



Insights on the Structural Variations of the Furin-Like Cleavage Site Found Among the December 2019–July 2020 SARS-CoV-2 Spike Glycoprotein: A Computational Study Linking Viral Evolution and Infection

OPEN ACCESS

Edited by:

Matteo Convertino,
Hokkaido University, Japan

Reviewed by:

Takahiro Watanabe,
Nagoya University, Japan
Vladimir N. Uversky,
University of South Florida,
United States

*Correspondence:

Marni E. Cueno
marni.cueno@nihon-u.ac.jp

Specialty section:

This article was submitted to
Infectious Diseases Surveillance,
Prevention and Treatment,
a section of the journal
Frontiers in Medicine

Received: 02 October 2020

Accepted: 16 February 2021

Published: 10 March 2021

Citation:

Cueno ME, Ueno M, Iguchi R,
Harada T, Miki Y, Yasumaru K, Kiso N,
Wada K, Baba K and Imai K (2021)
Insights on the Structural Variations of
the Furin-Like Cleavage Site Found
Among the December 2019–July
2020 SARS-CoV-2 Spike
Glycoprotein: A Computational Study
Linking Viral Evolution and Infection.
Front. Med. 8:613412.
doi: 10.3389/fmed.2021.613412

Marni E. Cueno^{1,2,3*}, Miu Ueno³, Rinako Iguchi³, Tsubasa Harada³, Yoshifumi Miki²,
Kanae Yasumaru², Natsumi Kiso², Kanta Wada², Koki Baba² and Kenichi Imai¹

¹ Department of Microbiology, Nihon University School of Dentistry, Tokyo, Japan, ² Immersion Biology Class, Department of Science, Tokyo Gakugei University International Secondary School, Tokyo, Japan, ³ Immersion Physics Class, Department of Science, Tokyo Gakugei University International Secondary School, Tokyo, Japan

The SARS-CoV-2 (SARS2) is the cause of the coronavirus disease 2019 (COVID-19) pandemic. One unique structural feature of the SARS2 spike protein is the presence of a furin-like cleavage site (FLC) which is associated with both viral pathogenesis and host tropism. Specifically, SARS2 spike protein binds to the host ACE-2 receptor which in-turn is cleaved by furin proteases at the FLC site, suggesting that SARS2 FLC structural variations may have an impact on viral infectivity. However, this has not yet been fully elucidated. This study designed and analyzed a COVID-19 genomic epidemiology network for December 2019 to July 2020, and subsequently generated and analyzed representative SARS2 spike protein models from significant node clusters within the network. To distinguish possible structural variations, a model quality assessment was performed before further protein model analyses and superimposition of the protein models, particularly in both the receptor-binding domain (RBD) and FLC. Mutant spike models were generated with the unique ⁶⁸¹PRRA⁶⁸⁴ amino acid sequence found within the deleted FLC. We found 9 SARS2 FLC structural patterns that could potentially correspond to nine node clusters encompassing various countries found within the COVID-19 genomic epidemiology network. Similarly, we associated this with the rapid evolution of the SARS2 genome. Furthermore, we observed that either in the presence or absence of the unique ⁶⁸¹PRRA⁶⁸⁴ amino acid sequence no structural changes occurred within the SARS2 RBD, which we believe would mean that the SARS2 FLC has no structural influence on SARS2 RBD and may explain why host tropism was maintained.

Keywords: furin-like cleavage site, infection clusters, SARS-CoV-2 (SARS2), spike glycoprotein, structural variations

INTRODUCTION

Coronaviruses (CoV) are enveloped positive-stranded RNA viruses that have the largest genome among all known RNA viruses and, at present, there are seven known CoVs capable of infecting humans (1–7). Among CoV structural proteins, the spike protein is a class I viral fusion protein that is involved in viral entry, host tropism determination, viral pathogenesis, and host immune response induction (8–11). The spike protein is comprised of three segments (large ectodomain, single-pass transmembrane anchor, and short intracellular tail) (11), with the ectodomain further divided into the S1 receptor-binding subunit and S2 membrane-fusion subunit (10, 11). During a typical CoV infection, S1 binds to an ideal host receptor enabling viral attachment and, consequently, S2 would fuse the host and viral membranes, allowing viral genetic material to enter host cells (10, 11).

Interestingly, prior to SARS-CoV-2 (SARS2), there were six human pathogenic coronaviruses (10), with SARS2 resulting in a pandemic causing the coronavirus disease 2019 (COVID-19) (12, 13). With regards to the homotrimeric spike protein, the SARS2 spike protein follows the same mechanism of viral entry used by SARS-CoV-1, wherein, the SARS2 spike protein binds to a functional receptor human angiotensin-converting enzyme 2 (ACE2) via the 6-residue (L455, F486, Q493, S494, N501, Y505) receptor-binding domain (RBD) (10, 14). One notable structural feature of the SARS2 spike protein is the presence of a polybasic (furin-like) cleavage site (⁶⁸²RRAR⁶⁸⁵) which has been found to be disordered (15, 16) and, likewise, linked to effective furin cleavage that could help determine viral pathogenesis and host tropism (17–19). Moreover, the comparative analysis of the intrinsic disorder predisposition of spike protein from SARS2, SARS, and Bat CoV revealed that the furin-like cleavage site of SARS spike is incorporated within the longer disordered region ⁶⁷⁶TQTNSPRRRARSVAS⁶⁹¹, which is not present in spike proteins from SARS and Bat CoV (20). The presence of disorder in a region containing a polybasic (furin-like) cleavage site is an extremely important point, as an intrinsic disorder at the cleavage site is crucial for efficient protease action (20, 21). Furthermore, aside from the presence of the polybasic cleavage site (⁶⁸²RRAR⁶⁸⁵), SARS2 likewise has an inserted leading proline (P681), which is suggested to improve protease active site accessibility not only by furin proteases but other proteases as well (21). Thus, this would mean that the inserted sequence unique for SARS2 is the ⁶⁸¹PRRA⁶⁸⁴ sequence (18).

The structural orientation of either individual or a series of amino acids plays an important role in establishing both protein configuration and protein-protein complexes (22), which likewise may affect protein function (23). This would imply that any probable changes in structural orientation occurring in the SARS2 spike furin-like cleavage (including P681) site (FLC) may have an impact on viral infectivity (24). However, to our knowledge, this has never been fully elucidated. A better understanding of the potential effects of the structural orientation changes occurring within the SARS2 FLC site may shed light on the occurrence of varying SARS2 variants and, more importantly,

its role in viral reinfection, potentially leading to novel drug design and therapeutic strategies.

