLDSKVG	GNYNLYRRLF	RKSNLKPFFER	DISTEIIYQAG
QHS34546.1	416	KIADYNYKLP	DDFTGCVIAW	NSNNLDSKVG	GNYNLYRRLF	RKSNLKPFFER	DISTEIIYQAG
QIS60906.1	417	KIADYNYKLP	DDFTGCVIAW	NSNNLDSKVG	GNYNLYRRLF	RKSNLKPFFER	DISTEIIYQAG
QIS60930.1	417	KIADYNYKLP	DDFTGCVIAW	NSNNLDSKVG	GNYNLYRRLF	RKSNLKPFFER	DISTEIIYQAG
QIS61338.1	417	KIADYNYKLP	DDFTGCVIAW	NSNNLDSKVG	GNYNLYRRLF	RKSNLKPFFER	DISTEIIYQAG
QIS60489.1	417	KIADYNYKLP	DDFTGCVIAW	NSNNLDSKVG	GNYNLYRRLF	RKSNLKPFFER	DISTEIIYQAG
QIS30625.1	417	KIADYNYKLP	DDFTGCVIAW	NSNNLDSKVG	GNYNLYRRLF	RKSNLKPFFER	DISTEIIYQAG
QIS30335.1	417	KIADYNYKLP	DDFTGCVIAW	NSNNLDSKVG	GNYNLYRRLF	RKSNLKPFFER	DISTEIIYQAG
QIC53204.1	417	KIADYNYKLP	DDFTGCVIAW	NSNNLDSKVG	GNYNLYRRLF	RKSNLKPFFER	DISTEIIYQAG
QHW06059.1	417	KIADYNYKLP	DDFTGCVIAW	NSNNLDSKVG	GNYNLYRRLF	RKSNLKPFFER	DISTEIIYQAG
QHZ00379.1	417	KIADYNYKLP	DDFTGCVIAW	NSNNLDSKVG	GNYNLYRRLF	RKSNLKPFFER	DISTEIIYQAG
QHR84449.1	417	KIADYNYKLP	DDFTGCVIAW	NSNNLDSKVG	GNYNLYRRLF	RKSNLKPFFER	DISTEIIYQAG
QIJ96493.1	417	KIADYNYKLP	DDFTGCVIAW	NSNNLDSKVG	GNYNLYRRLF	RKSNLKPFFER	DISTEIIYQAG
QIA98583.1	417	KIADYNYKLP	DDFTGCVIAW	NSNNLDSKVG	GNYNLYRRLF	RKSNLKPFFER	DISTEIIYQAG
QII57278.1	417	KIADYNYKLP	DDFTGCVIAW	NSNNLDSKVG	GNYNLYRRLF	RKSNLKPFFER	DISTEIIYQAG
QIA20044.1	417	KIADYNYKLP	DDFTGCVIAW	NSNNLDSKVG	GNYNLYRRLF	RKSNLKPFFER	DISTEIIYQAG
QIS60546.1	417	KIADYNYKLP	DDFTGCVIAW	NSNNLDSKVG	GNYNLYRRLF	RKSNLKPFFER	DISTEIIYQAG
QIS60582.1	417	KIADYNYKLP	DDFTGCVIAW	NSNNLDSKVG	GNYNLYRRLF	RKSNLKPFFER	DISTEIIYQAG
QIS61254.1	417	KIADYNYKLP	DDFTGCVIAW	NSNNLDSKVG	GNYNLYRRLF	RKSNLKPFFER	DISTEIIYQAG
QIS61422.1	417	KIADYNYKLP	DDFTGCVIAW	NSNNLDSKVG	GNYNLYRRLF	RKSNLKPFFER	DISTEIIYQAG
QIS60978.1	417	KIADYNYKLP	DDFTGCVIAW	NSNNLDSKVG	GNYNLYRRLF	RKSNLKPFFER	DISTEIIYQAG
QIO04367.1	417	KIADYNYKLP	DDFTGCVIAW	NSNNLDSKVG	GNYNLYRRLF	RKSNLKPFFER	DISTEIIYQAG
QIS30295.1	417	KIADYNYKLP	DDFTGCVIAW	NSNNLDSKVG	GNYNLYRRLF	RKSNLKPFFER	DISTEIIYQAG
QIS30615.1	417	KIADYNYKLP	DDFTGCVIAW	NSNNLDSKVG	GNYNLYRRLF	RKSNLKPFFER	DISTEIIYQAG
QIK50427.1	417	KIADYNYKLP	DDFTGCVIAW	NSNNLDSKVG	GNYNLYRRLF	RKSNLKPFFER	DISTEIIYQAG
QIS30165.1	417	KIADYNYKLP	DDFTGCVIAW	NSNNLDSKVG	GNYNLYRRLF	RKSNLKPFFER	DISTEIIYQAG
SARS-CoV	404	KIADYNYKLP	DDFTGCVIAW	NSNNLDSKVG	GNYNLYRRLF	RKSNLKPFFER	DISTEIIYQAG

RBD

YP_009724390	477	STPCNGVEGF	NCYFPLQSYG	FQPTNGVGYY	PYRVVLSFE	LLHAPATVCG	PKKSTNLVKN
QHS34546.1	476	STPCNGVEGF	NCYFPLQSYG	FQPTNGVGYY	PYRVVLSFE	LLHAPATVCG	PKKSTNLVKN
QIS60906.1	477	STPCNGVEGF	NCYFPLQSYG	FQPTNGVGYY	PYRVVLSFE	LLHAPATVCG	PKKSTNLVKN
QIS60930.1	477	STPCNGVEGF	NCYFPLQSYG	FQPTNGVGYY	PYRVVLSFE	LLHAPATVCG	PKKSTNLVKN
QIS61338.1	477	STPCNGVEGF	NCYFPLQSYG	FQPTNGVGYY	PYRVVLSFE	LLHAPATVCG	PKKSTNLVKN
QIS60489.1	477	STPCNGVEGF	NCYFPLQSYG	FQPTNGVGYY	PYRVVLSFE	LLHAPATVCG	PKKSTNLVKN
QIS30625.1	477	STPCNGVEGF	NCYFPLQSYG	FQPTNGVGYY	PYRVVLSFE	LLHAPATVCG	PKKSTNLVKN
QIS30335.1	477	STPCNGVEGF	NCYFPLQSYG	FQPTNGVGYY	PYRVVLSFE	LLHAPATVCG	PKKSTNLVKN
QIC53204.1	477	STPCNGVEGF	NCYFPLQSYG	FQPTNGVGYY	PYRVVLSFE	LLHAPATVCG	PKKSTNLVKN
QHW06059.1	477	STPCNGVEGF	NCYFPLQSYG	FQPTNGVGYY	PYRVVLSFE	LLHAPATVCG	PKKSTNLVKN
QHZ00379.1	477	STPCNGVEGF	NCYFPLQSYG	FQPTNGVGYY	PYRVVLSFE	LLHAPATVCG	PKKSTNLVKN
QHR84449.1	477	STPCNGVEGF	NCYFPLQSYG	FQPTNGVGYY	PYRVVLSFE	LLHAPATVCG	PKKSTNLVKN
QIJ96493.1	477	STPCNGVEGF	NCYFPLQSYG	FQPTNGVGYY	PYRVVLSFE	LLHAPATVCG	PKKSTNLVKN
QIA98583.1	477	STPCNGVEGF	NCYFPLQSYG	FQPTNGVGYY	PYRVVLSFE	LLHAPATVCG	PKKSTNLVKN
QII57278.1	477	STPCNGVEGF	NCYFPLQSYG	FQPTNGVGYY	PYRVVLSFE	LLHAPATVCG	PKKSTNLVKN
QIA20044.1	477	STPCNGVEGF	NCYFPLQSYG	FQPTNGVGYY	PYRVVLSFE	LLHAPATVCG	PKKSTNLVKN
QIS60546.1	477	STPCNGVEGF	NCYFPLQSYG	FQPTNGVGYY	PYRVVLSFE	LLHAPATVCG	PKKSTNLVKN
QIS60582.1	477	STPCNGVEGF	NCYFPLQSYG	FQPTNGVGYY	PYRVVLSFE	LLHAPATVCG	PKKSTNLVKN
QIS61254.1	477	STPCNGVEGF	NCYFPLQSYG	FQPTNGVGYY	PYRVVLSFE	LLHAPATVCG	PKKSTNLVKN
QIS61422.1	477	STPCNGVEGF	NCYFPLQSYG	FQPTNGVGYY	PYRVVLSFE	LLHAPATVCG	PKKSTNLVKN
QIS60978.1	477	STPCNGVEGF	NCYFPLQSYG	FQPTNGVGYY	PYRVVLSFE	LLHAPATVCG	PKKSTNLVKN
QIO04367.1	477	STPCNGVEGF	NCYFPLQSYG	FQPTNGVGYY	PYRVVLSFE	LLHAPATVCG	PKKSTNLVKN
QIS30295.1	477	STPCNGVEGF	NCYFPLQSYG	FQPTNGVGYY	PYRVVLSFE	LLHAPATVCG	PKKSTNLVKN
QIS30615.1	477	STPCNGVEGF	NCYFPLQSYG	FQPTNGVGYY	PYRVVLSFE	LLHAPATVCG	PKKSTNLVKN
QIK50427.1	477	STPCNGVEGF	NCYFPLQSYG	FQPTNGVGYY	PYRVVLSFE	LLHAPATVCG	PKKSTNLVKN
QIS30165.1	477	STPCNGVEGF	NCYFPLQSYG	FQPTNGVGYY	PYRVVLSFE	LLHAPATVCG	PKKSTNLVKN
SARS-CoV	464	GKPCCT-PPAL	NCYFPLQSYG	FQPTNGVGYY	PYRVVLSFE	LLHAPATVCG	PKKSTNLVKN

RBD

SD1

YP_009724390	537	KCVNFNFNGL	TGTGVLTESN	KKFLPFQQFG	RDIADTTDAV	RDPQTLEILD	ITPCSFGGVS
QHS34546.1	536	KCVNFNFNGL	TGTGVLTESN	KKFLPFQQFG	RDIADTTDAV	RDPQTLEILD	ITPCSFGGVS
QIS60906.1	537	KCVNFNFNGL	TGTGVLTESN	KKFLPFQQFG	RDIADTTDAV	RDPQTLEILD	ITPCSFGGVS
QIS60930.1	537	KCVNFNFNGL	TGTGVLTESN	KKFLPFQQFG	RDIADTTDAV	RDPQTLEILD	ITPCSFGGVS
QIS61338.1	537	KCVNFNFNGL	TGTGVLTESN	KKFLPFQQFG	RDIADTTDAV	RDPQTLEILD	ITPCSFGGVS
QIS60489.1	537	KCVNFNFNGL	TGTGVLTESN	KKFLPFQQFG	RDIADTTDAV	RDPQTLEILD	ITPCSFGGVS
QIS30625.1	537	KCVNFNFNGL	TGTGVLTESN	KKFLPFQQFG	RDIADTTDAV	RDPQTLEILD	ITPCSFGGVS
QIS30335.1	537	KCVNFNFNGL	TGTGVLTESN	KKFLPFQQFG	RDIADTTDAV	RDPQTLEILD	ITPCSFGGVS
QIC53204.1	537	KCVNFNFNGL	TGTGVLTESN	KKFLPFQQFG	RDIADTTDAV	RDPQTLEILD	ITPCSFGGVS
QHW06059.1	537	KCVNFNFNGL	TGTGVLTESN	KKFLPFQQFG	RDIADTTDAV	RDPQTLEILD	ITPCSFGGVS
QHZ00379.1	537	KCVNFNFNGL	TGTGVLTESN	KKFLPFQQFG	RDIADTTDAV	RDPQTLEILD	ITPCSFGGVS
QHR84449.1	537	KCVNFNFNGL	TGTGVLTESN	KKFLPFQQFG	RDIADTTDAV	RDPQTLEILD	ITPCSFGGVS
QIJ96493.1	537	KCVNFNFNGL	TGTGVLTESN	KKFLPFQQFG	RDIADTTDAV	RDPQTLEILD	ITPCSFGGVS
QIA98583.1	537	KCVNFNFNGL	TGTGVLTESN	KKFLPFQQFG	RDIADTTDAV	RDPQTLEILD	ITPCSFGGVS
QII57278.1	537	KCVNFNFNGL	TGTGVLTESN	KKFLPFQQFG	RDIADTTDAV	RDPQTLEILD	ITPCSFGGVS
QIA20044.1	537	KCVNFNFNGL	TGTGVLTESN	KKFLPFQQFG	RDIADTTDAV	RDPQTLEILD	ITPCSFGGVS
QIS60546.1	537	KCVNFNFNGL	TGTGVLTESN	KKFLPFQQFG	RDIADTTDAV	RDPQTLEILD	ITPCSFGGVS
QIS60582.1	537	KCVNFNFNGL	TGTGVLTESN	KKFLPFQQFG	RDIADTTDAV	RDPQTLEILD	ITPCSFGGVS
QIS61254.1	537	KCVNFNFNGL	TGTGVLTESN	KKFLPFQQFG	RDIADTTDAV	RDPQTLEILD	ITPCSFGGVS
QIS61422.1	537	KCVNFNFNGL	TGTGVLTESN	KKFLPFQQFG	RDIADTTDAV	RDPQTLEILD	ITPCSFGGVS
QIS60978.1	537	KCVNFNFNGL	TGTGVLTESN	KKFLPFQQFG	RDIADTTDAV	RDPQTLEILD	ITPCSFGGVS
QIO04367.1	537	KCVNFNFNGL	TGTGVLTESN	KKFLPFQQFG	RDIADTTDAV	RDPQTLEILD	ITPCSFGGVS
QIS30295.1	537	KCVNFNFNGL	TGTGVLTESN	KKFLPFQQFG	RDIADTTDAV	RDPQTLEILD	ITPCSFGGVS
QIS30615.1	537	KCVNFNFNGL	TGTGVLTESN	KKFLPFQQFG	RDIADTTDAV	RDPQTLEILD	ITPCSFGGVS
QIK50427.1	537	KCVNFNFNGL	TGTGVLTESN	KKFLPFQQFG	RDIADTTDAV	RDPQTLEILD	ITPCSFGGVS
QIS30165.1	537	KCVNFNFNGL	TGTGVLTESN	KKFLPFQQFG	RDIADTTDAV	RDPQTLEILD	ITPCSFGGVS
SARS-CoV	523	KCVNFNFNGL	TGTGVLTESN	KKFLPFQQFG	RDIADTTDAV	RDPQTLEILD	ITPCSFGGVS

