SARS-CoV-2 variants circulating in Vietnam, April 2020–October 2021

Nghiem Xuan Hoan*,,**
Phan Quoc Hoan*,,**
Nguyen Dang Manh*,,**
Le Huu Song*,**

**108 Institute of Clinical Medical and Pharmaceutical Sciences,*
***Vietnamese-German Center for Medical Research*

Summary

Objective: Several distinct severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) variants have emerged in different regions worldwide, including Vietnam. This study was conducted is to understand the molecularly genetic epidemiology of SARS-CoV-2 which has not yet been systematically investigated in Vietnamese clinical setting. *Subject and method:* We analyzed 671 Vietnamese full-length SARS-CoV-2 sequences available on the Global Initiative on Sharing All Influenza Data (GISAID) with data available by 19 October 2021. Sequence variation was done using CoVsurver mutations App (https://www.gisaid.org/epiflu-applications/covsurver-mutations-app/). Phylogenetic tree was built using nextclade algorithms. *Result:* Our report highlighted that there has been a substantial change in the molecular epidemiology of SARS-CoV-2 circulating through last three waves and ongoing wave of COVID-19. Currently, the Delta variant was dominant with widespread national level. The Alpha variant was almost disappeared and other VOCs, including Beta, Gamma were speculated to be not circulating yet in Vietnam. *Conclusion:* Surveillance of the emergent variants of SARS-CoV-2 requires an expanded research program to improve our understanding of emerging SARS-CoV-2 mutations profile and their impact on the protective immunity against variants with these mutations. In addition, our surveillance for molecular epidemiology of SARS-CoV-2 in Vietnam will contribute to the efforts at the global levels to fight the pandemic.

Keywords: SARS-CoV-2, covid-19, mutation, deletion, variants, Vietnam.

1. Background

During the Covid-19 pandemic, several distinct severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) lineages have emerged in different regions worldwide. We have seen that several variants of SARS-CoV-2 have emerged, spread, and changed rapidly. To date, four strains are listed as variants of global concern (VOCs): Alpha (B.1.1.7), Beta (B.1.351), Gamma (P.1) and Delta (B.1.617.2), which associate with increased transmissibility, disease severity and/or possible reduced vaccine efficacy [1].

As of 19 October 2021, Covid-19 cases caused by Alpha variant have been reported from 196 countries, Beta variant from 145 countries, Gamma variant from 99 countries, and Delta variant from 193 countries across all six WHO regions [2]. Even broader classes termed as variants of interest (VOI) and variants under investigation (VUI) include

Received: 29 October 2021, Accepted: 16 December 2021
Correspondence to: *Nghiem Xuan Hoan - Department of Molecular Biology, 108 Institute of Clinical Medical and Pharmaceutical Sciences, Hanoi, Vietnam*
Email: *nghiemxuanhoan@108-icid.com.*

lineages suspected of having altered phenotypes designated by their mutation profile. In addition, new strains still continue to emerge worldwide and may be reported as VOI and VUI variants if they are considered to have concerned epidemiological, immunological or pathogenic potential and thus to be closely monitored and investigated.

As evidence becomes available, classification for VOCs will be revised to reflect the continuous evolution of circulating variants and their changing epidemiology. Molecularly genetic epidemiology of SARS-CoV-2 has changed rapidly from the originating Wuhan strains to a predominance of Delta variant and other VOCs, with declining prevalence of other variants among SARS-CoV-2 sequences submitted to publicly available datasets (https://www.gisaid.org/). In addition to VOCs, many VOIs and VUIs variants are emerging, which share mutations and biological features. Genetic changes in genome, significantly in spike gene define the biological properties of the virus. As evident, the Spike protein is ultimately responsible for receptor binding, membrane penetration and also the main target for neutralizing antibodies. Thus, monitoring the emergence of new variants of SARS-CoV-2 is a priority worldwide as the presence of important non-synonymous amino acid substitutions as well as Insertion–deletion mutations play an in important role in the biological properties.