MATERIALS AND METHODS

COVID-19 Genomic Epidemiology Network Design and Analyses Between December 2019 and July 2020

Network analyses were performed in order to gather a holistic understanding of the phylogeny of the COVID-19 genomic epidemiology (25). For this study, network design followed the phylogenetic tree of the COVID-19 genomic epidemiology, based on the GISAID website (www.gisaid.org) between December 2019 and July 2020. A total of 2,793 genomes were used for both network design and analyses. We used Cytoscape for both network design and analyses (26). For network design, nodes were made to represent the countries (indicated as a box) and phylogenetic branch points (indicated as dots) while the edges represent the phylogenetic lineage originating from either a country or branch point. For network analyses, the following centrality measurements were initially analyzed: (1) stress centrality (identifying important nodes); (2) eccentricity centrality (identifying accessible nodes); (3) closeness centrality (identifying relevant nodes); (4) betweenness centrality (identifying crucial nodes); and (5) edge betweenness centrality (identifying significant edges) (27). Briefly, nodes (**Supplementary Figure 1**) and edges (**Supplementary Figure 2**) above a computed threshold for each centrality were considered significant. A unified network was designed based on all centrality measurements used for this study (both nodal and edge centralities) and, more importantly, nodes that were linked to either nodes or edges that are above the threshold based on all five centrality measurements used were determined.

SARS2 Spike Protein Modeling

Representative SARS2 spike amino acid sequences ($n = 263$) deposited between December 2019 and July 2020 were collected from the National Center for Biological Information (NCBI). The selection of sequences was based on the results obtained from our previous COVID-19 genomic epidemiology network analyses. Moreover, representative monomeric SARS2 spike models were selected using Tm align (28). Briefly, a minimum of 10 generated sequence models were initially obtained. Further structural analyses used spike models with similar Root Mean Square Deviation (RMSD) values and Template Modeling scores (Tm-scores) based on superimposition. In particular, the SARS2 spike models used for further structural analyses were based on structural variations in SARS2 FLC and have the following Genebank accession numbers: MT019529, MN994468, MT020781, MT825091, MT467261, MT658503, MT499218, MT549887, and MT461625. The Phyre2 web server (29) was used to generate all protein models while the Jmol applet (30) was used for protein visualization.

Protein Model Quality Assessment

To confirm the accuracy and suitability of the generated SARS2 spike protein models for further analyses, both contact

mapping and protein model:crystal structure superimposition were performed for model quality assessment. A protein contact map was made using the CMView applet to determine the common contact between the model and crystal (31). Moreover, higher common contact (>90%) would mean more structural similarities (32), which would mean that the generated model is suitable for further analyses. Subsequently, representative SARS2 spike cryo-EM structure (PDB ID: 6XR8) (15) and a monomeric 6XR8 model (cryo-EM model) generated using Phyre 2 were used for superimposition (using Tm align) to serve as a model quality check. For this study, SARS2 spike models were considered suitable for further analyses if superimposed sequence model:crystal and crystal model:crystal have RMSD < 1.50.

Comparison of SARS2 Spike Models

All structural comparisons conducted focused on both the SARS2 FLC and RBD. Moreover, two sets of structural comparisons were made. The first set of structural comparisons focused on contrasting the SARS2 FLC and RBD among all representative SARS2 spike models through superimposition. One of the representative models (generated from MT019529) was used as the common model for superimposition. The second set of structural comparisons involved producing mutants from all representative SARS2 spike models without the ⁶⁸¹PRRA⁶⁸⁴ sequence unique in SARS2. A protein threading approach (via Phyre 2) was used to generate the mutant models. Similarly, focusing on SARS2 FLC and RBD, the original model (with ⁶⁸¹PRRA⁶⁸⁴) was compared to the mutated model (without ⁶⁸¹PRRA⁶⁸⁴) through superimposition using Tm align. Model superimposition (focusing on SARS2 FLC and RBD), RMSD values, and Tm scores were established using Jmol and Tm align, respectively.

RESULTS

Nine Node Clusters From the COVID-19 Genomic Epidemiology Network Were Established Between December 2019 and July 2020

The SARS2 genome is constantly evolving, and genome distribution varies in terms of geographic location (33, 34). To establish possible node clusters within the COVID-19 genomic epidemiology network established between December 2019 and July 2020, network analytics was performed to elucidate the holistic and simultaneous analyses of complementary data (27, 35). One of the key points of network analytics is centrality analysis, which involves collecting network components in order to distinguish important elements and, likewise, requires several centrality measurements to be considered fully efficient for analyzing networks (27, 36). Considering this and the five different centrality measurements used to identify node clusters, this would suggest that the results obtained are reliable. Interestingly, we were able to identify nine node clusters, encompassing various SARS2 genomic clades classified by the GISAID website (Figure 1A). We observed that some of the countries identified among the nine node clusters are likewise

found in other node clusters (regardless of belonging to different SARS2 clades) (Figure 1B). These results could mean that the putative significant node clusters are not dependent on SARS2 clades, which coincidentally are based on viral genome mutations (34). This insinuates that there could be other similarities among the node clusters with regard to SARS2 pathogenesis. Considering that the SARS2 FLC is crucial for viral pathogenesis and host tropism (17–19), which we believe would imply that the SARS2 FLC is a conserved structural feature (18), we postulate that the SARS2 FLC could be a common structural feature among the node clusters. We wish to emphasize that our current study mainly focused on the SARS2 FLC structural feature. In possible future work, it would be interesting to recognize other possible spike protein structural features found among the node clusters identified.

SARS2 Spike Models Are Suitable for Structural Analyses

It has long been recommended that model quality assessment be performed prior to any downstream structural analyses using protein structures generated from either experimental (i.e., crystallized) or theoretical (i.e., computer-based) methods (37). To establish the reliability and suitability of all SARS2 spike models generated, both protein contact maps and structural superimpositions were performed. Representative SARS2 crystal structure (Figure 2A), SARS2 crystal model (Figure 2B), and SARS2 sequence model (Figure 2C) were used for all superimpositions conducted. We observed that protein contact map superimposition between crystal model:crystal structure (Figure 2D), sequence model:crystal structure (Figure 2E), and sequence model:crystal model (Figure 2F) have high common contact (>90%), which implies that there is high contact similarity between the superimposed structures. We only considered SARS2 spike monomers when examining structural superimpositions. We also observed that RMSD values between cryo-EM model:crystal structure [RMSD 0.75] (Figure 2G), sequence model:cryo-EM structure [RMSD 0.66] (Figure 2H), and sequence model:cryo-EM model [RMSD 1.07] (Figure 2I) were RMSD < 1.5 which in-turn were considered adequate for further analyses (38). These results (both protein contact map and structural superimpositions) would suggest that the generated SARS2 spike models are suitable for further structural analyses.