SD1

SD2

YP_009724390	597	VITPGTNTSN	QVAVLYQDVN	CTEVPVAIHA	DQLTPTWRVY	STGSNVFQTR	AGCLIGAEHV
QHS34546.1	596	VITPGTNTSN	QVAVLYQDVN	CTEVPVAIHA	DQLTPTWRVY	STGSNVFQTR	AGCLIGAEHV
QIS60906.1	597	VITPGTNTSN	QVAVLYQDVN	CTEVPVAIHA	DQLTPTWRVY	STGSNVFQTR	AGCLIGAEHV
QIS60930.1	597	VITPGTNTSN	QVAVLYQDVN	CTEVPVAIHA	DQLTPTWRVY	STGSNVFQTR	AGCLIGAEHV
QIS61338.1	597	VITPGTNTSN	QVAVLYQDVN	CTEVPVAIHA	DQLTPTWRVY	STGSNVFQTR	AGCLIGAEHV
QIS60489.1	597	VITPGTNTSN	QVAVLYQDVN	CTEVPVAIHA	DQLTPTWRVY	STGSNVFQTR	AGCLIGAEHV
QIS30625.1	597	VITPGTNTSN	QVAVLYQDVN	CTEVPVAIHA	DQLTPTWRVY	STGSNVFQTR	AGCLIGAEHV
QIS30335.1	597	VITPGTNTSN	QVAVLYQDVN	CTEVPVAIHA	DQLTPTWRVY	STGSNVFQTR	AGCLIGAEHV
QIC53204.1	597	VITPGTNTSN	QVAVLYQDVN	CTEVPVAIHA	DQLTPTWRVY	STGSNVFQTR	AGCLIGAEHV
QHW06059.1	597	VITPGTNTSN	QVAVLYQDVN	CTEVPVAIHA	DQLTPTWRVY	STGSNVFQTR	AGCLIGAEHV
QHZ00379.1	597	VITPGTNTSN	QVAVLYQDVN	CTEVPVAIHA	DQLTPTWRVY	STGSNVFQTR	AGCLIGAEHV
QHR84449.1	597	VITPGTNTSN	QVAVLYQDVN	CTEVPVAIHA	DQLTPTWRVY	STGSNVFQTR	AGCLIGAEHV
QIJ96493.1	597	VITPGTNTSN	QVAVLYQDVN	CTEVPVAIHA	DQLTPTWRVY	STGSNVFQTR	AGCLIGAEHV
QIA98583.1	597	VITPGTNTSN	QVAVLYQDVN	CTEVPVAIHA	DQLTPTWRVY	STGSNVFQTR	AGCLIGAEHV
QII57278.1	597	VITPGTNTSN	QVAVLYQDVN	CTEVPVAIHA	DQLTPTWRVY	STGSNVFQTR	AGCLIGAEHV
QIA20044.1	597	VITPGTNTSN	QVAVLYQDVN	CTEVPVAIHA	DQLTPTWRVY	STGSNVFQTR	AGCLIGAEHV
QIS60546.1	597	VITPGTNTSN	QVAVLYQDVN	CTEVPVAIHA	DQLTPTWRVY	STGSNVFQTR	AGCLIGAEHV
QIS60582.1	597	VITPGTNTSN	QVAVLYQDVN	CTEVPVAIHA	DQLTPTWRVY	STGSNVFQTR	AGCLIGAEHV
QIS61254.1	597	VITPGTNTSN	QVAVLYQDVN	CTEVPVAIHA	DQLTPTWRVY	STGSNVFQTR	AGCLIGAEHV
QIS61422.1	597	VITPGTNTSN	QVAVLYQDVN	CTEVPVAIHA	DQLTPTWRVY	STGSNVFQTR	AGCLIGAEHV
QIS60978.1	597	VITPGTNTSN	QVAVLYQDVN	CTEVPVAIHA	DQLTPTWRVY	STGSNVFQTR	AGCLIGAEHV
QIO04367.1	597	VITPGTNTSN	QVAVLYQDVN	CTEVPVAIHA	DQLTPTWRVY	STGSNVFQTR	AGCLIGAEHV
QIS30295.1	597	VITPGTNTSN	QVAVLYQGVN	CTEVPVAIHA	DQLTPTWRVY	STGSNVFQTR	AGCLIGAEHV
QIS30615.1	597	VITPGTNTSN	QVAVLYQGVN	CTEVPVAIHA	DQLTPTWRVY	STGSNVFQTR	AGCLIGAEHV
QIK50427.1	597	VITPGTNTSN	QVAVLYQGVN	CTEVPVAIHA	DQLTPTWRVY	STGSNVFQTR	AGCLIGAEHV
QIS30165.1	597	VITPGTNTSN	QVAVLYQDVN	CTEVPVAIHA	DQLTPTWRVY	STGSNVFQTR	AGCLIGAEHV
SARS-CoV	583	VITPGTNTSS	EVAVLYQDVN	CTEVPVAIHA	DQLTPTWRVY	STGSNVFQTR	AGCLIGAEHV

SD2

YP_009724390	657	NNSYECDIPI	GAGICASYQT	QTNSPRRARS	VASQSIIAYT	MSLGAENSVA	YSNNSIAIPT
QHS34546.1	656	NNSYECDIPI	GAGICASYQT	QTNSPRRARS	VASQSIIAYT	MSLGAENSVA	YSNNSIAIPT
QIS60906.1	657	NNSYECDIPI	GAGICASYQT	QTNSPRRARS	VASQSIIAYT	MSLGAENSVA	YSNNSIAIPT
QIS60930.1	657	NNSYECDIPI	GAGICASYQT	QTNSPRRARS	VASQSIIAYT	MSLGAENSVA	YSNNSIAIPT
QIS61338.1	657	NNSYECDIPI	GAGICASYQT	QTNSPRRARS	VASQSIIAYT	MSLGAENSVA	YSNNSIAIPT
QIS60489.1	657	NNSYECDIPI	GAGICASYQT	QTNSPRRARS	VASQSIIAYT	MSLGAENSVA	YSNNSIAIPT
QIS30625.1	657	NNSYECDIPI	GAGICASYQT	QTNSPRRARS	VASQSIIAYT	MSLGAENSVA	YSNNSIAIPT
QIS30335.1	657	NNSYECDIPI	GAGICASYQT	QTNSPRRARS	VASQSIIAYT	MSLGAENSVA	YSNNSIAIPT
QIC53204.1	657	NNSYECDIPI	GAGICASYQT	QTNSPRRARS	VASQSIIAYT	MSLGAENSVA	YSNNSIAIPT
QHW06059.1	657	NNSYECDIPI	GAGICASYQT	QTNSPRRARS	VASQSIIAYT	MSLGAENSVA	YSNNSIAIPT
QHZ00379.1	657	NNSYECDIPI	GAGICASYQT	QTNSPRRARS	VASQSIIAYT	MSLGAENSVA	YSNNSIAIPT
QHR84449.1	657	NNSYECDIPI	GAGICASYQT	QTNSPRRARS	VASQSIIAYT	MSLGAENSVA	YSNNSIAIPT
QIJ96493.1	657	NNSYECDIPI	GAGICASYQT	QTNSPRRARS	VASQSIIAYT	MSLGAENSVA	YSNNSIAIPT
QIA98583.1	657	NNSYECDIPI	GAGICASYQT	QTNSPRRARS	VASQSIIAYT	MSLGAENSVA	YSNNSIAIPT
QII57278.1	657	NNSYECDIPI	GAGICASYQT	QTNSPRRARS	VASQSIIAYT	MSLGAENSVA	YSNNSIAIPT
QIA20044.1	657	NNSYECDIPI	GAGICASYQT	QTNSPRRARS	VASQSIIAYT	MSLGAENSVA	YSNNSIAIPT
QIS60546.1	657	NNSYECDIPI	GAGICASYQT	QTNSPRRARS	VASQSIIAYT	MSLGAENSVA	YSNNSIAIPT
QIS60582.1	657	NNSYECDIPI	GAGICASYQT	QTNSPRRARS	VASQSIIAYT	MSLGAENSVA	YSNNSIAIPT
QIS61254.1	657	NNSYECDIPI	GAGICASYQT	QTNSPRRARS	VASQSIIAYT	MSLGAENSVA	YSNNSIAIPT
QIS61422.1	657	NNSYECDIPI	GAGICASYQT	QTNSPRRARS	VASQSIIAYT	MSLGAENSVA	YSNNSIAIPT
QIS60978.1	657	NNSYECDIPI	GAGICASYQT	QTNSPRRARS	VASQSIIAYT	MSLGAENSVA	YSNNSIAIPT
QIO04367.1	657	NNSYECDIPI	GAGICASYQT	QTNSPRRARS	VASQSIIAYT	MSLGAENSVA	YSNNSIAIPT
QIS30295.1	657	NNSYECDIPI	GAGICASYQT	QTNSPRRARS	VASQSIIAYT	MSLGAENSVA	YSNNSIAIPT
QIS30615.1	657	NNSYECDIPI	GAGICASYQT	QTNSPRRARS	VASQSIIAYT	MSLGAENSVA	YSNNSIAIPT
QIK50427.1	657	NNSYECDIPI	GAGICASYQT	QTNSPRRARS	VASQSIIAYT	MSLGAENSVA	YSNNSIAIPT
QIS30165.1	657	NNSYECDIPI	GAGICASYQT	QTNSPRRARS	VASQSIIAYT	MSLGAENSVA	YSNNSIAIPT
SARS-CoV	643	DTSYECDIPI	GAGICASYHT	VS----LLRS	TSQRSIVAYT	MSLGADSSIA	YSNNTIAIPT