Vietnam experienced four waves of COVID-19: wave 1 from 23 January-16 April 2020 with 106 cases in the community; wave 2 from 25 July to 1 December 2020 with 554 cases; wave 3 from 28 January to 25 March 2021 with 910 cases and wave 4 from April 2021-present with a total cumulative number of 870255 laboratory confirmed cases from all 63 provinces including 21344 deaths (2.45%); of those approximately 794846 had recovered (91.3%). At the time of writing, there are currently 36908 people being monitored and treated of those 5014 are severe cases (https://www.worldometers. info/coronavirus).

The COVID-19 situation in Vietnam has changed and increased rapidly since January 2021 due to the scenarios of emergence of vulnerable VOC variants. To better reflect the recent changes and current geographical distribution of SARS-CoV-2 variants and to characterize the mutation profile, especially in the spike gene, we downloaded and analyzed all available Vietnamese SARS-CoV-2 sequences to the Global Initiative on Sharing All Influenza Data (GISAID) (https://www. gisaid.org/) to understand the molecularly genetic epidemiology of SARS-CoV-2 which has not yet been systematically investigated in our clinical setting with data available by 19 October 2021.

2. Subject and method

Sequence retrieval, alignment and mutation analysis

To decipher the genetic variations, we retrieved 695 sequences of Vietnamese SARS-CoV-2 strains available at the Global Initiative on Sharing All Influenza Data (GISAID) (https://www.gisaid.org/) by 19 October 2021. We excluded incomplete full-length and low coverage sequences from 695 sequences. Finally, 671 complete full-length genome sequences were selected. We aligned the full-length sequences of SARS-CoV-2 by MAFFT online server [3] using default parameters, and the complete genome sequence hCoV-19/Wuhan/WIV04/2019 (Genbank: MN996528.1) was used as a reference genome. For each whole genome sequence, the GISAID clade assignment was done using CoVsurver mutations App (https://www.gisaid.org/epiflu-applications/ covsurver mutations-app/). Phylogenetic tree was built using nextclade algorithms (https://clades.nextstrain.org/).

3. Result and discussion

Demographic characteristics of 671 Vietnamese SARS-CoV-2 positive patients

The Figure 1 depicts the demographic characteristics of patients who provided SARS-CoV-2 whole genome sequences. Uploaded whole

genome sequences from patients from several provinces across Vietnam but larger proportions of patients lived in Ho Chi Minh and Hanoi city. As seen in the Figure 1A, there no significant difference of SARS-CoV-2 infection between men and women in each subgroup classified by WHO. The results could indicate that gender is not a susceptibility factor for SARS-CoV-2 infection as reported as the real situation of COVID-19 [4]. Most cases included in this survey aged between 18-60 years accounting for approximately 75% (Figure 1A).

SARS-CoV-2 variants emerged in 2020 belonged to the non-Alpha/-Delta clades (others), which appeared in the early stages of the pandemic include 19A, 19B, 20A, 20B, 20C, 20D, 20E as shown later in the Figure 1. Only one case infected with VOC Alpha was detected among the 671 patients identified in December 2020. Subsequently, variants emerged in January 2021 were all Alpha but were speedily replaced by Delta variants from April 2021 on with no Alpha variants identified. In total, Delta (GK clade according to the GISAID classification), Alpha variants (GRY) that responsible for 75.4% and 3.7%, respectively and 20.9% of strains belonged to the other clades (19A, 19B, 20A, 20B, 20C, 20D, 20E according to the nextclade classification - Figure 1C and 1D).

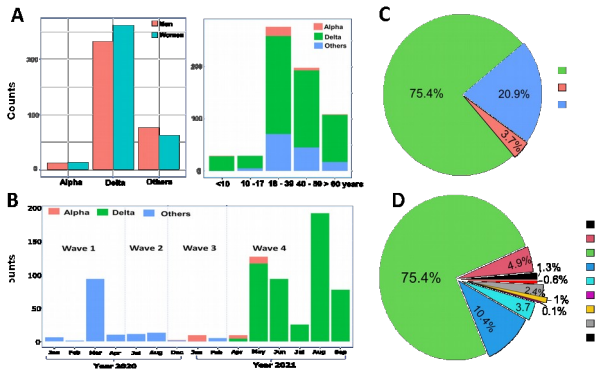

Figure 1. Distribution of 671 full-length sequences of SARS-CoV-2

(A). Distribution of SARS-CoV-2 variants according to the gender and age groups (A) and according to the waves of Covid-19 (B) and the proportion of SARS-CoV-2 variants according to the WHO classification (C) and GISAID clades (D).