Nine SARS2 FLC Structural Patterns Were Identified Among the Nine Node Clusters

Protein structure and conformation dynamics have often been correlated to biological function, which emphasizes the importance of protein structural pattern variations (23). To elucidate the possible SARS2 FLC structural variations among the 9 node clusters, representative SARS2 models from each node cluster were superimposed with the SARS2 model generated from MT019529 (Wuhan, China) as a comparison. Since SARS2 FLC also affects host tropism, SARS2 RBD was similarly checked.

As seen in Figure 3A, both SARS2 RBD (box dash lines) and FLC (box solid lines) structural changes were the focus of

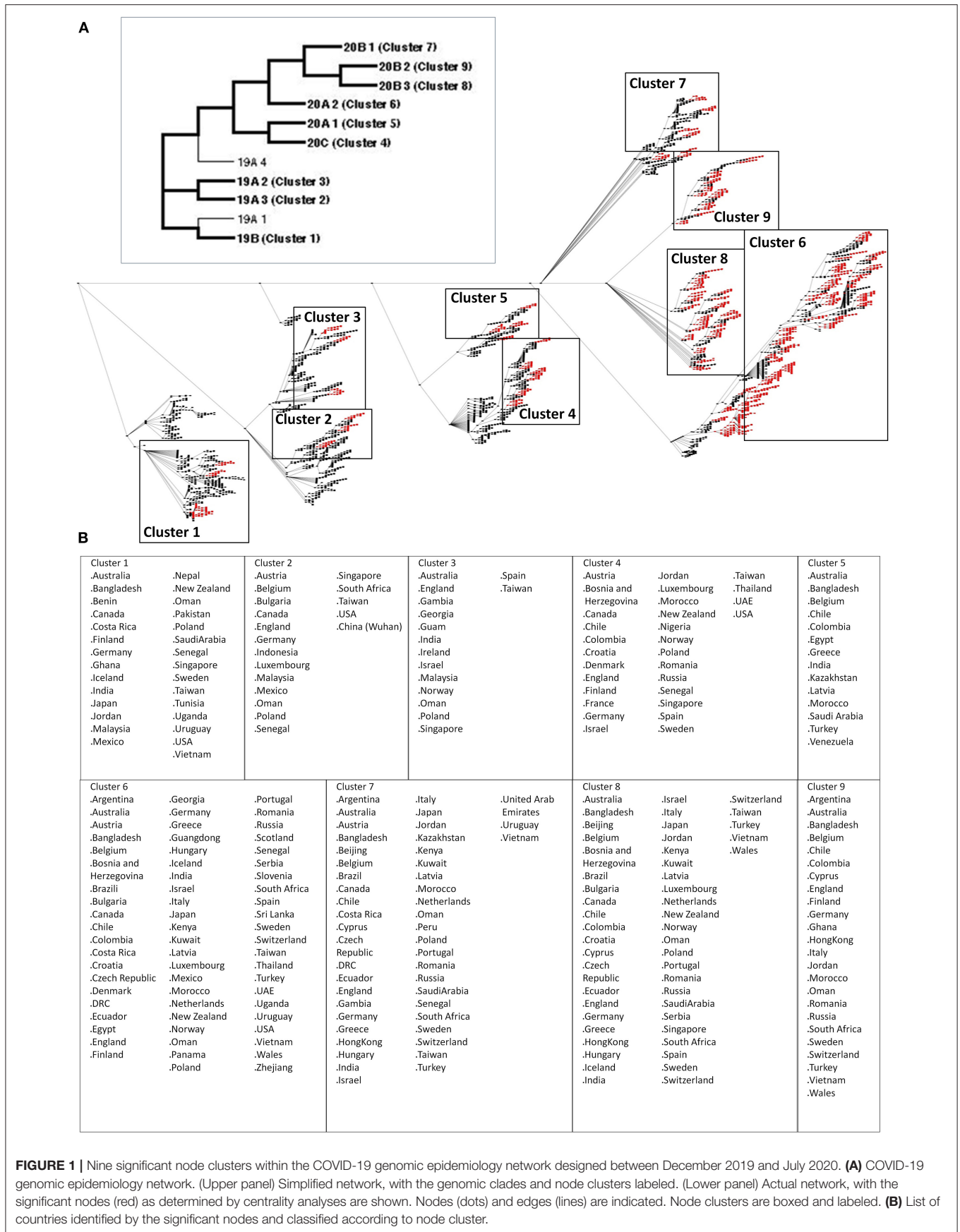
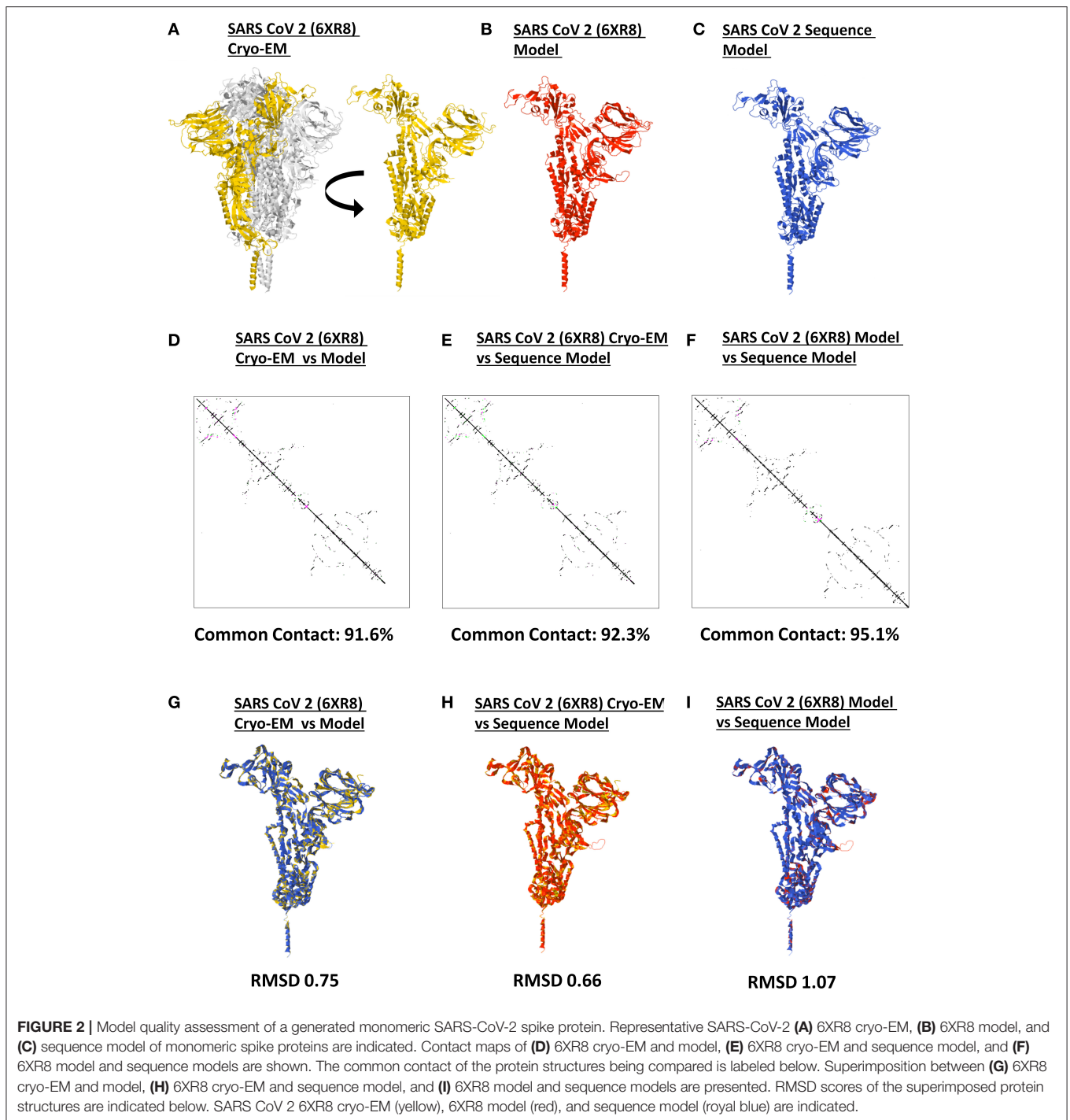