SD2

S1/S2

YP_009724390	717	NFTISVTTEI	LPVSMTKTSV	DCTMYICGDS	TECSNLLLQY	GSFCTQLNRA	LTGIAVEQDK
QHS34546.1	716	NFTISVTTEI	LPVSMTKTSV	DCTMYICGDS	TECSNLLLQY	GSFCTQLNRA	LTGIAVEQDK
QIS60906.1	717	NFTISVTTEI	LPVSMTKTSV	DCTMYICGDS	TECSNLLLQY	GSFCTQLNRA	LTGIAVEQDK
QIS60930.1	717	NFTISVTTEI	LPVSMTKTSV	DCTMYICGDS	TECSNLLLQY	GSFCTQLNRA	LTGIAVEQDK
QIS61338.1	717	NFTISVTTEI	LPVSMTKTSV	DCTMYICGDS	TECSNLLLQY	GSFCTQLNRA	LTGIAVEQDK
QIS60489.1	717	NFTISVTTEI	LPVSMTKTSV	DCTMYICGDS	TECSNLLLQY	GSFCTQLNRA	LTGIAVEQDK
QIS30625.1	717	NFTISVTTEI	LPVSMTKTSV	DCTMYICGDS	TECSNLLLQY	GSFCTQLNRA	LTGIAVEQDK
QIS30335.1	717	NFTISVTTEI	LPVSMTKTSV	DCTMYICGDS	TECSNLLLQY	GSFCTQLNRA	LTGIAVEQDK
QIC53204.1	717	NFTISVTTEI	LPVSMTKTSV	DCTMYICGDS	TECSNLLLQY	GSFCTQLNRA	LTGIAVEQDK
QHW06059.1	717	NFTISVTTEI	LPVSMTKTSV	DCTMYICGDS	TECSNLLLQY	GSFCTQLNRA	LTGIAVEQDK
QHZ00379.1	717	NFTISVTTEI	LPVSMTKTSV	DCTMYICGDS	TECSNLLLQY	GSFCTQLNRA	LTGIAVEQDK
QHR84449.1	717	NFTISVTTEI	LPVSMTKTSV	DCTMYICGDS	TECSNLLLQY	GSFCTQLNRA	LTGIAVEQDK
QIJ96493.1	717	NFTISVTTEI	LPVSMTKTSV	DCTMYICGDS	TECSNLLLQY	GSFCTQLNRA	LTGIAVEQDK
QIA98583.1	717	NFTISVTTEI	LPVSMTKTSV	DCTMYICGDS	TECSNLLLQY	GSFCTQLNRA	LTGIAVEQDK
QII57278.1	717	NFTISVTTEI	LPVSMTKTSV	DCTMYICGDS	TECSNLLLQY	GSFCTQLNRA	LTGIAVEQDK
QIA20044.1	717	NFTISVTTEI	LPVSMTKTSV	DCTMYICGDS	TECSNLLLQY	GSFCTQLNRA	LTGIAVEQDK
QIS60546.1	717	NFTISVTTEI	LPVSMTKTSV	DCTMYICGDS	TECSNLLLQY	GSFCTQLNRA	LTGIAVEQDK
QIS60582.1	717	NFTISVTTEI	LPVSMTKTSV	DCTMYICGDS	TECSNLLLQY	GSFCTQLNRA	LTGIAVEQDK
QIS61254.1	717	NFTISVTTEI	LPVSMTKTSV	DCTMYICGDS	TECSNLLLQY	GSFCTQLNRA	LTGIAVEQDK
QIS61422.1	717	NFTISVTTEI	LPVSMTKTSV	DCTMYICGDS	TECSNLLLQY	GSFCTQLNRA	LTGIAVEQDK
QIS60978.1	717	NFTISVTTEI	LPVSMTKTSV	DCTMYICGDS	TECSNLLLQY	GSFCTQLNRA	LTGIAVEQDK
QIO04367.1	717	NFTISVTTEI	LPVSMTKTSV	DCTMYICGDS	TECSNLLLQY	GSFCTQLNRA	LTGIAVEQDK
QIS30295.1	717	NFTISVTTEI	LPVSMTKTSV	DCTMYICGDS	TECSNLLLQY	GSFCTQLNRA	LTGIAVEQDK
QIS30615.1	717	NFTISVTTEI	LPVSMTKTSV	DCTMYICGDS	TECSNLLLQY	GSFCTQLNRA	LTGIAVEQDK
QIK50427.1	717	NFTISVTTEI	LPVSMTKTSV	DCTMYICGDS	TECSNLLLQY	GSFCTQLNRA	LTGIAVEQDK
QIS30165.1	717	NFTISVTTEI	LPVSMTKTSV	DCTMYICGDS	TECSNLLLQY	GSFCTQLNRA	LTGIAVEQDK
SARS-CoV	699	NFTISVTTEI	LPVSMTKTSV	DCTMYICGDS	TECSNLLLQY	GSFCTQLNRA	LTGIAVEQDK

YP_009724390	777	NTQEVFAQVK	QIYKTPPIKD	FGGFNFSQIL	PDPSKPSKRS	FIEDLLFNKV	TLADAGFIKQ
QHS34546.1	776	NTQEVFAQVK	QIYKTPPIKD	FGGFNFSQIL	PDPSKPSKRS	FIEDLLFNKV	TLADAGFIKQ
QIS60906.1	777	NTQEVFAQVK	QIYKTPPIKD	FGGFNFSQIL	PDPSKPSKRS	FIEDLLFNKV	TLADAGFIKQ
QIS60930.1	777	NTQEVFAQVK	QIYKTPPIKD	FGGFNFSQIL	PDPSKPSKRS	FIEDLLFNKV	TLADAGFIKQ
QIS61338.1	777	NTQEVFAQVK	QIYKTPPIKD	FGGFNFSQIL	PDPSKPSKRS	FIEDLLFNKV	TLADAGFIKQ
QIS60489.1	777	NTQEVFAQVK	QIYKTPPIKD	FGGFNFSQIL	PDPSKPSKRS	FIEDLLFNKV	TLADAGFIKQ
QIS30625.1	777	NTQEVFAQVK	QIYKTPPIKD	FGGFNFSQIL	PDPSKPSKRS	FIEDLLFNKV	TLADAGFIKQ
QIS30335.1	777	NTQEVFAQVK	QIYKTPPIKD	FGGFNFSQIL	PDPSKPSKRS	FIEDLLFNKV	TLADAGFIKQ
QIC53204.1	777	NTQEVFAQVK	QIYKTPPIKD	FGGFNFSQIL	PDPSKPSKRS	FIEDLLFNKV	TLADAGFIKQ
QHW06059.1	777	NTQEVFAQVK	QIYKTPPIKD	FGGFNFSQIL	PDPSKPSKRS	FIEDLLFNKV	TLADAGFIKQ
QHZ00379.1	777	NTQEVFAQVK	QIYKTPPIKD	FGGFNFSQIL	PDPSKPSKRS	FIEDLLFNKV	TLADAGFIKQ
QHR84449.1	777	NTQEVFAQVK	QIYKTPPIKD	FGGFNFSQIL	PDPSKPSKRS	FIEDLLFNKV	TLADAGFIKQ
QIJ96493.1	777	NTQEVFAQVK	QIYKTPPIKD	FGGFNFSQIL	PDPSKPSKRS	FIEDLLFNKV	TLADAGFIKQ
QIA98583.1	777	NTQEVFAQVK	QIYKTPPIKD	FGGFNFSQIL	PDPSKPSKRS	FIEDLLFNKV	TLADAGFIKQ
QII57278.1	777	NTQEVFAQVK	QIYKTPPIKD	FGGFNFSQIL	PDPSKPSKRS	FIEDLLFNKV	TLADAGFIKQ
QIA20044.1	777	NTQEVFAQVK	QIYKTPPIKD	FGGFNFSQIL	PDPSKPSKRS	FIEDLLFNKV	TLADAGFIKQ
QIS60546.1	777	NTQEVFAQVK	QIYKTPPIKD	FGGFNFSQIL	PDPSKPSKRS	FIEDLLFNKV	TLADAGFIKQ
QIS60582.1	777	NTQEVFAQVK	QIYKTPPIKD	FGGFNFSQIL	PDPSKPSKRS	FIEDLLFNKV	TLADAGFIKQ
QIS61254.1	777	NTQEVFAQVK	QIYKTPPIKD	FGGFNFSQIL	PDPSKPSKRS	FIEDLLFNKV	TLADAGFIKQ
QIS61422.1	777	NTQEVFAQVK	QIYKTPPIKD	FGGFNFSQIL	PDPSKPSKRS	FIEDLLFNKV	TLADAGFIKQ
QIS60978.1	777	NTQEVFAQVK	QIYKTPPIKD	FGGFNFSQIL	PDPSKPSKRS	FIEDLLFNKV	TLADAGFIKQ
QIO04367.1	777	NTQEVFAQVK	QIYKTPPIKD	FGGFNFSQIL	PDPSKPSKRS	FIEDLLFNKV	TLADAGFIKQ
QIS30295.1	777	NTQEVFAQVK	QIYKTPPIKD	FGGFNFSQIL	PDPSKPSKRS	FIEDLLFNKV	TLADAGFIKQ
QIS30615.1	777	NTQEVFAQVK	QIYKTPPIKD	FGGFNFSQIL	PDPSKPSKRS	FIEDLLFNKV	TLADAGFIKQ
QIK50427.1	777	NTQEVFAQVK	QIYKTPPIKD	FGGFNFSQIL	PDPSKPSKRS	FIEDLLFNKV	TLADAGFIKQ
QIS30165.1	777	NTQEVFAQVK	QIYKTPPIKD	FGGFNFSQIL	PDPSKPSKRS	FIEDLLFNKV	TLADAGFIKQ
SARS-CoV	759	NTQEVFAQVK	QIYKTPPIKD	FGGFNFSQIL	PDPSKPSKRS	FIEDLLFNKV	TLADAGFIKQ

S2' FP

YP_009724390	837	YGDCLGDIAA	RDLICAQKFN	GLTVLPPLLT	DEMIAQY TSA	LLAGTITSGW	TFGAGAALQI
QHS34546.1	836	YGDCLGDIAA	RDLICAQKFN	GLTVLPPLLT	DEMIAQY TSA	LLAGTITSGW	TFGAGAALQI
QIS60906.1	837	YGDCLGDIAA	RDLICAQKFN	GLTVLPPLLT	DEMIAQY TSA	LLAGTITSGW	TFGAGAALQI
QIS60930.1	837	YGDCLGDIAA	RDLICAQKFN	GLTVLPPLLT	DEMIAQY TSA	LLAGTITSGW	TFGAGAALQI

QIS61338.1	837	YGDCLGDIAA	RDLICAQKFN	GLTVLPPLLT	DEMIAQY TSA	LLAGTITSGW	TFGAGAALQI
QIS60489.1	837	YGDCLGDIAA	RDLICAQKFN	GLTVLPPLLT	DEMIAQY TSA	LLAGTITSGW	TFGAGAALQI
QIS30625.1	837	YGDCLGDIAA	RDLICAQKFN	GLTVLPPLLT	DEMIAQY TSA	LLAGTITSGW	TFGAGAALQI
QIS30335.1	837	YGDCLGDIAA	RDLICAQKFN	GLTVLPPLLT	DEMIAQY TSA	LLAGTITSGW	TFGAGAALQI
QIC53204.1	837	YGDCLGDIAA	RDLICAQKFN	GLTVLPPLLT	DEMIAQY TSA	LLAGTITSGW	TFGAGAALQI
QHW06059.1	837	YGDCLGDIAA	RDLICAQKFN	GLTVLPPLLT	DEMIAQY TSA	LLAGTITSGW	TFGAGAALQI
QHZ00379.1	837	YGDCLGDIAA	RDLICAQKFN	GLTVLPPLLT	DEMIAQY TSA	LLAGTITSGW	TFGAGAALQI
QHR84449.1	837	YGDCLGDIAA	RDLICAQKFN	GLTVLPPLLT	DEMIAQY TSA	LLAGTITSGW	TFGAGAALQI
QIJ96493.1	837	YGDCLGDIAA	RDLICAQKFN	GLTVLPPLLT	DEMIAQY TSA	LLAGTITSGW	TFGAGAALQI
QIA98583.1	837	YGDCLGDIAA	RDLICAQKFN	GLTVLPPLLT	DEMIAQY TSA	LLAGTITSGW	TFGAGAALQI
QII57278.1	837	YGDCLGDIAA	RDLICAQKFN	GLTVLPPLLT	DEMIAQY TSA	LLAGTITSGW	TFGAGAALQI
QIA20044.1	837	YGDCLGDIAA	RDLICAQKFN	GLTVLPPLLT	DEMIAQY TSA	LLAGTITSGW	TFGAGAALQI
QIS60546.1	837	YGDCLGDIAA	RDLICAQKFN	GLTVLPPLLT	DEMIAQY TSA	LLAGTITSGW	TFGAGAALQI
QIS60582.1	837	YGDCLGDIAA	RDLICAQKFN	GLTVLPPLLT	DEMIAQY TSA	LLAGTITSGW	TFGAGAALQI
QIS61254.1	837	YGDCLGDIAA	RDLICAQKFN	GLTVLPPLLT	DEMIAQY TSA	LLAGTITSGW	TFGAGAALQI
QIS61422.1	837	YGDCLGDIAA	RDLICAQKFN	GLTVLPPLLT	DEMIAQY TSA	LLAGTITSGW	TFGAGAALQI
QIS60978.1	837	YGDCLGDIAA	RDLICAQKFN	GLTVLPPLLT	DEMIAQY TSA	LLAGTITSGW	TFGAGAALQI
QIO04367.1	837	YGDCLGDIAA	RDLICAQKFN	GLTVLPPLLT	DEMIAQY TSA	LLAGTITSGW	TFGAGAALQI
QIS30295.1	837	YGDCLGDIAA	RDLICAQKFN	GLTVLPPLLT	DEMIAQY TSA	LLAGTITSGW	TFGAGAALQI
QIS30615.1	837	YGDCLGDIAA	RDLICAQKFN	GLTVLPPLLT	DEMIAQY TSA	LLAGTITSGW	TFGAGAALQI
QIK50427.1	837	YGDCLGDIAA	RDLICAQKFN	GLTVLPPLLT	DEMIAQY TSA	LLAGTITSGW	TFGAGAALQI
QIS30165.1	837	YGDCLGDIAA	RDLICAQKFN	GLTVLPPLLT	DEMIAQY TSA	LLAGTITSGW	TFGAGAALQI
SARS-CoV	819	YGECLGDINA	RDLICAQKFN	GLTVLPPLLT	DEMIAQY TAA	LVSGTATAGW	TFGAGAALQI