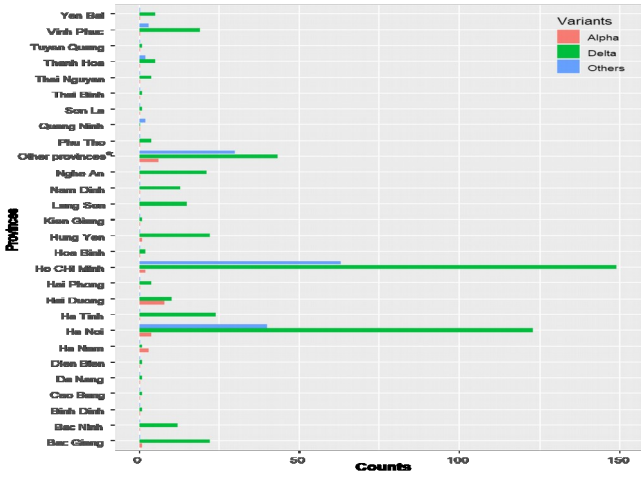

Figure 2. Distribution of different SARS-CoV-2 variants across Vietnam

Phylogenetic analysis of 671 full-length genome sequences of SARS-CoV-2

The Figure 3A shows the phylogenetic tree that was built based on analysis of whole genome strains from different regions/countries to differentiate the clades or variants VOC Delta, Alpha, Beta, Gamma and other variants. In the Figure 3B that is only visible for Vietnamese strains), analyzed the 671 Vietnamese whole genome sequences of SARS-CoV-2, which are available on the GISIAD (on 19th October 2021) presented the different clades along with a number of amino acid mutations on different genes of the virus. We identified 11 clades with a dominance of 21I that accounted for 73.5% (493/671) cases.

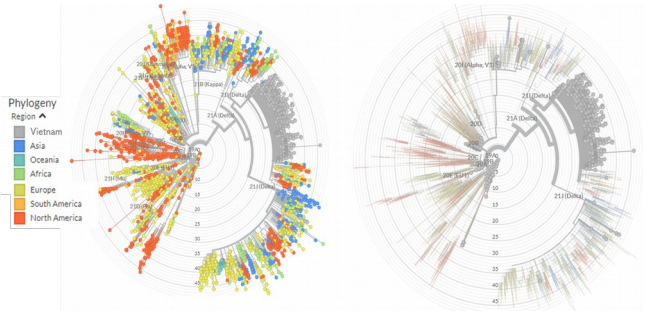

Figure 3. Phylogenetic tree of whole genome sequences of SARS-CoV-2 from different countries and Vietnam

(A). Phylogenetic tree built by nextclade algorithm using all strains from different countries presenting the different clades. Vietnamese strains are in grey color.

(B) Only Vietnamese strains are visible and mostly defined as VOC Delta. A smaller proportion is allocated as VOC Alpha and others including clades 19A, 19B, 20A, 20B, 20C, 20D, 20E (or non-VOCs).

Mutation's profile of SARS-CoV-2 circulating in Vietnam

Since its emergence in late 2019, SARS-CoV-2 has diversified into several different co-circulating variants. Several mutations in different regions of the SARS-CoV-2 genome emerged overtime and across variants or clades with significantly higher number of mutations, including nucleotide substitutions, amino acid substitutions, and amino acid deletions in VOC Alpha (20I) and VOC Delta (21A, 21I, 21J) variants compared to other variants (Figure 4A & 4B, Figure 5). Variants were grouped into clades which are defined by specific signature mutations. At the time or writing, 22 major clades are defined (the clades 19A and 19B are close to the original Wuhan strains during early pandemic in China and have been dominating the early outbreak. Thereafter, clade 20A emerged from 19A out of dominated the European outbreak in March and has since spread globally. 20B and 20C are large genetically distinct subclades 20A emerged in early 2020. 20D to 20J have emerged globally over the summer of 2020 and include three VOCs: 20H (Beta), 20I (Alpha), 20J (Gamma). 21A to 21F include the VOC delta and several variants of interest.