FIGURE 1 | Nine significant node clusters within the COVID-19 genomic epidemiology network designed between December 2019 and July 2020. **(A)** COVID-19 genomic epidemiology network. (Upper panel) Simplified network, with the genomic clades and node clusters labeled. (Lower panel) Actual network, with the significant nodes (red) as determined by centrality analyses are shown. Nodes (dots) and edges (lines) are indicated. Node clusters are boxed and labeled. **(B)** List of countries identified by the significant nodes and classified according to node cluster.



the study. Interestingly, we found nine SARS2 FLC structural patterns (Figures 3B–J, left panel), which coincidentally match with the nine node clusters identified earlier (Figure 1A). This insinuates that the SARS2 FLC structural pattern identified in each node cluster is a unique structural feature for the node cluster. However, we emphasize that the SARS2 FLC might not be the only factor determining the nine node clusters. In this regard and as possible future works, additional experimental

evidence is needed to further prove the presence of the nine SARS2 FLC structural patterns from the nine nodal clusters, and, equally important, it would be interesting to likewise determine other factors that may explain the presence of the nine node clusters. Subsequently, we observed that no structural changes occurred in the SARS2 RBD (Figures 3B–J, right panel). In all the superimpositions made, no significant structural changes (RMSD < 1.0; Tm align > 0.96) occurred between

superimposed SARS2 models (**Figures 3B–J, lower panel**), which is consistent with SARS2 maintaining its genomic integrity across propagation (34).

It was previously reported that the SARS2 FLC naturally undergoes polymorphisms, which in-turn affects viral transmissibility and tropism (39). In this regard, we suspect that the putative nine SARS2 FLC structural patterns are a product of natural polymorphism and, similarly, finding one of the SARS2 FLC structural patterns in one of the node clusters identified could suggest that certain countries (or continents) with overlapping node clusters may have varying levels of viral transmissibility and virulence (33, 34, 39). Since cleavage of the SARS2 FLC is a prerequisite for pathogenesis (17–19), we think that cleavage among the nine SARS2 FLC structural patterns may likewise vary (possibly depending on how exposed the FLC is), which in turn, could directly affect viral transmissibility. Additionally, with regards to host tropism, there seems to be no noticeable structural change in the SARS2 RBD, insinuating that host tropism is unchanged. This indicates that, regardless of any structural variations in SARS2 FLC, host tropism will not be consistently affected by genomic integrity (34). However, it is unclear whether the absence of SARS2 FLC (particularly ⁶⁸¹PRRA⁶⁸⁴) would affect SARS2 RBD.

SARS2 RBD Residues Did Not Change in the Absence of the Unique ⁶⁸¹PRRA⁶⁸⁴ Sequence

SARS2 has been reported to infect multiple species as well as humans due to variations in ACE2 receptors across species (40), which emphasizes the potential significance of the SARS2 RBD with regards to host tropism. Similarly, SARS2 FLC was found to likewise affect host tropism (17–19). This may suggest that SARS2 FLC (particularly ⁶⁸¹PRRA⁶⁸⁴) could affect SARS2 RBD. To establish the possible structural influence of the unique ⁶⁸¹PRRA⁶⁸⁴ amino acid sequence on SARS2 RBD structural orientation, we generated mutant SARS2 models with the unique ⁶⁸¹PRRA⁶⁸⁴ amino acid sequence deleted in all nine SARS2 FLC structural patterns and, subsequently, superimposed each mutant to the original model for comparison. This study undertook a side-by-side comparison of an original (*left panel*) and mutant (*right panel*) SARS2 model with a focus on SARS2 RBD (box dash lines) and FLC (box solid lines) structural changes (**Figure 4A**). As expected, in the absence of the ⁶⁸¹PRRA⁶⁸⁴ amino acid sequence we observed structural variations in the SARS2 FLC (**Figures 4B–J, left panel**). Nevertheless, no significant structural changes were observed (RMSD < 1.0; Tm align > 0.82) between superimposed original and mutated SARS2 models (**Figures 4B–J, lower panel**). Most surprisingly, no structural variations were observed in the SARS2 RBD (**Figures 4B–J, right panel**). This would suggest that SARS2 FLC (particularly ⁶⁸¹PRRA⁶⁸⁴) has no structural influence on SARS2 RBD, which is consistent with earlier works (41) that showed that SARS2 FLC may not be as critical as previously thought for the high fusion capacity of SARS2. However, it is worth mentioning that regions with high levels of the disorder typically do not have stable structures, and thus, would not have much of an effect on

the remaining structured parts of the protein (20) consistent with our observations. Taken together, the lack of a stable structure in the FLC site and its surroundings may explain why no structural changes occurred within the SARS2 RBD after the removal of a unique ⁶⁸¹PRRA⁶⁸⁴ region. Nevertheless, we presume that regardless of the absence of any structural variations within the SARS2 RBD, viral pathogenesis was unaffected since one important factor that determines virulence is high-affinity virus receptor interaction and, likewise, takes into account multiple host factors (40). This may explain why SARS2 infection in humans varies among COVID-19 infected patients. Additional experiments are needed to further prove this point.