CR

YP_009724390	897	PFAMQMAYRF	NGIGVTQNVL	YENQKLIANQ	FNSAIGKIQD	SLSSTASALG	KLQDVVNQNA
QHS34546.1	896	PFAMQMAYRF	NGIGVTQNVL	YENQKLIANQ	FNSAIGKIQD	SLSSTASALG	KLQDVVNQNA
QIS60906.1	897	PFAMQMAYRF	NGIGVTQNVL	YENQKLIANQ	FNSAIGKIQD	SLSSTASALG	KLQDVVNQNA
QIS60930.1	897	PFAMQMAYRF	NGIGVTQNVL	YENQKLIANQ	FNSAIGKIQD	SLSSTASALG	KLQDVVNQNA
QIS61338.1	897	PFAMQMAYRF	NGIGVTQNVL	YENQKLIANQ	FNSAIGKIQD	SLSSTASALG	KLQDVVNQNA
QIS60489.1	897	PFAMQMAYRF	NGIGVTQNVL	YENQKLIANQ	FNSAIGKIQD	SLSSTASALG	KLQDVVNQNA
QIS30625.1	897	PFAMQMAYRF	NGIGVTQNVL	YENQKLIANQ	FNSAIGKIQD	SLSSTASALG	KLQDVVNQNA
QIS30335.1	897	PFAMQMAYRF	NGIGVTQNVL	YENQKLIANQ	FNSAIGKIQD	SLSSTASALG	KLQDVVNQNA
QIC53204.1	897	PFAMQMAYRF	NGIGVTQNVL	YENQKLIANQ	FNSAIGKIQD	SLSSTASALG	KLQDVVNQNA
QHW06059.1	897	PFAMQMAYRF	NGIGVTQNVL	YENQKLIANQ	FNSAIGKIQD	SLSSTASALG	KLQDVVNQNA
QHZ00379.1	897	PFAMQMAYRF	NGIGVTQNVL	YENQKLIANQ	FNSAIGKIQD	SLSSTASALG	KLQDVVNQNA
QHR84449.1	897	PFAMQMAYRF	NGIGVTQNVL	YENQKLIANQ	FNSAIGKIQD	SLSSTASALG	KLQDVVNQNA
QIJ96493.1	897	PFAMQMAYRF	NGIGVTQNVL	YENQKLIANQ	FNSAIGKIQD	SLSSTASALG	KLQDVVNQNA
QIA98583.1	897	PFAMQMAYRF	NGIGVTQNVL	YENQKLIANQ	FNSAIGKIQD	SLSSTASALG	KLQDVVNQNA
QII57278.1	897	PFAMQMAYRF	NGIGVTQNVL	YENQKLIANQ	FNSAIGKIQD	SLSSTASALG	KLQDVVNQNA
QIA20044.1	897	PFAMQMAYRF	NGIGVTQNVL	YENQKLIANQ	FNSAIGKIQD	SLSSTASALG	KLQDVVNQNA
QIS60546.1	897	PFAMQMAYRF	NGIGVTQNVL	YENQKLIANQ	FNSAIGKIQD	SLSSTASALG	KLQDVVNQNA
QIS60582.1	897	PFAMQMAYRF	NGIGVTQNVL	YENQKLIANQ	FNSAIGKIQD	SLSSTASALG	KLQDVVNQNA
QIS61254.1	897	PFAMQMAYRF	NGIGVTQNVL	YENQKLIANQ	FNSAIGKIQD	SLSSTASALG	KLQDVVNQNA
QIS61422.1	897	PFAMQMAYRF	NGIGVTQNVL	YENQKLIANQ	FNSAIGKIQD	SLSSTASALG	KLQDVVNQNA
QIS60978.1	897	PFAMQMAYRF	NGIGVTQNVL	YENQKLIANQ	FNSAIGKIQD	SLSSTASALG	KLQDVVNQNA
QIO04367.1	897	PFAMQMAYRF	NGIGVTQNVL	YENQKLIANQ	FNSAIGKIQD	SLSSTASALG	KLQDVVNQNA
QIS30295.1	897	PFAMQMAYRF	NGIGVTQNVL	YENQKLIANQ	FNSAIGKIQD	SLSSTASALG	KLQDVVNQNA
QIS30615.1	897	PFAMQMAYRF	NGIGVTQNVL	YENQKLIANQ	FNSAIGKIQY	SLSSTASALG	KLQDVVNQNA
QIK50427.1	897	PFAMQMAYRF	NGIGVTQNVL	YENQKLIANQ	FNSAIGKIQD	SLSSTASALG	KLQDVVNQNA
QIS30165.1	897	PFAMQMAYRF	NGIGVTQNVL	YENQKLIANQ	FNSAIGKIQD	SLSSTASALG	KLQDVVNQNA
SARS-CoV	879	PFAMQMAYRF	NGIGVTQNVL	YENQKLIANQ	FNSAIGKIQE	SLSSTASALG	KLQDVVNQNA

CR

HR1

YP_009724390	957	QALNTLVKQL	SSNFGAISSV	LNDILSRDLK	VEAEVQIDRL	ITGRQLSLQT	YVTQQLIRAA
QHS34546.1	956	QALNTLVKQL	SSNFGAISSV	LNDILSRDLK	VEAEVQIDRL	ITGRQLSLQT	YVTQQLIRAA
QIS60906.1	957	QALNTLVKQL	SSNFGAISSV	LNDILSRDLK	VEAEVQIDRL	ITGRQLSLQT	YVTQQLIRAA
QIS60930.1	957	QALNTLVKQL	SSNFGAISSV	LNDILSRDLK	VEAEVQIDRL	ITGRQLSLQT	YVTQQLIRAA
QIS61338.1	957	QALNTLVKQL	SSNFGAISSV	LNDILSRDLK	VEAEVQIDRL	ITGRQLSLQT	YVTQQLIRAA
QIS60489.1	957	QALNTLVKQL	SSNFGAISSV	LNDILSRDLK	VEAEVQIDRL	ITGRQLSLQT	YVTQQLIRAA

QIS30625.1	957	QALNTLVKQL	SSNFGAISSV	LNDILSRDLK	VEAEVQIDRL	ITGRLQSLQT	YVTQQLIRAA
QIS30335.1	957	QALNTLVKQL	SSNFGAISSV	LNDILSRDLK	VEAEVQIDRL	ITGRLQSLQT	YVTQQLIRAA
QIC53204.1	957	QALNTLVKQL	SSNFGAISSV	LNDILSRDLK	VEAEVQIDRL	ITGRLQSLQT	YVTQQLIRAA
QHW06059.1	957	QALNTLVKQL	SSNFGAISSV	LNDILSRDLK	VEAEVQIDRL	ITGRLQSLQT	YVTQQLIRAA
QHZ00379.1	957	QALNTLVKQL	SSNFGAISSV	LNDILSRDLK	VEAEVQIDRL	ITGRLQSLQT	YVTQQLIRAA
QHR84449.1	957	QALNTLVKQL	SSNFGAISSV	LNDILSRDLK	VEAEVQIDRL	ITGRLQSLQT	YVTQQLIRAA
QIJ96493.1	957	QALNTLVKQL	SSNFGAISSV	LNDILSRDLK	VEAEVQIDRL	ITGRLQSLQT	YVTQQLIRAA
QIA98583.1	957	QALNTLVKQL	SSNFGAISSV	LNDILSRDLK	VEAEVQIDRL	ITGRLQSLQT	YVTQQLIRAA
QII57278.1	957	QALNTLVKQL	SSNFGAISSV	LNDILSRDLK	VEAEVQIDRL	ITGRLQSLQT	YVTQQLIRAA
QIA20044.1	957	QALNTLVKQL	SSNFGAISSV	LNDILSRDLK	VEAEVQIDRL	ITGRLQSLQT	YVTQQLIRAA
QIS60546.1	957	QALNTLVKQL	SSNFGAISSV	LNDILSRDLK	VEAEVQIDRL	ITGRLQSLQT	YVTQQLIRAA
QIS60582.1	957	QALNTLVKQL	SSNFGAISSV	LNDILSRDLK	VEAEVQIDRL	ITGRLQSLQT	YVTQQLIRAA
QIS61254.1	957	QALNTLVKQL	SSNFGAISSV	LNDILSRDLK	VEAEVQIDRL	ITGRLQSLQT	YVTQQLIRAA
QIS61422.1	957	QALNTLVKQL	SSNFGAISSV	LNDILSRDLK	VEAEVQIDRL	ITGRLQSLQT	YVTQQLIRAA
QIS60978.1	957	QALNTLVKQL	SSNFGAISSV	LNDILSRDLK	VEAEVQIDRL	ITGRLQSLQT	YVTQQLIRAA
QIO04367.1	957	QALNTLVKQL	SSNFGAISSV	LNDILSRDLK	VEAEVQIDRL	ITGRLQSLQT	YVTQQLIRAA
QIS30295.1	957	QALNTLVKQL	SSNFGAISSV	LNDILSRDLK	VEAEVQIDRL	ITGRLQSLQT	YVTQQLIRAA
QIS30615.1	957	QALNTLVKQL	SSNFGAISSV	LNDILSRDLK	VEAEVQIDRL	ITGRLQSLQT	YVTQQLIRAA
QIK50427.1	957	QALNTLVKQL	SSNFGAISSV	LNDILSRDLK	VEAEVQIDRL	ITGRLQSLQT	YVTQQLIRAA
QIS30165.1	957	QALNTLVKQL	SSNFGAISSV	LNDILSRDLK	VEAEVQIDRL	ITGRLQSLQT	YVTQQLIRAA
SARS-CoV	939	QALNTLVKQL	SSNFGAISSV	LNDILSRDLK	VEAEVQIDRL	ITGRLQSLQT	YVTQQLIRAA