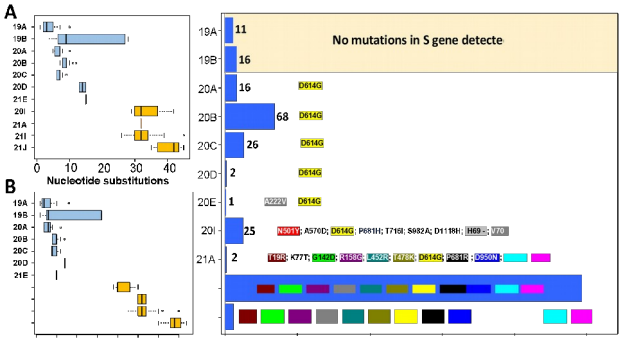

Figure 4. Distribution of mutations and specific amino acid mutations
in spike protein according to the nextclades classification.

Figure 4A and 4B show the number of nucleotide substitutions and amino acid substitutions according to the different clades. Data were presented by median and range (min, max). The Figure 4C shows the distribution of specific substitution mutations in the Spike protein in distinguishing the different SARS-CoV clades.

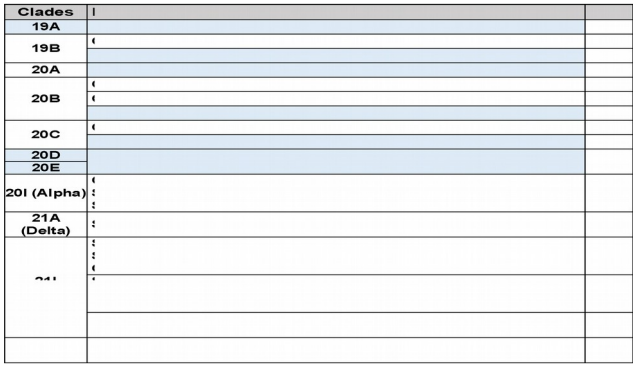

Figure 5. Deletion mutations in different regions of SARS-CoV-2 genome

Distribution of specific amino acid deletions in the ORFs and in the Spike protein which can combine with other mutations in different regions of full-length sequences in discriminating the different variants.

We analyzed the whole genome of 671 SARS-CoV-2 sequences and mutations on the spike gene, including amino acid substitutions and deletions were detailed in the Figure 4C and Figure 5. The D614G mutation was not existed in the 19A and 19B clades, which are close to the original Wuhan strains during early pandemic in China and have been dominating the early outbreak. As stated in several previous studies, we also reported that the D614G mutation emerged and existed in all remaining clades. There have been immense evidences showing that the D614G mutation significantly increases infectivity and transmissibility of the virus by mean of several mechanisms [5-8]. This mutation locates closely to the RBD region playing a crucial role in the fusion mechanism of virus into targeted host cells (Figure 6). Upon viral binding to the angiotensin-converting enzyme 2 (ACE2) receptor, the spike protein undergoes cleavage at the S1/S2 junction and the two subunits dissociate. This activates a series of conformational changes that lead to the fusion of viral and cellular membranes to

initiate infection. The mutant viruses are more resistant to proteolytic cleavage during production of the protein in host cells as well as on pseudotyped lentiviral particles, suggesting that replicated virus produced in human cells may be more infectious due to a greater proportion of functional spike protein per virion (uncleaved with receptor-binding domain) [9]. Additional evidence showed that SARS-CoV-2 spike glycoprotein D614G increases the infection efficiency of cellular entry for the virus across a broad range of human cell types, including cells from lung, liver, and colon [9].