DISCUSSION

SARS2 FLC is a conserved structural feature that is crucial for viral entry to host cells (39, 42) and, more importantly, can influence viral pathogenesis and host tropism (17–19, 40). In addition, the SARS2 FLC was found to have a naturally occurring polymorphism that can affect both transmissibility and host tropism (39). Throughout this study, we attempted to show that the SARS2 FLC has structural orientation variations putatively associated with the SARS2 genomic distribution particularly between December 2019 and July 2020.

SARS2 genome has continued to mutate since its emergence in December 2019 and SARS2 was found to have a >7.23 actual mutation rate with genetic changes occurring every other week (33, 34). These mutational changes are made possible through host-dependent RNA editing associated with the APOBEC mechanism (43). Cluster infections have also been associated with SARS2 incubation period infection and, likewise, play an important role in the rapid evolution of COVID-19 transmission (44, 45). This highlights how quickly the SARS2 genome is changing and, similarly, may explain how multiple variants of the virus can evolve easily and spread worldwide (33, 34). Several of the SARS2 nucleotide changes are nonsynonymous, thus, amino acid changes likewise occur (33) that may result in protein structural changes among SARS2 viral proteins. In particular, several structural changes have been reported with regards to the SARS2 spike protein (39, 42, 46, 47). Considering that we observed nine SARS2 FLC structural patterns from nine node clusters distributed worldwide, we postulate that this observation is putatively correlated to mutational changes that occurred within the SARS2 spike genome during the timeframe studied which in-turn affected the resulting amino acid sequence and, subsequently, lead to structural changes that may affect virulence and tropism.

It is worth mentioning that COVID-19 symptoms vary in the human population and, similarly, animal species (40). SARS2 infection in the human population often affects the lower respiratory tract (48) and follows a distinguishable order of symptom onset with varying levels of severity (49–51). COVID-19 reinfection has been clinically observed (52–56) and we suspect it is associated with varying SARS2 variants. In this regard, we hypothesize that COVID-19 reinfection could potentially be linked to SARS2 FLC structural variations

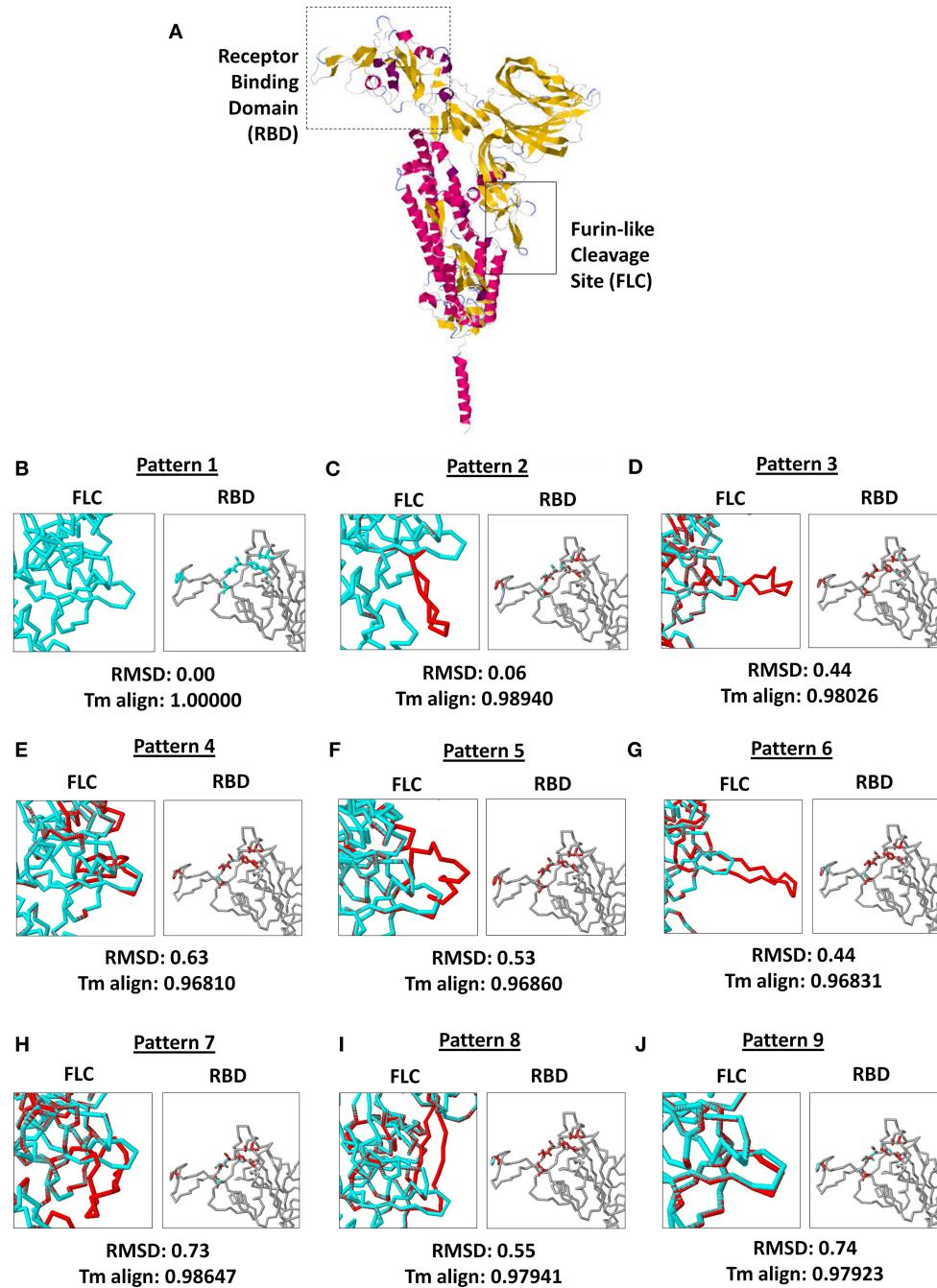
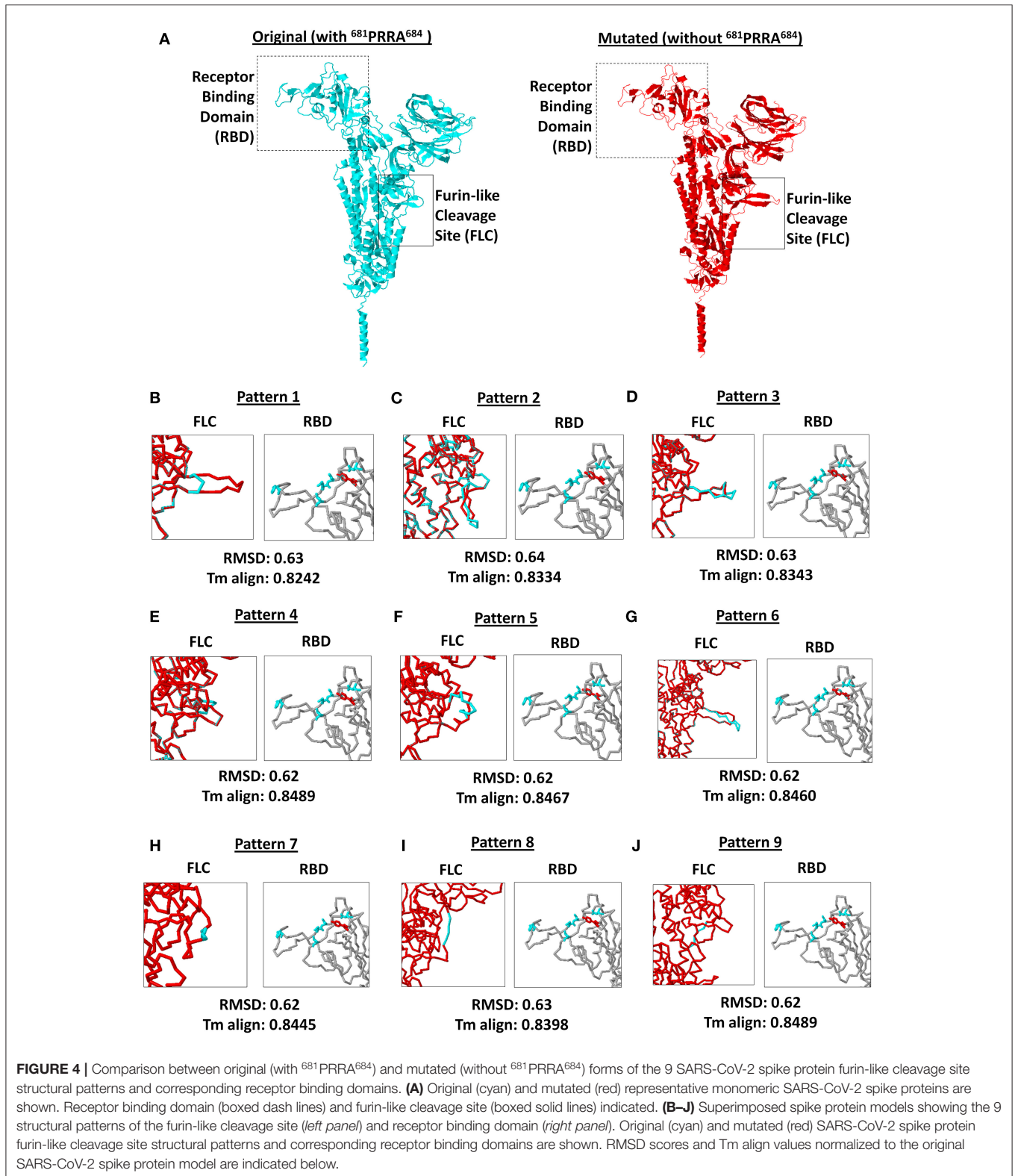