HR1

CH

YP_009724390	1017	EIRASANLAA	TKMSECVLGQ	SKRVDFCGKG	YHLMSFPQSA	PHGVVFLHVT	YVPAQEKNFT
QHS34546.1	1016	EIRASANLAA	TKMSECVLGQ	SKRVDFCGKG	YHLMSFPQSA	PHGVVFLHVT	YVPAQEKNFT
QIS60906.1	1017	EIRASANLAA	TKMSECVLGQ	SKRVDFCGKG	YHLMSFPQSA	PHGVVFLHVT	YVPAQEKNFT
QIS60930.1	1017	EIRASANLAA	TKMSECVLGQ	SKRVDFCGKG	YHLMSFPQSA	PHGVVFLHVT	YVPAQEKNFT
QIS61338.1	1017	EIRASANLAA	TKMSECVLGQ	SKRVDFCGKG	YHLMSFPQSA	PHGVVFLHVT	YVPAQEKNFT
QIS60489.1	1017	EIRASANLAA	TKMSECVLGQ	SKRVDFCGKG	YHLMSFPQSA	PHGVVFLHVT	YVPAQEKNFT
QIS30625.1	1017	EIRASANLAA	TKMSECVLGQ	SKRVDFCGKG	YHLMSFPQSA	PHGVVFLHVT	YVPAQEKNFT
QIS30335.1	1017	EIRASANLAA	TKMSECVLGQ	SKRVDFCGKG	YHLMSFPQSA	PHGVVFLHVT	YVPAQEKNFT
QIC53204.1	1017	EIRASANLAA	TKMSECVLGQ	SKRVDFCGKG	YHLMSFPQSA	PHGVVFLHVT	YVPAQEKNFT
QHW06059.1	1017	EIRASANLAA	TKMSECVLGQ	SKRVDFCGKG	YHLMSFPQSA	PHGVVFLHVT	YVPAQEKNFT
QHZ00379.1	1017	EIRASANLAA	TKMSECVLGQ	SKRVDFCGKG	YHLMSFPQSA	PHGVVFLHVT	YVPAQEKNFT
QHR84449.1	1017	EIRASANLAA	TKMSECVLGQ	SKRVDFCGKG	YHLMSFPQSA	PHGVVFLHVT	YVPAQEKNFT
QIJ96493.1	1017	EIRASANLAA	TKMSECVLGQ	SKRVDFCGKG	YHLMSFPQSA	PHGVVFLHVT	YVPAQEKNFT
QIA98583.1	1017	EIRASANLAA	TKMSECVLGQ	SKRVDFCGKG	YHLMSFPQSA	PHGVVFLHVT	YVPAQEKNFT
QII57278.1	1017	EIRASANLAA	TKMSECVLGQ	SKRVDFCGKG	YHLMSFPQSA	PHGVVFLHVT	YVPAQEKNFT
QIA20044.1	1017	EIRASANLAA	TKMSECVLGQ	SKRVDFCGKG	YHLMSFPQSA	PHGVVFLHVT	YVPAQEKNFT
QIS60546.1	1017	EIRASANLAA	TKMSECVLGQ	SKRVDFCGKG	YHLMSFPQSA	PHGVVFLHVT	YVPAQEKNFT
QIS60582.1	1017	EIRASANLAA	TKMSECVLGQ	SKRVDFCGKG	YHLMSFPQSA	PHGVVFLHVT	YVPAQEKNFT
QIS61254.1	1017	EIRASANLAA	TKMSECVLGQ	SKRVDFCGKG	YHLMSFPQSA	PHGVVFLHVT	YVPAQEKNFT
QIS61422.1	1017	EIRASANLAA	TKMSECVLGQ	SKRVDFCGKG	YHLMSFPQSA	PHGVVFLHVT	YVPAQEKNFT
QIS60978.1	1017	EIRASANLAA	TKMSECVLGQ	SKRVDFCGKG	YHLMSFPQSA	PHGVVFLHVT	YVPAQEKNFT
QIO04367.1	1017	EIRASANLAA	TKMSECVLGQ	SKRVDFCGKG	YHLMSFPQSA	PHGVVFLHVT	YVPAQEKNFT
QIS30295.1	1017	EIRASANLAA	TKMSECVLGQ	SKRVDFCGKG	YHLMSFPQSA	PHGVVFLHVT	YVPAQEKNFT
QIS30615.1	1017	EIRASANLAA	TKMSECVLGQ	SKRVDFCGKG	YHLMSFPQSA	PHGVVFLHVT	YVPAQEKNFT
QIK50427.1	1017	EIRASANLAA	TKMSECVLGQ	SKRVDFCGKG	YHLMSFPQSA	PHGVVFLHVT	YVPAQEKNFT
QIS30165.1	1017	EIRASANLAA	TKMSECVLGQ	SKRVDFCGKG	YHLMSFPQSA	PHGVVFLHVT	YVPAQEKNFT
SARS-CoV	999	EIRASANLAA	TKMSECVLGQ	SKRVDFCGKG	YHLMSFPQAA	PHGVVFLHVT	YVPSQERNFT

CH

BH

SD3

YP_009724390	1077	TAPAICHDGK	AHFPREGVFF	SNGTHWFTVQ	RNFYEPQIIT	TDNTFVSGNC	DVVIGIVNNT
QHS34546.1	1076	TAPAICHDGK	AHFPREGVFF	SNGTHWFTVQ	RNFYEPQIIT	TDNTFVSGNC	DVVIGIVNNT
QIS60906.1	1077	TAPAICHDGK	AHFPREGVFF	SNGTHWFTVQ	RNFYEPQIIT	TDNTFVSGNC	DVVIGIVNNT
QIS60930.1	1077	TAPAICHDGK	AHFPREGVFF	SNGTHWFTVQ	RNFYEPQIIT	TDNTFVSGNC	DVVIGIVNNT
QIS61338.1	1077	TAPAICHDGK	AHFPREGVFF	SNGTHWFTVQ	RNFYEPQIIT	TDNTFVSGNC	DVVIGIVNNT
QIS60489.1	1077	TAPAICHDGK	AHFPREGVFF	SNGTHWFTVQ	RNFYEPQIIT	TDNTFVSGNC	DVVIGIVNNT
QIS30625.1	1077	TAPAICHDGK	AHFPREGVFF	SNGTHWFTVQ	RNFYEPQIIT	TDNTFVSGNC	DVVIGIVNNT

QIS30335.1	1077	TAPAICH	DGK	AHFPREG	VFV	SNGTHW	FVTQ	RNFYEP	QIIT	TDNTFV	SGNC	DVVIGIV	NNNT
QIC53204.1	1077	TAPAICH	DGK	AHFPREG	VFV	SNGTHW	FVTQ	RNFYEP	QIIT	TDNTFV	SGNC	DVVIGIV	NNNT
QHW06059.1	1077	TAPAICH	DGK	AHFPREG	VFV	SNGTHW	FVTQ	RNFYEP	QIIT	TDNTFV	SGNC	DVVIGIV	NNNT
QHZ00379.1	1077	TAPAICH	DGK	AHFPREG	VFV	SNGTHW	FVTQ	RNFYEP	QIIT	TDNTFV	SGNC	DVVIGIV	NNNT
QHR84449.1	1077	TAPAICH	DGK	AHFPREG	VFV	SNGTHW	FVTQ	RNFYEP	QIIT	TDNTFV	SGNC	DVVIGIV	NNNT
QIJ96493.1	1077	TAPAICH	DGK	AHFPREG	VFV	SNGTHW	FVTQ	RNFYEP	QIIT	TDNTFV	SGNC	DVVIGIV	NNNT
QIA98583.1	1077	TAPAICH	DGK	AHFPREG	VFV	SNGTHW	FVTQ	RNFYEP	QIIT	TDNTFV	SGNC	DVVIGIV	NNNT
QII57278.1	1077	TAPAICH	DGK	AHFPREG	VFV	SNGTHW	FVTQ	RNFYEP	QIIT	TDNTFV	SGNC	DVVIGIV	NNNT
QIA20044.1	1077	TAPAICH	DGK	AHFPREG	VFV	SNGTHW	FVTQ	RNFYEP	QIIT	TDNTFV	SGNC	DVVIGIV	NNNT
QIS60546.1	1077	TAPAICH	DGK	AHFPREG	VFV	SNGTHW	FVTQ	RNFYEP	QIIT	TDNTFV	SGNC	DVVIGIV	NNNT
QIS60582.1	1077	TAPAICH	DGK	AHFPREG	VFV	SNGTHW	FVTQ	RNFYEP	QIIT	TDNTFV	SGNC	DVVIGIV	NNNT
QIS61254.1	1077	TAPAICH	DGK	AHFPREG	VFV	SNGTHW	FVTQ	RNFYEP	QIIT	TDNTFV	SGNC	DVVIGIV	NNNT
QIS61422.1	1077	TAPAICH	DGK	AHFPREG	VFV	SNGTHW	FVTQ	RNFYEP	QIIT	TDNTFV	SGNC	DVVIGIV	NNNT
QIS60978.1	1077	TAPAICH	DGK	AHFPREG	VFV	SNGTHW	FVTQ	RNFYEP	QIIT	TDNTFV	SGNC	DVVIGIV	NNNT
QIO04367.1	1077	TAPAICH	DGK	AHFPREG	VFV	SNGTHW	FVTQ	RNFYEP	QIIT	TDNTFV	SGNC	DVVIGIV	NNNT
QIS30295.1	1077	TAPAICH	DGK	AHFPREG	VFV	SNGTHW	FVTQ	RNFYEP	QIIT	TDNTFV	SGNC	DVVIGIV	NNNT
QIS30615.1	1077	TAPAICH	DGK	AHFPREG	VFV	SNGTHW	FVTQ	RNFYEP	QIIT	TDNTFV	SGNC	DVVIGIV	NNNT
QIK50427.1	1077	TAPAICH	DGK	AHFPREG	VFV	SNGTHW	FVTQ	RNFYEP	QIIT	TDNTFV	SGNC	DVVIGIV	NNNT
QIS30165.1	1077	TAPAICH	DGK	AHFPREG	VFV	SNGTHW	FVTQ	RNFYEP	QIIT	TDNTFV	SGNC	DVVIGIV	NNNT
SARS-CoV	1059	TAPAICH	DGK	AHFPREG	VFV	SNGTHW	FVTQ	RNFYEP	QIIT	TDNTFV	SGNC	DVVIGIV	NNNT

SD3

YP_009724390	1137	VYDPLQ	PELD	SFKEELD	KYF	KNHTSP	DVDL	GDISGIN	ASV	VNIQKE	IDRL	NEVAKNL	NES
QHS34546.1	1136	VYDPLQ	PELD	SFKEELD	KYF	KNHTSP	DVDL	GDISGIN	ASV	VNIQKE	IDRL	NEVAKNL	NES
QIS60906.1	1137	VYDPLQ	PELD	SFKEELD	KYF	KNHTSP	DVDL	GDISGIN	ASV	VNIQKE	IDRL	NEVAKNL	NES
QIS60930.1	1137	VYDPLQ	PELD	SFKEELD	KYF	KNHTSP	DVDL	GDISGIN	ASV	VNIQKE	IDRL	NEVAKNL	NES
QIS61338.1	1137	VYDPLQ	PELD	SFKEELD	KYF	KNHTSP	DVDL	GDISGIN	ASV	VNIQKE	IDRL	NEVAKNL	NES
QIS60489.1	1137	VYDPLQ	PELD	SFKEELD	KYF	KNHTSP	DVDL	GDISGIN	ASV	VNIQKE	IDRL	NEVAKNL	NES
QIS30625.1	1137	VYDPLQ	PELD	SFKEELD	KYF	KNHTSP	DVDL	GDISGIN	ASV	VNIQKE	IDRL	NEVAKNL	NES
QIS30335.1	1137	VYDPLQ	PELD	SFKEELD	KYF	KNHTSP	DVDL	GDISGIN	ASV	VNIQKE	IDRL	NEVAKNL	NES
QIC53204.1	1137	VYDPLQ	PELD	SFKEELD	KYF	KNHTSP	DVDL	GDISGIN	ASV	VNIQKE	IDRL	NEVAKNL	NES
QHW06059.1	1137	VYDPLQ	PELD	SFKEELD	KYF	KNHTSP	DVDL	GDISGIN	ASV	VNIQKE	IDRL	NEVAKNL	NES
QHZ00379.1	1137	VYDPLQ	PELD	SFKEELD	KYF	KNHTSP	DVDL	GDISGIN	ASV	VNIQKE	IDRL	NEVAKNL	NES
QHR84449.1	1137	VYDPLQ	PELD	SFKEELD	KYF	KNHTSP	DVDL	GDISGIN	ASV	VNIQKE	IDRL	NEVAKNL	NES
QIJ96493.1	1137	VYDPLQ	PELD	SFKEELD	KYF	KNHTSP	DVDL	GDISGIN	ASV	VNIQKE	IDRL	NEVAKNL	NES
QIA98583.1	1137	VYDPLQ	PELD	SFKEELD	KYF	KNHTSP	DVDL	GDISGIN	ASV	VNIQKE	IDRL	NEVAKNL	NES
QII57278.1	1137	VYDPLQ	PELD	SFKEELD	KYF	KNHTSP	DVDL	GDISGIN	ASV	VNIQKE	IDRL	NEVAKNL	NES
QIA20044.1	1137	VYDPLQ	PELD	SFKEELD	KYF	KNHTSP	DVDL	GDISGIN	ASV	VNIQKE	IDRL	NEVAKNL	NES
QIS60546.1	1137	VYDPLQ	PELD	SFKEELD	KYF	KNHTSP	DVDL	GDISGIN	ASV	VNIQKE	IDRL	NEVAKNL	NES
QIS60582.1	1137	VYDPLQ	PELD	SFKEELD	KYF	KNHTSP	DVDL	GDISGIN	ASV	VNIQKE	IDRL	NEVAKNL	NES
QIS61254.1	1137	VYDPLQ	PELD	SFKEELD	KYF	KNHTSP	DVDL	GDISGIN	ASV	VNIQKE	IDRL	NEVAKNL	NES
QIS61422.1	1137	VYDPLQ	PELD	SFKEELD	KYF	KNHTSP	DVDL	GDISGIN	ASV	VNIQKE	IDRL	NEVAKNL	NES
QIS60978.1	1137	VYDPLQ	PELD	SFKEELD	KYF	KNHTSP	DVDL	GDISGIN	ASV	VNIQKE	IDRL	NEVAKNL	NES
QIO04367.1	1137	VYDPLQ	PELD	SFKEELD	KYF	KNHTSP	DVDL	GDISGIN	ASV	VNIQKE	IDRL	NEVAKNL	NES
QIS30295.1	1137	VYDPLQ	PELD	SFKEELD	KYF	KNHTSP	DVDL	GDISGIN	ASV	VNIQKE	IDRL	NEVAKNL	NES
QIS30615.1	1137	VYDPLQ	PELD	SFKEELD	KYF	KNHTSP	DVDL	GDISGIN	ASV	VNIQKE	IDRL	NEVAKNL	NES
QIK50427.1	1137	VYDPLQ	PELD	SFKEELD	KYF	KNHTSP	DVDL	GDISGIN	ASV	VNIQKE	IDRL	NEVAKNL	NES
QIS30165.1	1137	VYDPLQ	PELD	SFKEELD	KYF	KNHTSP	DVDL	GDISGIN	ASV	VNIQKE	IDRL	NEVAKNL	NES
SARS-CoV	1119	VYDPLQ	PELD	SFKEELD	KYF	KNHTSP	DVDL	GDISGIN	ASV	VNIQKE	IDRL	NEVAKNL	NES