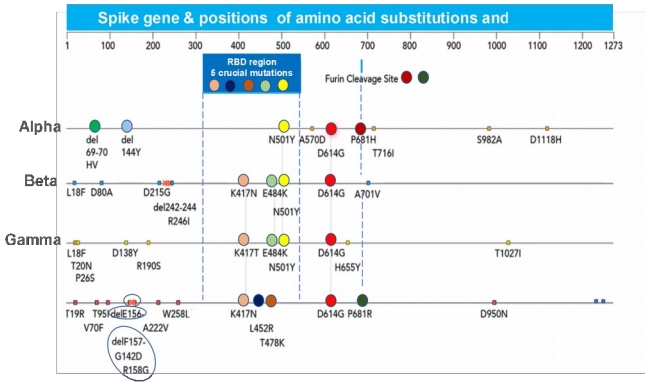

Figure 6. Mutations and positions in the Spike protein in VOCs variants (adapted from https://covariants.org/variants)

Mutations in the receptor binding domain (RBD) are extremely important to ACE2 binding and antibody recognition, and neutralizing activity [10]. Five amino acid substitutions found in this region are K417N, L452R, T478K, E484K, and N501Y. These specific mutations are found in the VOCs variants reported at the end of 2020 including Alpha, Beta, Gamma, and Delta (Figure 6). In our current report, the N501Y mutation is uniquely found in the Alpha variant as Beta and Gamma were not identified in all 671 uploaded sequences. This specific mutation can enhance virus's affinity to the human ACE2 receptor [11, [12] and the 501Y mutant virus exhibited consistent fitness gains for replication in the upper airway in the hamster model as well as primary human airway epithelial cells [13].

The Delta variant with mutations at the main locations in RBD (K417N, L452R, T478K) and several combined mutations in other regions of S protein as T19R, G142D, R158G, A222V, D614G, P681R, D950N has adversely affected the areas of many regions over the world and are currently dominating in Vietnam.

Among these common and important aa substitution mutations, the G142D mutation in the N-terminal domain is an escape mutant to some antibodies and has appeared in viruses grown in the presence of a monoclonal antibody [14, 15]. Mutations S: P681H and P618R have emerged in the VOCs Alpha and Delta as well as in several circulating VOIs (https://outbreak.info). The position of these mutations P681H/R is adjacent to the S1/S2 furin cleavage site which is an important region for SARS-CoV-2 transmission. The previous study has shown that Delta variants with P681R mutation augments SARS-CoV-2 fitness over Alpha variant with P681H mutation [16]. In addition, functional studies confirm that P681R mutation facilitates the furin-mediated spike cleavage and enhances viral infectivity and the efficacy of viral fusion and cell-cell viral spread, particularly when it occurs in the background of other S changes [17] In this study, we have shown that the P681H was found only in Alpha variant (20I) but not in other variants. However, P681R was only found in Delta variants.

Regarding the amino acid deletion profile, there were no amino acid deletions identified in all non-VOC variants (or clades: 19A, 19B, 20A, 20B, 20C, 20D, 20E). In the Alpha variants, the important deletion mutations in spike protein were H69-, V70-, and Y144- and in Delta variants were E156-, F157-, Y144- (Figure 5 and Figure 6). Alpha variants bear six aa deletions in all 25 uploaded seuqences that include S3675-; G3676-; F3677- in ORF1a and H69-; V70-; Y144- in S protein. While Delta variants are featured by specific amino acid deletions including E156-; F157- in S protein and D119- ; F120- in ORF8a. The mutation H69-/V70- in the S1 N-terminal domain of the spike protein may alter the recognition by antibodies, possibly impacting some antibody-therapy treatments, or immunity [18],

whereas the Y144- is associated with antibody escape [19]. The deletion mutations E156- and F157- in S protein are cliamed to associate with increased immune evasion [20].