FIGURE 3 | Comparison of the 9 SARS-CoV-2 spike protein furin-like cleavage site structural patterns and corresponding receptor binding domains. **(A)** Representative monomeric SARS-CoV-2 spike protein model with the receptor binding domain (boxed dash lines) and furin-like cleavage site (boxed solid lines) indicated. **(B–J)** Superimposed spike protein models showing the nine structural patterns of the furin-like cleavage site (*left panel*) and receptor binding domain (*right panel*). Pattern 1 SARS-CoV-2 spike protein model (cyan) and the eight other structural patterns (red) are shown. RMSD scores and Tm align values normalized to Pattern 1 SARS-CoV-2 spike protein model are indicated below.

since SARS2 FLC affects viral pathogenesis, tropism, and transmissibility. Admittedly, additional experiments are needed to further prove this hypothesis.

In summary, we propose that between December 2019 and July 2020, nine SARS2 FLC structural patterns could

putatively correspond to the nine node clusters found within the COVID-19 genomic epidemiology network. Similarly, we associated this with the rapid evolution of the SARS2 genome. We observed that either in the presence or absence of the unique ⁶⁸¹PRRA⁶⁸⁴ amino acid sequence no



structural changes occurred within the SARS2 RBD, which we believe could mean that the SARS2 FLC has no structural

influence on SARS2 RBD and may explain why host tropism was maintained.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

MC and KI conceptualized the idea, provided feedback, helped in both structural and network analyses, and wrote the paper. MU, RI, and TH generated the protein models and analyzed the structural changes. YM, KY, NK, KW, and KB designed and analyzed the network. All authors contributed to the article and approved the submitted version.