HR2

YP_009724390	1197	LIDLQEL	GKY	EQYIKWP	WYI	WLGFIAG	LIA	IVMVTIM	LCC	MTSCCS	CLKG	CCSCGS	CCKF
QHS34546.1	1196	LIDLQEL	GKY	EQYIKWP	WYI	WLGFIAG	LIA	IVMVTIM	LCC	MTSCCS	CLKG	CCSCGS	CCKF
QIS60906.1	1197	LIDLQEL	GKY	EQYIKWP	WYI	WLGFIAG	LIA	IVMVTIM	LCC	MTSCCS	CLKG	CCSCGS	CCKF
QIS60930.1	1197	LIDLQEL	GKY	EQYIKWP	WYI	WLGFIAG	LIA	IVMVTIM	LCC	MTSCCS	CLKG	CCSCGS	CCKF
QIS61338.1	1197	LIDLQEL	GKY	EQYIKWP	WYI	WLGFIAG	LIA	IVMVTIM	LCC	MTSCCS	CLKG	CCSCGS	CCKF
QIS60489.1	1197	LIDLQEL	GKY	EQYIKWP	WYI	WLGFIAG	LIA	IVMVTIM	LCC	MTSCCS	CLKG	CCSCGS	CCKF
QIS30625.1	1197	LIDLQEL	GKY	EQYIKWP	WYI	WLGFIAG	LIA	IVMVTIM	LCC	MTSCCS	CLKG	CCSCGS	CCKF
QIS30335.1	1197	LIDLQEL	GKY	EQYIKWP	WYI	WLGFIAG	LIA	IVMVTIM	LCC	MTSCCS	CLKG	CCSCGS	CCKF

QIC53204.1	1197	LIDLQELGKY	EQYIKWPWYI	WLGFIAGLIA	IVMVTIMLCC	MTSCCSCLKG	CCSCGSCKKF
QHW06059.1	1197	LIDLQELGKY	EQYIKWPWYI	WLGFIAGLIA	IVMVTIMLCC	MTSCCSCLKG	CCSCGSCKKF
QHZ00379.1	1197	LIDLQELGKY	EQYIKWPWYI	WLGFIAGLIA	IVMVTIMLCC	MTSCCSCLKG	CCSCGSCKKF
QHR84449.1	1197	LIDLQELGKY	EQYIKWPWYI	WLGFIAGLIA	IVMVTIMLCC	MTSCCSCLKG	CCSCGSCKKF
QIJ96493.1	1197	LIDLQELGKY	EQYIKWPWYI	WLGFIAGLIA	IVMVTIMLCC	MTSCCSCLKG	CCSCGSCKKF
QIA98583.1	1197	LIDLQELGKY	EQYIKWPWYI	WLGFIAGLIA	IVMVTIMLCC	MTSCCSCLKG	CCSCGSCKKF
QII57278.1	1197	LIDLQELGKY	EQYIKWPWYI	WLGFIAGLIA	IVMVTIMLCC	MTSCCSCLKG	CCSCGSCKKF
QIA20044.1	1197	LIDLQELGKY	EQYIKWPWYI	WLGFIAGLIA	IVMVTIMLCC	MTSCCSCLKG	CCSCGSCKKF
QIS60546.1	1197	LIDLQELGKY	EQYIKWPWYI	WLGFIAGLIA	IVMVTIMLCC	MTSCCSCLKG	CCSCGSCKKF
QIS60582.1	1197	LIDLQELGKY	EQYIKWPWYI	WLGFIAGLIA	IVMVTIMLCC	MTSCCSCLKG	CCSCGSCKKF
QIS61254.1	1197	LIDLQELGKY	EQYIKWPWYI	WLGFIAGLIA	IVMVTIMLCC	MTSCCSCLKG	CCSCGSCKKF
QIS61422.1	1197	LIDLQELGKY	EQYIKWPWYI	WLGFIAGLIA	IVMVTIMLCC	MTSCCSCLKG	CCSCGSCKKF
QIS60978.1	1197	LIDLQELGKY	EQYIKWPWYI	WLGFIAGLIA	IVMVTIMLCC	MTSCCSCLKG	CCSCGSCKKF
QIO04367.1	1197	LIDLQELGKY	EQYIKWPWYI	WLGFIAGLIA	IVMVTIMLCC	MTSCCSCLKG	CCSCGSCKKF
QIS30295.1	1197	LIDLQELGKY	EQYIKWPWYI	WLGFIAGLIA	IVMVTIMLCC	MTSCCSCLKG	CCSCGSCKKF
QIS30615.1	1197	LIDLQELGKY	EQYIKWPWYI	WLGFIAGLIA	IVMVTIMLCC	MTSCCSCLKG	CCSCGSCKKF
QIK50427.1	1197	LIDLQELGKY	EQYIKWPWYI	WLGFIAGLIA	IVMVTIMLCC	MTSCCSCLKG	CCSCGSCKKF
QIS30165.1	1197	LIDLQELGKY	EQYIKWPWYI	WLGFIAGLIA	IVMVTIMLCC	MTSCCSCLKG	CCSCGSCKKF
SARS-CoV	1179	LIDLQELGKY	EQYIKWPWYI	WLGFIAGLIA	IVMVTIMLCC	MTSCCSCLKG	CCSCGSCKKF

HR2

TM

CT

YP_009724390	1257	DEDDSEPVLK	GVKLHYT
QHS34546.1	1256	DEDDSEPVLK	GVKLHYT
QIS60906.1	1257	DEDDSEPVLK	GVKLHYT
QIS60930.1	1257	DEDDSEPVLK	GVKLHYT
QIS61338.1	1257	DEDDSEPVLK	GVKLHYT
QIS60489.1	1257	DEDDSEPVLK	GVKLHYT
QIS30625.1	1257	DEDDSEPVLK	GVKLHYT
QIS30335.1	1257	DEDDSEPVLK	GVKLHYT
QIC53204.1	1257	DEDDSEPVLK	GVKLHYT
QHW06059.1	1257	DEDDSEPVLK	GVKLHYT
QHZ00379.1	1257	DEDDSEPVLK	GVKLHYT
QHR84449.1	1257	DEDDSEPVLK	GVKLHYT
QIJ96493.1	1257	DEDDSEPVLK	GVKLHYT
QIA98583.1	1257	DEDDSEPVLK	GVKLHYT
QII57278.1	1257	DEDDSEPVLK	GVKLHYT
QIA20044.1	1257	DEDDSEPVLK	GVKLHYT
QIS60546.1	1257	DEDDSEPVLK	GVKLHYT
QIS60582.1	1257	DEDDSEPVLK	GVKLHYT
QIS61254.1	1257	DEDDSEPVLK	GVKLHYT
QIS61422.1	1257	DEDDSEPVLK	GVKLHYT
QIS60978.1	1257	DEDDSEPVLK	GVKLHYT
QIO04367.1	1257	DEDDSEPVLK	GVKLHYT
QIS30295.1	1257	DEDDSEPVLK	GVKLHYT
QIS30615.1	1257	DEDDSEPVLK	GVKLHYT
QIK50427.1	1257	DEDDSEPVLK	GVKLHYT
QIS30165.1	1257	DEDDSEPVLK	GVKLHYT
SARS-CoV	1239	DEDDSEPVLK	GVKLHYT

CT

Supplementary Fig 3. **Multiple sequence alignment of unique SARS-CoV-2 spike protein variants with SARS-CoV spike protein.** Identical regions are shaded throughout all the sequences.

Supplementary Table 1. **List of model quality assessment results of SARS-CoV-2 S protein variants in RAMPAGE server.**

Variants	Number of residues		
	Favoured Region	Allowed Region	Outlier Region
Template (pdb: 6VSB)	892 (95.8%)	38 (4.1%)	1 (0.1%)
Y28N	1025 (91.7%)	74 (6.6%)	19 (1.7%)
T29I	1024 (91.6%)	74 (6.6%)	20 (1.8%)
H49Y	1019 (91.1%)	77 (6.9%)	22 (2.0%)
L54F	1023 (91.5%)	73 (6.5%)	22 (2.0%)
N74K	1014 (90.7%)	79 (7.1%)	25 (2.2%)
E96D	1021 (91.3%)	73 (6.5%)	24 (2.1%)
D111N	1026 (91.8%)	72 (6.4%)	20 (1.8%)
F157L	1028 (91.9%)	71 (6.4%)	19 (1.9%)
G181V	1023 (91.5%)	73 (6.5%)	22 (2.0%)
S247R	1024 (91.6%)	74 (6.6%)	20 (1.8%)
A348T	1019 (91.1%)	81 (7.2%)	18 (1.6%)
R408I	1022 (91.5%)	77 (6.9%)	18 (1.6%)
G476S	1024 (91.6%)	74 (6.6%)	20 (1.8%)
V483A	1030 (92.1%)	74 (6.6%)	14 (1.3%)
H519Q	1023 (91.5%)	71 (6.4%)	24 (2.1%)
A520S	1018 (91.1%)	76 (6.8%)	24 (2.1%)
D614G	1020 (91.2%)	75 (6.7%)	23 (2.1%)
F797C	1016 (90.9%)	77 (6.9%)	25 (2.2%)
A930V	1022 (91.4%)	73 (6.5%)	23 (2.1%)
D936Y	1027 (91.1%)	71 (6.4%)	20 (1.8%)
A1078V	1023 (91.5%)	72 (6.4%)	23 (2.1%)

Supplementary Table 2. **List of variations among 320 SARS-CoV-2 whole genomes.**

Codon Variation	Mutation type	ORF
8782C>T	Synonymous	orf1ab
17747C>T	P5828L (17747C>T)	orf1ab
17858A>G	Y5865C (17858A>G)	orf1ab
18060C>T	Synonymous	orf1ab
28144T>C	Synonymous	orf8
8602G>A	Synonymous	orf1ab
26152G>A	G254R (26152G>A)	orf3a
23120G>T	A520S	S
2094C>T	S610L (2094C>T)	orf1ab
19203T>C	Synonymous	orf1ab
22984G>A	Synonymous	S
27384T>C	Synonymous	orf6
241C>T		5' UTR
1059C>T	T265I (1059C>T)	orf1ab
3037C>T	Synonymous	orf1ab
14408C>T	P4715L (14408C>T)	orf1ab
23403A>G	D614G (23403A>G)	S
25563G>T	Q57H (25563G>T)	orf3a
29553G>A	Synonymous	orf10
34A>T		5' UTR
35A>T		5' UTR
36C>T		5' UTR
37C>A		5' UTR
15927T>G	Synonymous	orf1ab
18401C>T	P6046L (18401C>T)	orf1ab
24795C>T	A1078V (24795C>T)	S
26936C>T	Synonymous	M
11320T>C	Synonymous	orf1ab
13845T>C	Synonymous	orf1ab
16912G>T	V5550L (16912G>T)	orf1ab
8768G>A	D2835N	orf1ab
27525A>G	Synonymous	orf7a
27722T>C	I110T (27722T>C)	orf7a
8945A>G	N2894D (8945A>G)	orf1ab
24022T>C	Synonymous	S
28881G>A	R203K (28881G>A 28882G>A)	N
28882G>A	R203K (28881G>A 28882G>A)	N

