

International Journal of Biosciences | IJB | ISSN: 2220-6655 (Print), 2222-5234 (Online) http://www.innspub.net Vol. 21, No. 5, p. 83-92, 2022

RESEARCH PAPER

OPEN ACCESS

Variations in spike gene of SARS-CoV-2 isolated in Vietnam

Quan Ke Thai^{*1}, Phuoc Huynh², Duc Minh Le¹

¹Saigon University, 273 An Duong Vuong, Ward 3 District 5, Ho Chi Minh city, Vietnam ²VNU HCMC University of Science, 227 Nguyen Van Cu, Ward 4 District 5, Ho Chi Minh city, Vietnam

Key words: Vietnam, Covid-19, Genetic diversity, SARS-CoV-2, Spike gene

http://dx.doi.org/10.12692/ijb/21.5.83-92

Article published on November 06, 2022

Abstract

The S gene encodes the spike protein, which plays an important role in the life cycle of SARS-CoV-2. Many mutations in the S gene demonstrate the increase in the transmission rate, helping viral can escape the immune system and monoclonal antibodies. This study aims to track the genetic diversity of the S gene of SARS-CoV-2 isolated in Vietnam during the four waves of the pandemic. A total of 5478 S gene sequences were extracted and classified into 1093 haplotypes, giving the mean number of pairwise differences at 27.491490 ± 12.039763 and the haplotype diversity index at 0.9089 ± 0.0031 . In addition, we recorded the nucleotide diversity index at 0.007176 ± 0.003475 with 802 polymorphic sites. We found that indel mutations appear mainly in the NTD coding region, and amino acid substitutions appear most in the RBD coding region. In the RBD region, we also recorded mutations that increase the transmission rate for SARS-CoV-2 with a large proportion in Vietnam: G22578A (G339D), C22686T (S375F), G22813T (K417N), T22917G (L452R), C22995A (T478K) and A23063T (N501Y). In addition, the furin cleavage site also recorded three important mutations in enhancing transmissibility namely A23403G (D614G) and C23604A/G (P681H/R). The mean nucleotide difference between haplotypes in the cluster of Delta variants was 1.989772 ± 1.123948 nucleotides. The Omicron cluster has a tremendous nucleotide difference of 17.951297 ± 7.970441 nucleotides. The analysis results showed the whole picture of variations in the spike gene of SARS-CoV-2 in Vietnam, supporting the management of new variants imported into Vietnam.

* Corresponding Author: Quan Ke Thai 🖂 tkquan@sgu.edu.vn.

Introduction

More than two years, from the first case infection of a novel coronavirus in Wuhan, China, to the worldwide COVID-19 pandemic. The emergence of SARS-CoV-2 led to hundreds of millions of infections and mortality worldwide. Similarly, SARS-CoV and MERS-CoV, SARS-CoV-2 belong to the Betacoronavirus genus and is the seventh member of the Coronaviridae family infected with humans. However, the parameters of risk factors for viral transmission and hospitalization of SARS-CoV-2 are higher than SAR-CoV and MERS-CoV (Zhou et al., 2021). SARS-CoV-2 contains ssRNA+ without replication repair systems, leading to a high mutation rate. During the transmission, SARS-CoV-2 evolved to increase adaptation with the host by introducing and storing positive mutations. Thus, numerous dangerous variants of SARS-CoV-2 bear the fitness more transmission and escape the immune system. Understanding the riskiness, WHO had to track the changes in the genome and, dependent on that, classify novel variants of SARS-CoV-2 into three groups: VOI, VUM, and VOCs. Among those, the most notable is the VOC because more evidence indicates this variant's fitness, and more transmission and implication escape the monoclonal antibody and vaccine (WHO). The first VOCs is Alpha, a demonstrated variant with a transmission rate higher than the other viral isolated at that time by previous research (Walker et al., 2021) (Lyngse et al., 2021). After that, the emergence of the Delta variant raised horrible waves of pandemics worldwide. The United States of America and India were mentioned as the two countries with the worst impact by the Delta variant (Yang and Shaman, 2022) (Bolze et al., 2022). The success of worldwide vaccination strategies may promote the recession in Delta. Currently, a new variant was put in VOC - Omicron, first identified in South Africa in December 2021. The Omicron variant holds the capability to escape the immune system and vaccine, thereby causing reinfections. Recently, Omicron has gradually taken over Delta and become the dominant variant worldwide (Chaguza et al., 2022). In the early stage, despite COVID-19 destroying the pandemic preventions of countries, Vietnam was one of the few countries controlling the