REFERENCES

- Hamre D, Procknow JJ. A new virus isolated from the human respiratory tract. *Proc Soc Exp Biol Med.* (1966) 121:190–3. doi: 10.3181/00379727-121-30734
- Kapikian AZ, James HD Jr, Kelly SJ, Dees JH, Turner HC, McIntosh K, et al. Isolation from man of “avian infectious bronchitis virus-like” viruses (coronaviruses) similar to 229E virus, with some epidemiological observations. *J Infect Dis.* (1969) 119:282–90. doi: 10.1093/infdis/119.3.282
- Ksiazek TG, Erdman D, Goldsmith CS, Zaki SR, Peret T, Emery S, et al. A novel coronavirus associated with severe acute respiratory syndrome. *N Engl J Med.* (2003) 348:1953–66. doi: 10.1056/NEJMoa030781
- Fouchier RA, Hartwig NG, Bestebroer TM, Niemeyer B, De Jong JC, Simon JH, et al. A previously undescribed coronavirus associated with respiratory disease in humans. *Proc Natl Acad Sci USA.* (2004) 101:6212–6. doi: 10.1073/pnas.0400762101
- Woo PC, Lau SK, Chu CM, Chan KH, Tsoi HW, Huang Y, et al. Characterization and complete genome sequence of a novel coronavirus, coronavirus HKU1, from patients with pneumonia. *J Virol.* (2005) 79:884–95. doi: 10.1128/JVI.79.2.884-895.2005
- Zaki AM, Van Boheemen S, Bestebroer TM, Osterhaus AD, Fouchier RA. Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *N Engl J Med.* (2012) 367:1814–20. doi: 10.1056/NEJMoa1211721
- Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, et al. A novel coronavirus from patients with pneumonia in China, 2019. *N Engl J Med.* (2020) 382:727–33. doi: 10.1056/NEJMoa2001017
- Lu G, Wang Q, Gao GF. Bat-to-human: spike features determining ‘host jump’ of coronaviruses SARS-CoV, MERS-CoV, and beyond. *Trends Microbiol.* (2015) 23:468–78. doi: 10.1016/j.tim.2015.06.003
- Millet JK, Whittaker GR. Host cell proteases: critical determinants of coronavirus tropism and pathogenesis. *Virus Res.* (2015) 202:120–34. doi: 10.1016/j.virusres.2014.11.021
- Hulswit RJ, De Haan CA, Bosch BJ. Coronavirus spike protein and tropism changes. *Adv Virus Res.* (2016) 96:29–57. doi: 10.1016/bs.aivir.2016.08.004
- Li F. Structure, function, and evolution of coronavirus spike proteins. *Annu Rev Virol.* (2016) 3:237–61. doi: 10.1146/annurev-virology-110615-042301
- Moore JB, June CH. Cytokine release syndrome in severe COVID-19. *Science.* (2020) 368:473–4. doi: 10.1126/science.abb8925
- Tay MZ, Poh CM, Renia L, Macary PA, Ng LFP. The trinity of COVID-19: immunity, inflammation and intervention. *Nat Rev Immunol.* (2020) 20:363–74. doi: 10.1038/s41577-020-03111-8
- Wang N, Shang J, Jiang S, Du L. Subunit vaccines against emerging pathogenic human coronaviruses. *Front Microbiol.* (2020) 11:298. doi: 10.3389/fmicb.2020.00298
- Cai Y, Zhang J, Xiao T, Peng H, Sterling SM, Walsh RM Jr, et al. Distinct conformational states of SARS-CoV-2 spike protein. *Science.* (2020) 369:1586–92. doi: 10.1126/science.abd4251
- Ord M, Faustova I, Loog M. The sequence at Spike S1/S2 site enables cleavage by furin and phospho-regulation in SARS-CoV2 but not in SARS-CoV1 or MERS-CoV. *Sci Rep.* (2020) 10:16944. doi: 10.1038/s41598-020-74101-0

FUNDING

This work was supported by JSPS KAKENHI Grant Numbers 19K10078 and 19K10097 and Uemura Fund, Dental Research Center, Nihon University School of Dentistry.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmed.2021.613412/full#supplementary-material>

- Nao N, Yamagishi J, Miyamoto H, Igarashi M, Manzoor R, Ohnuma A, et al. Genetic predisposition to acquire a polybasic cleavage site for highly pathogenic avian influenza virus hemagglutinin. *MBio.* (2017) 8:e02298–16. doi: 10.1128/mBio.02298-16
- Andersen KG, Rambaut A, Lipkin WI, Holmes EC, Garry RF. The proximal origin of SARS-CoV-2. *Nat Med.* (2020) 26:450–2. doi: 10.1038/s41591-020-0820-9
- Coutard B, Valle C, De Lamballerie X, Canard B, Seidah NG, Decroly E. The spike glycoprotein of the new coronavirus 2019-nCoV contains a furin-like cleavage site absent in CoV of the same clade. *Antiviral Res.* (2020) 176:104742. doi: 10.1016/j.antiviral.2020.104742
- Giri R, Bhardwaj T, Shegane M, Gehi BR, Kumar P, Gadhave K, et al. Understanding COVID-19 via comparative analysis of dark proteomes of SARS-CoV-2, human SARS and bat SARS-like coronaviruses. *Cell Mol Life Sci.* (2020) 78:1655–88. doi: 10.1007/s00018-020-03603-x
- Jaimes J, Millet J, Whittaker G. Proteolytic cleavage of the SARS-CoV-2 spike protein and the role of the novel S1/S2 Site. *SSRN.* (2020) 3581359. doi: 10.2139/ssrn.3581359
- Kortemme T, Morozov AV, Baker D. An orientation-dependent hydrogen bonding potential improves prediction of specificity and structure for proteins and protein-protein complexes. *J Mol Biol.* (2003) 326:1239–59. doi: 10.1016/S0022-2836(03)00021-4
- Chen SC, Bahar I. Mining frequent patterns in protein structures: a study of protease families. *Bioinformatics.* (2004) 20 (Suppl. 1):i77–85. doi: 10.1093/bioinformatics/bth912
- Seyran M, Takayama K, Uversky VN, Lundstrom K, Palu G, Sherchan SP, et al. The structural basis of accelerated host cell entry by SARS-CoV-2 dagger. *FEBS J.* (2020). doi: 10.1111/febs.15651
- Gilman A, Arkin AP. Genetic “code”: representations and dynamical models of genetic components and networks. *Annu Rev Genomics Hum Genet.* (2002) 3:341–69. doi: 10.1146/annurev.genom.3.030502.111004
- Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* (2003) 13:2498–504. doi: 10.1101/gr.1239303
- Koschutski D, Schreiber F. Centrality analysis methods for biological networks and their application to gene regulatory networks. *Gene Regul Syst Bio.* (2008) 2:193–201. doi: 10.4137/GRSB.S702
- Zhang Y, Skolnick J. TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res.* (2005) 33:2302–9. doi: 10.1093/nar/gki524
- Kelley LA, Sternberg MJ. Protein structure prediction on the Web: a case study using the Phyre server. *Nat Protoc.* (2009) 4:363–71. doi: 10.1038/nprot.2009.2
- Herraez A. Biomolecules in the computer: Jmol to the rescue. *Biochem Mol Biol Educ.* (2006) 34:255–61. doi: 10.1002/bmb.2006.494034042644
- Vehlow C, Stehr H, Winkelmann M, Duarte JM, Petzold L, Dinse J, et al. CMView: interactive contact map visualization and analysis. *Bioinformatics.* (2011) 27:1573–4. doi: 10.1093/bioinformatics/btr163