28883G>C	G204R (28883G>C)	N
16887C>T	Synonymous	orf1ab
25850G>A	C153Y (25850G>A)	orf3a
21893G>A	D111N (21893G>A)	S
27276C>T	Synonymous	orf6
1238C>T	H325Y (1238C>T)	orf1ab
25494G>T	Synonymous	orf3a
28924T>C	Synonymous	N
16299T>C	Synonymous	orf1ab
16329C>T	Synonymous	orf1ab
23010T>C	V483A (23010T>C)	S
4236A>G	N1324S (4236A>G)	orf1ab
11083G>T	L3606F (11083G>T)	orf1ab
14805C>T	Synonymous	orf1ab
25655T>C	V88A (25655T>C)	orf3a
26144G>T	G251V (26144G>T)	orf3a
26028C>T	Synonymous	orf3a
29866A>G		3' UTR
29868G>C		3' UTR
29869A>T		3' UTR
8078C>T	P2605S (8078C>T)	orf1ab
10319C>T	L3352F (10319C>T)	orf1ab
15720C>T	Synonymous	orf1ab
27964C>T	S24L (27964C>T)	orf8
17326C>T	P5688S (17326C>T)	orf1ab
17884A>G	T5874A (17884A>G)	orf1ab
19645G>T	V6461F (19645G>T)	orf1ab
2232C>T	A656V (2232C>T)	orf1ab
7518C>T	A2418V (7518C>T)	orf1ab
23119T>A	H519Q (23119T>A)	S
673C>T	Synonymous	orf1ab
4683C>T	A1473V (4683C>T)	orf1ab
29881A>C		3' UTR
29888A>C		3' UTR
25064G>C	D1168H (25064G>C)	S
5A>T		5' UTR
3593C>T	L1110F (3593C>T)	orf1ab
27200A>G	Synonymous	M
29862G>A		3' UTR
29864G>A		3' UTR
29867T>A		3' UTR

29868G>A		3' UTR
29870C>A		3' UTR
1440G>A	G392D (1440G>A)	orf1ab
2891G>A	A876T (2891G>A)	orf1ab
28854C>T	S194L (28854C>T)	N
2416C>T	Synonymous	orf1ab
26233G>T		unknown
12809C>T	Synonymous	orf1ab
2041T>C	Synonymous	orf1ab
11704C>T	Synonymous	orf1ab
20256T>C	Synonymous	orf1ab
29866A>T		3' UTR
219G>T		5' UTR
490T>A	D75E (490T>A)	orf1ab
3177C>T	P971L (3177C>T)	orf1ab
9113C>T	P2950S (9113C>T)	orf1ab
18736T>C	F6158L (18736T>C)	orf1ab
21648C>T	T29I (21648C>T)	S
24034C>T	Synonymous	S
26729T>C	Synonymous	M
27635C>T	S81L (27635C>T)	orf7a
28077G>C	V62L (28077G>C)	orf8
29567A>C	I4L (29567A>C)	orf10
29700A>G		3' UTR
2676C>T	P804L (2676C>T)	orf1ab
22606A>T	Synonymous	S
28878G>A	S202N (28878G>A)	N
29742G>A		3' UTR
25337G>C	D1259H (25337G>C)	S
29869A>C		3' UTR
10323A>G	K3353R (10323A>G)	orf1ab
79A>G		5' UTR
23G>C		5' UTR
84C>T		5' UTR
85T>C		5' UTR
90G>C		5' UTR
29410T>C	Synonymous	N
29557G>T		unknown
1A>T		5' UTR
2T>G		5' UTR
29880A>C		3' UTR

6310C>T	Synonymous	orf1ab
27327G>T	K42N (27327G>T)	ORF6
28378G>T	Synonymous	N
1401G>A	G379E (1401G>A)	orf1ab
6693A>G	K2143R (6693A>G)	orf1ab
14877C>T	Synonymous	orf1ab
29188A>T	Synonymous	N
619C>T	Synonymous	orf1ab
14912A>G	N4883S (14912A>G)	orf1ab
6040C>T	Synonymous	orf1ab
12478G>A	M4071I (12478G>A)	orf1ab
21910C>T	Synonymous	S
28896C>G	A208G (28896C>G)	N
29902A>C		3' UTR
29903A>C		3' UTR
9534C>A	T3090N (9534C>A)	orf1ab
18877C>T	Synonymous	orf1ab
28968G>C	S232T (28968G>C 28969C>T)	N
28969C>T	S232T (28968G>C 28969C>T)	N
2110C>T	Synonymous	orf1ab
21575C>T	L5F (21575C>T)	S
21850G>T	E96D (21850G>T)	S
27804C>T	Synonymous	orf7b
5000C>T	L1579F (5000C>T)	orf1ab
29085C>T	T271I (29085C>T)	N
5183C>T	P1640S (5183C>T)	orf1ab
25691G>T	G100V (25691G>T)	orf3a
28708C>T	Synonymous	N
11916C>T	S3884L (11916C>T)	orf1ab
55A>C		5' UTR
1912C>T	Synonymous	orf1ab
25452C>T	Synonymous	orf3a
27246A>C	L15F (27246A>C)	orf6
27247C>T	L16S (27247C>T 27248T>C)	orf6
27248T>C	Synonymous	orf6
27255T>G	Synonymous	orf6
27340A>C	N47Q (27340A>C 27342T>G)	orf6
27342T>G	Synonymous	orf6
27344A>G	K48R (27344A>G)	orf6
8937C>T	A2891V (8937C>T)	orf1ab
25523G>T	G44V (25523G>T)	orf3a

27869T>C	Synonymous	orf7b
13748A>G	K4495R (13748A>G)	orf1ab
1191C>T	P309L (1191C>T)	orf1ab
21724G>C	L54F (21724G>C)	S
12111G>A	S3949N (12111G>A)	orf1ab
10818C>T	A3518V (10818C>T)	orf1ab
14937C>T	Synonymous	orf1ab
22988G>A	G476S (22988G>A)	S
37C>T		5' UTR
26433A>G	Synonymous	E
28099C>T	S69L (28099C>T)	orf8
833T>C	F190L (833T>C)	orf1ab
19137A>T	L6291F (19137A>T)	orf1ab
8090C>T	L2609F (8090C>T)	orf1ab
17247T>C	Synonymous	orf1ab
565T>C	Synonymous	orf1ab
17825C>T	T5854I (17825C>T)	orf1ab
29573G>A	V6I (29573G>A)	orf10
29861G>A		3' UTR
28280G>T	D3Y (28280G>T)	N
29515T>A	Synonymous	N
34A>C		5' UTR
24368G>T	D936Y (24368G>T)	S
6424A>C	E2053D (6424A>C)	orf1ab
11042G>T	V3593F (11042G>T)	orf1ab
22207T>C	Synonymous	S
24694A>T	Synonymous	S
28283A>G	N4D (28283A>G)	N
20692C>T	P6810S (20692C>T)	orf1ab
29845T>G		3' UTR
29846T>A		3' UTR
3T>C		5' UTR
4A>T		5' UTR
140C>T		5' UTR
12213C>T	S3983F (12213C>T)	orf1ab
19269C>T	Synonymous	orf1ab
11674C>T	Synonymous	orf1ab
28826C>T	R185C (28826C>T)	N
197C>T		5' UTR
198G>T		5' UTR
200T>C		5' UTR

201T>A		5' UTR
202T>G		5' UTR
204G>C		5' UTR
205T>G		5' UTR
206C>G		5' UTR
208G>T		5' UTR
210G>A		5' UTR
2632G>T	M789I (2632G>T)	orf1ab
13225C>T	Synonymous	orf1ab
14836A>G	I4858V (14836A>G)	orf1ab
9180C>T	S2972F (9180C>T)	orf1ab
29774C>T		3' UTR
354G>A	G30D (354G>A)	orf1ab
5716G>T	K1817N (5716G>T)	orf1ab
19018C>T	P6252S (19018C>T)	orf1ab
4642G>A	Synonymous	orf1ab
27807C>T	Synonymous	orf7b
14229G>T	K4655N (14229G>T)	orf1ab
832C>T	Synonymous	orf1ab
1545C>T	A427V (1545C>T)	orf1ab
3992C>T	Synonymous	orf1ab
27806G>A	Synonymous	orf7b
28229T>G	H112Q (28229T>G)	orf8
14604A>G	Synonymous	orf1ab
21911C>T	Synonymous	S
19684G>T	V6474L (19684G>T)	orf1ab
27999C>T	P36S (27999C>T)	orf8
19632T>A	N6456K (19632T>A)	orf1ab
23989A>G	Synonymous	S
17373C>T	Synonymous	orf1ab
10572C>T	A3436V (10572C>T)	orf1ab
11538T>C	I3758T (11538T>C)	orf1ab
23242G>T	Synonymous	S
28376G>A	A35T (28376G>A)	N
6691T>C	Synonymous	orf1ab
7843C>T	Synonymous	orf1ab
13768A>G	M4502V (13768A>G)	orf1ab
15869A>T	H5202L (15869A>T)	orf1ab
17639C>T	S5792L (17639C>T)	orf1ab
19542G>T	M6426I (19542G>T)	orf1ab
28915C>T	Synonymous	N

3955G>T	K1230N (3955G>T)	orf1ab
8681A>G	I2806V (8681A>G)	orf1ab
22604G>A	A348T (22604G>A)	S
150T>C		5' UTR
153T>G		5' UTR
29747G>A		3' UTR
18433G>A	D6057N (18433G>A)	orf1ab
29167C>T	Synonymous	N
15888T>C	Synonymous	orf1ab
313C>T	Synonymous	orf1ab
1623T>C	I453T (1623T>C)	orf1ab
3299T>C	Synonymous	orf1ab
25156C>T	Synonymous	S
29844A>G		3' UTR
29847T>G		3' UTR
29848T>G		3' UTR
3099C>T	T945I (3099C>T)	orf1ab
6636C>T	T2124I (6636C>T)	orf1ab
11750C>T	L3829F (11750C>T)	orf1ab
11956C>T	Synonymous	orf1ab
16075G>T	D5271Y (16075G>T)	orf1ab
18689T>A	F6142Y (18689T>A)	orf1ab
20031C>A	Synonymous	orf1ab
27046C>T	T175M (27046C>T)	M
28409C>T	P46S (28409C>T)	N
29736G>T		3' UTR
3259G>T	Q998H (3259G>T)	orf1ab
2269A>T	Synonymous	orf1ab
29635C>T	Synonymous	orf10
9474C>T	A3070V (9474C>T)	orf1ab
1385C>T	H374Y (1385C>T)	orf1ab
29230C>T	Synonymous	N
6026C>T	P1921S (6026C>T)	orf1ab
12473C>T	Synonymous	orf1ab
1469C>T	R402C (1469C>T)	orf1ab
29254G>T	Synonymous	N
10036C>T	Synonymous	orf1ab
3T>G		5' UTR
4A>G		5' UTR
6A>T		5' UTR
332G>T	V23F (332G>T)	orf1ab

5784C>T	T1840I (5784C>T)	orf1ab
20823C>T	Synonymous	orf1ab
2446T>C	Synonymous	orf1ab
3411C>T	A1049V (3411C>T)	orf1ab
5572G>T	M1769I (5572G>T)	orf1ab
8092C>T	Synonymous	orf1ab
26037C>T	Synonymous	orf3a
27877G>T	C41F (27877G>T)	orf7b
20281T>C	F6673L (20281T>C)	orf1ab
254C>T	Synonymous	orf1ab
29751G>C		3' UTR
9157T>C	Synonymous	orf1ab
22104G>T	G181V (22104G>T)	S
28916G>A	G215S (28916G>A)	N
10507C>T	Synonymous	orf1ab
29871A>G		3' UTR
6035A>G	S1924G (6035A>G)	orf1ab
16467A>G	Synonymous	orf1ab
21386C>T	S7041F (21386C>T)	orf1ab
23185C>T	Synonymous	S
29867T>C		3' UTR
29871A>T		3' UTR
29872A>C		3' UTR
29873A>G		3' UTR
29898A>G		3' UTR
29902A>G		3' UTR
29903A>G		3' UTR
5052C>T	P1596L (5052C>T)	orf1ab
17474C>T	T5737I (17474C>T)	orf1ab
27226G>T	V9F (27226G>T)	orf6
1102C>T	Synonymous	orf1ab
21147T>C	Synonymous	orf1ab
22432C>T	Synonymous	S
29837C>A		3' UTR
514T>C	Synonymous	orf1ab
17410C>T	R5716C (17410C>T)	orf1ab
10232C>T	R3323C (10232C>T)	orf1ab
221T>C		5' UTR
27A>T		5' UTR
18814C>T	Synonymous	orf1ab
9477T>A	F3071Y (9477T>A)	orf1ab