4. Conclusion

In conclusion, our report highlighted that there has been an obvious change in the molecular epidemiology of SARS-CoV-2 circulating in Vietnam. Currently, the Delta variant is dominant with widespread national level as also seen at the global levels. The Alpha variant is still circulating in Vietnam, however, at a small proportion. Other VOCs, including Beta, Gamma are considered to be absent in Vietnam till date. We have seen that several variants of SARS-CoV-2 have emerged and spread rapidly, thus at the national level, surveillance of the emergent variants of SARS-CoV-2 requires an expanded research program to improve our understanding of emerging SARS-CoV-2 mutations and the correlates of protective immunity against variants with these mutations. In addition, our surveillance for molecular epidemiology of SARS-CoV-2 in Vietnam will contribute to the efforts at the global levels to fight the pandemic.

Declarations

Funding

This study does not have funding

Data availability

We used the complete genome sequence hCoV-19/Wuhan/WIV04/2019 strain (Genbank: MN996528.1) as a reference genome. The annotated genomes (n=671) have been retrieved from publicly open database (GISAID: https://www.gisaid.org/) up to October 19, 2021. Further information regarding the accession numbers of the studied genomes could be found in the Supplementary data file.

Acknowledgements

We gratefully acknowledge the Authors from the Originating laboratories responsible for obtaining the specimens, as well as the Submitting laboratories where the genome data were generated and shared via GISAID, on which this report is based

References

1. WHO (2021) Tracking SARS-CoV-2 variants.
2. WHO (2021) *Weekly epidemiological update on COVID-19 - 19 October 2021*.
3. Katoh K, Misawa K, Kuma K, Miyata T (2002) *MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform.* Nucleic Acids Res 30(14): 3059-3066. doi: 10.1093/nar/gkf436.
4. Ahmed SB, Dumanski SM (2020) *Sex, gender and COVID-19: A call to action.* Can J Public Health. 111(6): 980-983. doi: 10.17269/s41997-020-00417-z.
5. Korber B, Fischer WM, Gnanakaran S, Yoon H, Theiler J, Abfalterer W et al (2020) *Tracking changes in SARS-CoV-2 spike: Evidence that D614G increases infectivity of the COVID-19 virus*. Cell 182(4): 812-827. doi: 10.1016/j.cell.2020 .06.043.
6. Yang TJ, Yu PY, Chang YC, Hsu SD (2021) *D614G mutation in the SARS-CoV-2 spike protein enhances viral fitness by desensitizing it to temperature-dependent denaturation.* J Biol Chem: 101238. doi: 10.1016/j.jbc.2021.101238.
7. Zhang L, Jackson CB, Mou H, Ojha A, Peng H, Quinlan BD et al (2020) *SARS-CoV-2 spike-protein D614G mutation increases virion spike density and infectivity.* Nat Commun 11(1): 6013. doi: 10.1038/s41467-020-19808-4.
8. Binquan Luan HW, Tien Huynh (2021) *Molecular mechanism of the N501Y Mutation for enhanced binding between SARS-CoV-2's spike protein and human ACE2 receptor.* BioRxiv. doi: 10.1101/2021.01.04.425316.
9. Daniloski Z, Jordan TX, Ilmain JK, Guo X, Bhabha G, tenOever BR et al (2021) *The Spike D614G mutation increases SARS-CoV-2 infection of multiple human cell types.* Elife 10. doi: 10.7554/eLife.65365.
10. Lu L, Chu AW, Zhang RR, Chan WM, Ip JD, Tsoi HW et al (2021) *The impact of spike N501Y mutation on neutralizing activity and RBD binding of SARS-CoV-2 convalescent serum*. EbioMedicine 71: 103544. doi: 10.1016/j.ebiom.2021.103544.