32. Holm L, Sander C. Mapping the protein universe. *Science*. (1996) 273:595–603. doi: 10.1126/science.273.5275.595
33. Day T, Gandon S, Lion S, Otto SP. On the evolutionary epidemiology of SARS-CoV-2. *Curr Biol*. (2020) 30:R849–R857. doi: 10.1016/j.cub.2020.06.031
34. Mercatelli D, Giorgi FM. Geographic and genomic distribution of SARS-CoV-2 mutations. *Front Microbiol*. (2020) 11:1800. doi: 10.3389/fmicb.2020.01800
35. Przulj N, Malod-Dognin N. NETWORK ANALYSIS. Network analytics in the age of big data. *Science*. (2016) 353:123–4. doi: 10.1126/science.aah3449
36. Wuchty S, Stadler PF. Centers of complex networks. *J Theor Biol*. (2003) 223:45–53. doi: 10.1016/S0022-5193(03)00071-7
37. Berman HM, Burley SK, Chiu W, Sali A, Adzhubei A, Bourne PE, et al. Outcome of a workshop on archiving structural models of biological macromolecules. *Structure*. (2006) 14:1211–7. doi: 10.1016/j.str.2006.06.005
38. Hevener KE, Zhao W, Ball DM, Babaoglu K, Qi J, White SW, et al. Validation of molecular docking programs for virtual screening against dihydropteroate synthase. *J Chem Inf Model*. (2009) 49:444–60. doi: 10.1021/ci800293n
39. Xing Y, Li X, Gao X, Dong Q. Natural Polymorphisms Are Present in the Furin Cleavage Site of the SARS-CoV-2 Spike Glycoprotein. *Front Genet*. (2020) 11:783. doi: 10.3389/fgene.2020.00783
40. Sarkar J, Guha R. Infectivity, virulence, pathogenicity, host-pathogen interactions of SARS and SARS-CoV-2 in experimental animals: a systematic review. *Vet Res Commun*. (2020) 44:101–10. doi: 10.1007/s11259-020-09778-9
41. Xia S, Lan Q, Su S, Wang X, Xu W, Liu Z, et al. The role of furin cleavage site in SARS-CoV-2 spike protein-mediated membrane fusion in the presence or absence of trypsin. *Signal Transduct Target Ther*. (2020) 5:92. doi: 10.1038/s41392-020-0184-0
42. Walls AC, Park YJ, Tortorici MA, Wall A, McGuire AT, Veesler D. Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. *Cell*. (2020) 181:281–92 e286. doi: 10.1016/j.cell.2020.02.058
43. Di Giorgio S, Martignano F, Torcia MG, Mattiuz G, Conticello SG. Evidence for host-dependent RNA editing in the transcriptome of SARS-CoV-2. *Sci Adv*. (2020) 6:eabb5813. doi: 10.1126/sciadv.abb5813
44. Gao Y, Shi C, Chen Y, Shi P, Liu J, Xiao Y, et al. A cluster of the Corona Virus Disease 2019 caused by incubation period transmission in Wuxi, China. *J Infect*. (2020) 80:666–70. doi: 10.1016/j.jinf.2020.03.042
45. Liu T, Gong D, Xiao J, Hu J, He G, Rong Z, et al. Cluster infections play important roles in the rapid evolution of COVID-19 transmission: a systematic review. *Int J Infect Dis*. (2020) 99:374–80. doi: 10.1016/j.ijid.2020.07.073
46. Davidson AD, Williamson MK, Lewis S, Shoemark D, Carroll MW, Heesom KJ, et al. Characterisation of the transcriptome and proteome of SARS-CoV-2 reveals a cell passage induced in-frame deletion of the furin-like cleavage site from the spike glycoprotein. *Genome Med*. (2020) 12:68. doi: 10.1186/s13073-020-00763-0
47. Korber B, Fischer WM, Gnanakaran S, Yoon H, Theiler J, Abfalterer W, et al. Tracking Changes in SARS-CoV-2 spike: evidence that D614G Increases Infectivity of the COVID-19 Virus. *Cell*. (2020) 182:812–27 e819. doi: 10.1016/j.cell.2020.06.043
48. Liu SL, Saif L. Emerging Viruses without Borders: The Wuhan coronavirus. *Viruses*. (2020) 12:130. doi: 10.3390/v12020130
49. Grant MC, Geoghegan L, Arbyn M, Mohammed Z, McGuinness L, Clarke EL, et al. The prevalence of symptoms in 24,410 adults infected by the novel coronavirus (SARS-CoV-2; COVID-19): a systematic review and meta-analysis of 148 studies from 9 countries. *PLoS ONE*. (2020) 15:e0234765. doi: 10.1371/journal.pone.0234765
50. Koutsakos M, Kedzierska K. A race to determine what drives COVID-19 severity. *Nature*. (2020) 583:366–8. doi: 10.1038/d41586-020-01915-3
51. Larsen JR, Martin MR, Martin JD, Kuhn P, Hicks JB. Modeling the onset of symptoms of COVID-19. *Front Public Health*. (2020) 8:473. doi: 10.3389/fpubh.2020.00473
52. Bonifacio LP, Pereira APS, Araujo D, Balbao V, Fonseca B, Passos ADC, et al. Are SARS-CoV-2 reinfection and Covid-19 recurrence possible? A case report from Brazil. *Rev Soc Bras Med Trop*. (2020) 53:e20200619. doi: 10.1590/0037-8682-0619-2020
53. Gousseff M, Penot P, Gallay L, Batisse D, Benech N, Bouiller K, et al. Clinical recurrences of COVID-19 symptoms after recovery: Viral relapse, reinfection or inflammatory rebound? *J Infect*. (2020) 81:816–46. doi: 10.1016/j.jinf.2020.06.073
54. Madan M, Kunal S. COVID-19 reinfection or relapse: an intriguing dilemma. *Clin Rheumatol*. (2020) 39:3189. doi: 10.1007/s10067-020-05427-3
55. Parry J. Covid-19: Hong Kong scientists report first confirmed case of reinfection. *BMJ*. (2020) 370:m3340. doi: 10.1136/bmj.m3340
56. To KK, Hung IF, Chan KH, Yuan S, To WK, Tsang DN, et al. Serum antibody profile of a patient with COVID-19 reinfection. *Clin Infect Dis*. (2020) ciaa1368. doi: 10.1093/cid/ciaa1368

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Cueno, Ueno, Iguchi, Harada, Miki, Yasumaru, Kiso, Wada, Baba and Imai. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.