25979G>T	G196V (25979G>T)	orf3a
28657C>T	Synonymous	N
28863C>T	S197L (28863C>T)	N
4288G>T	E1341D (4288G>T)	orf1ab
4307A>C	K1348Q (4307A>C)	orf1ab
7479A>G	N2405S (7479A>G)	orf1ab
11207G>C	A3648P (11207G>C)	orf1ab
11233T>G	Synonymous	orf1ab
12041G>C	D3926H (12041G>C)	orf1ab
12160G>C	E3965D (12160G>C)	orf1ab
12202G>C	K3979N (12202G>C)	orf1ab
12208G>T	K3981N (12208G>T)	orf1ab
12355G>C	Q4030H (12355G>C)	orf1ab
12378G>A	R4038K (12378G>A)	orf1ab
12464G>T	A4067S (12464G>T)	orf1ab
12467G>T	A4068S (12467G>T)	orf1ab
12491G>T	D4076Y (12491G>T)	orf1ab
12514G>C	Synonymous	orf1ab
12572G>T	D4103Y (12572G>T)	orf1ab
12578G>T	D4105Y (12578G>T)	orf1ab
12582G>T	S4106I (12582G>T)	orf1ab
12600G>A	S4112N (12600G>A)	orf1ab
12660G>C	R4132T (12660G>C)	orf1ab
12685G>C	Q4140H (12685G>C)	orf1ab
12773G>T	A4170S (12773G>T)	orf1ab
12793G>T	K4176N (12793G>T)	orf1ab
20980G>C	D6906H (20980G>C)	orf1ab
21784T>A	N74K (21784T>A)	S
9511T>C	Synonymous	orf1ab
7G>A		5' UTR
8G>C		5' UTR
10T>A		5' UTR
11161G>A	Synonymous	orf1ab
18149G>T	Synonymous	orf1ab
29865A>G		3' UTR
10632C>T	A3456V (10632C>T)	orf1ab
884C>T	R207C (884C>T)	orf1ab
1348C>T	Synonymous	orf1ab
1397G>A	V378I (1397G>A)	orf1ab
9159C>T	P2965L (9159C>T)	orf1ab
3373C>A	D1036E (3373C>A)	orf1ab

9561C>T	S3099L (9561C>T)	orf1ab
15607T>C	Synonymous	orf1ab
29095C>T	Synonymous	N
17000C>T	T5579I (17000C>T)	orf1ab
1548G>A	S428N (1548G>A)	orf1ab
28792A>T	Synonymous	N
8001A>C	D2579A (8001A>C)	orf1ab
9534C>T	T3090I (9534C>T)	orf1ab
7016G>A	G2251S (7016G>A)	orf1ab
21137A>G	K6958R (21137A>G)	orf1ab
21316G>A	D7018N (21316G>A)	orf1ab
24325A>G	Synonymous	S
19065T>C	Synonymous	orf1ab
22303T>G	S247R (22303T>G)	S
7866G>T	G2534V (7866G>T)	orf1ab
104T>A		5' UTR
111T>C		5' UTR
112T>G		5' UTR
119C>G		5' UTR
120T>C		5' UTR
124G>A		5' UTR
6996T>C	I2244T (6996T>C)	orf1ab
614G>A	A117T (614G>A)	orf1ab
5084A>G	I1607V (5084A>G)	orf1ab
2091C>T	T609I (2091C>T)	orf1ab
21707C>T	H49Y (21707C>T)	S
2971G>T	M902I (2971G>T)	orf1ab
6031C>T	Synonymous	orf1ab
12115C>T	Synonymous	orf1ab
15597T>C	Synonymous	orf1ab
20936C>T	T6891M (20936C>T)	orf1ab
22224C>G	S221W (22224C>G)	S
25775G>T	W128L (25775G>T)	orf3a
26354T>A	L37H (26354T>A)	E
3518G>T	V1085F (3518G>T)	orf1ab
17423A>G	Y5720C (17423A>G)	orf1ab
75C>A		5' UTR
11083G>C	L3606F (11083G>C)	orf1ab
21644T>A	Y28N (21644T>A)	S
9034A>G	H3076Y (9491C>T)	orf1ab
9491C>T	Synonymous	orf1ab

186C>T		5' UTR
2717G>A	G818S (2717G>A)	orf1ab
9274A>G	Synonymous	orf1ab
13225C>G	F4321L (13226T>C)	orf1ab
13226T>C	Synonymous	orf1ab
17376A>G	Synonymous	orf1ab
23952T>G	F797C (23952T>G)	S
18603T>C	Synonymous	orf1ab
18975T>A	Synonymous	orf1ab
19175A>C	D6304A (19175A>C)	orf1ab
27925C>T	T11I (27925C>T)	orf8
9924C>T	A3220V (9924C>T)	orf1ab
1A>C		5' UTR
12534C>T	T4090I (12534C>T)	orf1ab
13072C>T	Synonymous	orf1ab
29873A>T		3' UTR
29876A>G		3' UTR
29877A>G		3' UTR
29878A>T		3' UTR
29880A>G		3' UTR
29882A>G		3' UTR
29886A>T		3' UTR
29887A>T		3' UTR
29889A>C		3' UTR
29890A>T		3' UTR
29891A>T		3' UTR
29892A>G		3' UTR
29893A>T		3' UTR
29894A>C		3' UTR
29897A>C		3' UTR
29898A>T		3' UTR
29901A>T		3' UTR
29902A>T		3' UTR
6819G>T	S2185I (6819G>T)	orf1ab
19610C>T	T6449I (19610C>T)	orf1ab
29527G>A	Synonymous	N
15324C>T	Synonymous	orf1ab
29303C>T	P344S (29303C>T)	N
29884A>G		3' UTR
29885A>G		3' UTR
29887A>G		3' UTR

654G>A	G130E (654G>A)	orf1ab
4402T>C	Synonymous	orf1ab
5062G>T	L1599F (5062G>T)	orf1ab
29301A>T	D343V (29301A>T)	N
1691A>G	I476V (1691A>G)	orf1ab
6501C>T	P2079L (6501C>T)	orf1ab
16877C>T	T5538I (16877C>T)	orf1ab
24351C>T	A930V (24351C>T)	S
2277T>C	I671T (2277T>C)	orf1ab
6695C>T	P2144S (6695C>T)	orf1ab
14657C>T	A4798V (14657C>T)	orf1ab
22785G>T	R408I (22785G>T)	S
3738C>T	P1158L (3738C>T)	orf1ab
11410G>A	Synonymous	orf1ab
26326C>T	Synonymous	E
22033C>A	F157L (22033C>A)	S
5845A>T	K1860N (5845A>T)	orf1ab
3778A>G	Synonymous	orf1ab
8388A>G	N2708S (8388A>G)	orf1ab
8987T>A	F2908I (8987T>A)	orf1ab

Supplementary Table 3. **List of SARS-CoV-2 Spike Protein Variants.**

Variants	Frequency	Accession	SARS Residue ^b
L5F	3	QIS60906 ^a	L5
Y28N	1	QIA20044	Y30
T29I	1	QIS60546	T31
H49Y	1	QHW06059	Y53
L54F	1	QIS30295	L58
N74K	1	QIO04367	Gap
E96D	1	QIS60930	E93
D111N	1	QIS61338	N108
Y145Del	1	QHS34546	-
F157L	1	QII57278	T150
G181V	1	QIJ96493	E174
S221W	1	QHZ00379	N214
S247R	1	QHR84449	Gap
A348T	1	QIS30335	P335
R408I	1	QHS34546	R395
G476S	3	QIS30625 ^a	D463
V483A	4	QIS30165 ^a	P469
H519Q	1	QIS61422	N505
A520S	1	QIS60489	A506
D614G	71	QIK50427 ^a	D601
F797C	1	QIC53204	F779
A930V	1	QIA98583	A912
D936Y	1	QIS30615	E918
A1078V	1	QIS61254	A1060
D1168H	1	QIS60978	D1150
D1259H	1	QIS60582	D1241

^aSupplementary Table 5 contain accession numbers of identical sequences of QIS60906, QIK50427, QIS30625, and QIS30165.

^bRespective residue with position was counted from the MSA between spike protein of SARS-CoV and that of SARS-CoV-2.

Supplementary Table 4. **List of SARS-CoV-2 S protein sequences identical to SARS-CoV-2 reference S protein sequence (YP_009724390).**

YP_009724390					
QHU79194	QHU79204	QIK50448	QIQ49992	QIS60399	QIS61386
QHZ00358	QHZ87582	QIA98554	QIQ50132	QIS60419	QIS60858
QHO62112	QII87818	QII57188	QIQ50122	QIS60469	QIS61434
QHQ71973	QII87782	QII57228	QIQ49952	QIS60429	QIS60510
QHU36824	QII87794	QIK02944	QIQ50162	QIS60276	QIS60522
QHU36834	QII87806	QIK02964	QIQ50192	QIS60439	QIS60534
QHZ00389	QIC53213	QIJ96463	QIQ49862	QIK02954	QIS60570
QHZ00399	QIE07461	QIJ96513	QIQ50052	QIH45033	QIS60594
QHU36864	QIJ96473	QII57248	QIQ49842	QHD43416	QIS60678
QHN73795	QIE07451	QII57298	QIQ50182	QIH45043	QIS60690
QHQ71963	QIK50438	QII57238	QIQ49852	QIH45053	QIS60714
QHW06049	QIE07471	QII57308	QIQ49792	QHO62107	QIS61230
QHQ82464	QIJ96483	QII57288	QIQ49812	QHR63260	QIS61326
QHO62877	QHR63250	QIG55994	QIQ49822	QHR63290	QIS61350
QHU36844	QII57258	QII57168	QIQ50032	QHW06039	QIS60798
QHU36854	QIS30065	QII57218	QIQ50042	QID21068	QIS61398
QIA98596	QIS30018	QII57268	QIQ50062	QHR63270	QIS61410
QIA98606	QIS30030	QII57338	QIQ49912	QHR63280	QIS60966
QIB84673	QIS30255	QIH45023	QIQ49932	QIQ08790	QIS60990

QHN73810	QIS30325	QID98794	QIQ50002	QIQ08830	QIS61002
QHZ87592	QIS30345	QID21058	QIQ49892	QIQ50012	QIS61014
QIE07481	QIS30375	QID21048	QIQ68564	QIQ50022	QIS61026
QIJ96503	QIS30465	QIS30545	QIQ22760	QIQ49962	QIS61038
QIJ96523	QIS29994	QIS30555	QIS60726	QIQ49982	QIS61086
QII57161	QIS29982	QIS30655	QIS60738	QIS60329	QIS61098
QIH55221	QIS30006	QIS30475	QIS61266	QIS60339	QIS61146
QIK50417	QIS30175	QIS30215	QIS61302	QIS60349	QIS61158
QII57178	QIS30495	QIS30042	QIS61314	QIS60449	QIS61218
QII57318	QIS30515	QIS30275	QIS60786	QIS60409	QIS61482
QII57208	QIS30605	QIS60479	QIS61362	QIS60459	QIS61494
QII57198	QIS30485	QIS60499	QIS60359	QIS60299	QIS61506
QII57328	QIS30675	QIS60309	QIS60369	QIS60319	QIS61566
QIQ68634	QIQ68604	QIQ68514	QIS60379	QIQ68684	QIQ68554
QHO60594	QIQ68614	QIQ68574	QIS60389	QIQ68544	QIQ68644
QIQ68494	QIQ68654	QIQ68584	QIQ68674	QIM47467	QIQ68524
QIQ68664	QIQ68694	QIQ68594	QIQ68464	QIQ68624	

Supplementary Table 5. **List of SARS-CoV-2 S protein sequences having similar variation.**

Variant	Identical Sequences
D614G	QIK50427, QIS30235, QIS30245, QIS30265, QIS30365, QIS30385, QIS30435, QIS30445, QIS30155, QIS30305, QIS30095, QIS30195, QIS30415, QIS30315, QIS30405, QIS30535, QIS30585, QIS30665, QIS30125, QIS30075, QIS30085, QIS30285, QIS60288, QIQ49772, QIQ49972, QIQ50142, QIQ50112, QIQ50172, QIQ49872, QIQ50072, QIQ49762, QIQ49782, QIQ49802, QIQ50082, QIQ49942, QIS61290, QIS60870, QIS60618, QIS60630, QIS60642, QIS60666, QIS61242, QIS61278, QIS60762, QIS60810, QIS60834, QIS60894, QIS60918, QIS60942, QIS61050, QIS61074, QIS61122, QIS61134, QIS61182, QIS61194, QIS61206, QIS61446, QIS61518, QIS61542, QIS61554, QIQ68484, QIQ68474, QIQ68504, QIQ68534
L5F	QIS60906, QIQ50152, QIQ49882
V483A	QIS30165, QIS60774, QIS60882, QIS60954
G476S	QIQ50152, QIQ49882, QIS30625