SARS-CoV-2 spread in the community. Undergoing three waves of COVID-19, Vietnam only recorded about 1500 cases of confirmed SARS-CoV-2 infection and 35 fatalities (Minh et al., 2021) despite the emergence of dangerous variants - A.23.1 and Alpha (B.1.1.7) (Chau et al., 2021). However, the appearance of the Delta variant raised the country's fourth wave of the pandemic and was the primary cause of breaking through the government's COVID-19 prevention strategies. The number of death by the Delta variant accounted for 99% of total deaths (Hoang et al., 2022), and the statistics of case-fatality ratio (CFR) was 2.50% in Vietnam is higher than the average in the world (Minh, Khoi Quan et al., 2021). The successful vaccine strategies supported reducing the community's gradual transmission and fatality rates. Therefore, the Vietnamese Government eased the measures of controlling the pandemic situation and established a "new normal" societal condition. In this condition, easy travel is the best opportunity for novel variants of SARS-CoV-2 entry into the country. The first case was a confirmed infection of Omicron variants in Vietnam on December 19, 2021. The emergence of Omicron variants invigorated for the fourth wave is still going on, and this variant will replace Delta to become dominant in Vietnam. Quitely, the Omicron variant was the main infection source in Hanoi capital in March 2022. At the same time, Vietnam noted 454,179 cases in a day. This is more numerous infection cases than in the period October to November 2021 (According Ministry of Health). It is shown that intimidation to the health of Omicron is very considered. The beneficial fitness of Omicron could be from the amino acid substitutions in the spike protein.

The high level of mutation rate is the primary motivation for the changes in the S gene that encoded the spike protein plays an important role in the life cycle because it directly interacts with hACE2 in the human cell (Ovsyannikova *et al.*, 2020). Previous studies to monitor the nucleotide changes in the genome of SARS-CoV-2 indicated that the S gene had a higher nucleotide substitution rate than other gene regions (Berrio *et al.*, 2020). Especially, the gene S

might be a positive selection gene while the opposite of the other genes (Berrio, Gartner et al., 2020). Indeed, the most dangerous variant - Delta bore double mutants (L452R, T478K) in receptor binding domain and P681R in furin cleavage site were demonstrated as the main mutations increasing the transmission rate of viral (Kumar et al., 2021) (Fan et al., 2021). Currently, the omicron variant contains more than 30 amino acid substitutions in this protein (Wei et al., 2021) (Kumar et al., 2022) that supply omiron's ability to break through the immune system and make the reinfection events. Moreover, spike protein is a gold target for developing vaccines. Thus, tracking the nucleotide change in the S gene is very important (Durmaz et al., 2020). Continuously researching the novel mutation and analyzing the genetic diversity of the gene S plays an essential role in providing information on nucleotide change in this gene. Hence, this study aims to examine the variations in the Spike gene of SARS-CoV-2 isolated in Vietnam. The analysis would support the management of new variant imported into Vietnam.

Methods

Data collection

We collected the sequences of SARS-CoV-2 isolated in Vietnam from the GISAID database. The other information of the sequences consists of accession number, place of sample collection, gender, collection date, and classification of GISAID, PANGO Lineages... Until 17, July 2022, recording 5484 sequences of SARS-CoV-2 of Vietnam in GISAID. However, data had three sequences that were not the whole genome (accession number in GISAID: EPI_ISL_483625, EPI_ISL_2629583, EPI_ISL_1827728). Therefore, these sequences were removed from the analysis data.