11. Tian F, Tong B, Sun L, Shi S, Zheng B, Wang Z et al (2021) *N501Y mutation of spike protein in SARS-CoV-2 strengthens its binding to receptor ACE2.* Elife 10. doi: 10.7554/eLife.69091.
12. Liu Y, Liu J, Plante KS, Plante JA, Xie X, Zhang X et al (2021) *The N501Y spike substitution enhances SARS-CoV-2 transmission.* bioRxiv. 2021. doi: 10.1101/2021.03.08.434499.
13. Liu Y, Liu J, Plante KS, Plante JA, Xie X, Zhang X, Ku Z, An Z, Scharton D, Schindewolf C, Menachery VD, Shi PY, Weaver SC (2021) *The N501Y spike substitution enhances SARS-CoV-2 transmission.* bioRxiv [Preprint]. 2021 Mar 9:2021.03.08.434499. doi: 10.1101/2021.03.08.434499.
14. McCallum M, De Marco A, Lempp FA, Tortorici MA, Pinto D, Walls AC, Beltramello M, Chen A, Liu Z, Zatta F, Zepeda S, di Iulio J, Bowen JE, Montiel-Ruiz M, Zhou J, Rosen LE, Bianchi S, Guarino B, Fregni CS, Abdelnabi R, Foo SC, Rothlauf PW, Bloyet LM, Benigni F, Cameroni E, Neyts J, Riva A, Snell G, Telenti A, Whelan SPJ, Virgin HW, Corti D, Pizzuto MS, Veesler D (2021*) N-terminal domain antigenic mapping reveals a site of vulnerability for SARS-CoV-2*. Cell. 184(9):2332-2347.e16. doi: 10.1016/j.cell.2021.03.028.
15. Suryadevara N, Shrihari S, Gilchuk P, VanBlargan LA, Binshtein E, Zost SJ et al (2021) *Neutralizing and protective human monoclonal antibodies recognizing the N-terminal domain of the SARS-CoV-2 spike protein.* Cell 184(9): 2316-2331. doi: 10.1016/j.cell.2021.03.029.
16. Liu Y, Liu J, Johnson BA, Xia H, Ku Z, Schindewolf C et al (2021) *Delta spike P681R mutation enhances SARS-CoV-2 fitness over Alpha variant*. bioRxiv.. doi: 10.1101/2021.08.12.456173.
17. Akatsuki Saito TI, Rigel Suzuki, Tadashi Maemura, Hesham Nasser, Keiya Uriu, Yusuke Kosugi, Kotaro Shirakawa, Kenji Sadamasu, Izumi Kimura, Jumpei Ito, Jiaqi Wu, Kiyoko Iwatsuki-Horimoto, Mutsumi Ito, Seiya Yamayoshi, Seiya Ozono, Erika P Butlertanaka, Yuri L Tanaka, Ryo Shimizu, Kenta Shimizu, Kumiko Yoshimatsu, Ryoko Kawabata, Takemasa Sakaguchi, Kenzo Tokunaga, Isao Yoshida, Hiroyuki Asakura, Mami Nagashima, Yasuhiro Kazuma, Ryosuke Nomura, Yoshihito Horisawa, Kazuhisa Yoshimura, Akifumi Takaori-Kondo, Masaki Imai, So Nakagawa, Terumasa Ikeda, Takasuke Fukuhara, Yoshihiro Kawaoka, Kei Sato (2021) *SARS-CoV-2 spike P681R mutation, a hallmark of the Delta variant, enhances viral fusogenicity and pathogenicity*. bioRxiv 448820; doi: https://doi.org/10.1101/2021.06.17.448820.
18. Kemp SA, Collier DA, Datir RP, Ferreira I, Gayed S, Jahun A et al (2021) *SARS-CoV-2 evolution during treatment of chronic infection.* Nature 592(7853): 277-282. doi: 10.1038/s41586-021-03291-y.
19. McCallum M, De Marco A, Lempp FA, Tortorici MA, Pinto D, Walls AC et al (2021) *N-terminal domain antigenic mapping reveals a site of vulnerability* for *SARS-CoV-2.* Cell 184(9): 2332-2347 e16. doi: 10.1016/j.cell.2021.03.028.
20. Armi M Chaudhari, Dinesh Kumar, Madhvi Joshi, Amrutlal Patel, Chaitanya Joshi (2021) *E156G and Arg158, Phe-157/del mutation in NTD of spike protein in B.1.617.2 lineage of SARS-CoV-2 leads to immune evasion through antibody escape.* BioRxiv. 2021. doi: https://doi.org/10.1101/2021.06.07.447321.