Multiple sequences alignment (MSA) analysis

We used the S gene sequence of Wuhan-Hu-1, accession number NC_045512.2 was extracted in GenBank as reference sequences. And then, the SARS-CoV-2 full-length MSA tool available in MAFFT version 7 web server (Katoh and Standley, 2013) was used in order to split the S gene from the whole genome. All of the S gene sequences are available to perform MSA by alignment algorithm, which is also in the MAFFT web server. The FFT-NS-1 strategy alignment, which is an excellent property of MAFFT was used to quickly provide results. Ultimately, the visualization of the S gene sequences was alignment through MEGA software (Tamura *et al.*, 2021) in order to identify the substitutions in DNA and protein sequences.

Examination of haplotype, phylogenetic tree, haplotype network, and genetic diversity

In order to identify of haplotype, we used the Haplotype tool available in DnaSP ver 6.12.03 (Rozas et al., 2017) to record the haplotype of the S gene in Vietnam. These haplotypes will be built in a phylogenetic tree based on the Phylogeny algorithm available in MAFFT server with the most common method is Neighbor-Joining (NJ) - All of gap-free sites in the Jukes-Cantor substitution model and use the reference sequences (NC_045512.2) as the root of the tree. In order to determine the trust level of the branch of the phylogenetic tree, we use the bootstrap value with 1000 repetitions. In the final step, we used the Figtree ver v1.4.4 software (http://tree.bio.ed.ac.uk/software/figtree/) in order to visualize and design the tree.

The detail of the genetic relationship between all haplotypes appeared in Vietnam based on the combination of two methods phylogenetic tree and haplotype network. The phylogenetic tree might determine different clusters appearing in the tree, while the haplotype network could exhibit detail of the nucleotide changes of each cluster in that tree. In this study, the haplotype networks were built by POPART software (Leigh and Bryant, 2015) based on the minimum spanning network method. Ultimately, Arlequin software ver 3.5.2.2 (Excoffier and Lischer, 2010) was used in order to determine genetic diversity indexes.

Results

Until 17th July 2022, we extracted 5481 of the S gene sequences from the whole genome isolated in

Vietnam. In addition, we recorded three samples (GISAID accession number: EPI_ISL_11776067, EPI_ISL_11776066, EPI_ISL_11776065) that appear the single insertion nucleotide in the S gene. As a result, a single insertion nucleotide makes a frameshift in the open reading frame of the S gene.



Fig. 1. Frequency of major variants in Vietnam.

Therefore, the sequences have a single insertion that will be rejected in the data analysis. Then, an examination of the S gene data indicated 1093 haplotypes with numerous polymorphic sites, a high level of genetic diversity, and nucleotide diversity. Detailed haplotype analysis shows that almost the variants that appeared in Vietnam belonged to Delta and Omicron variants that account for a relative proportion, respectively, 50.91% and 45.89%.

The rest of the other haplotypes only account for a low percentage (Fig. 1). Calculating the haplotype diversity by Arlequin software presented the Hd index at 0.9089 ± 0.0031 and the mean number of pairwise differences at 27.491490 ± 12.039763 . Furthermore, the analysis changes of nucleotides in the S gene confirmed the nucleotide diversity index at 0.007176 ± 0.003475 and recorded 802 polymorphic sites, consisting of 503 sites and 281 sites for nucleotide transitions and transversions, respectively.



Fig. 3. The frequency of the kind of nucleotide substitutions in the S gene.

The frequency of C > T nucleotide substitution is highest at 22.47%, followed by the G > T substitution at 13.33% (Fig. 2). The frequency of deletion mutations at 13.22% only once lower G > T substitution. The presentation of the S gene's mutation position indicated that almost the indels site had just appeared in the region encodes for NTD, and the missense mutations were concentrated in the RBD region (Fig. 3).



Fig. 3. The frequency of the kind of nucleotide substitutions in the S gene.

Then, we used the phylogenetic tree to determine the relationship of the haplotype appearing in Vietnam based on the S gene sequence. The phylogenetic tree of the S gene sequence in Vietnam might demonstrate the genetic relationship of the VOCs variants circulating in this country (Fig. 4). In the tree's topology, we recorded an apparent clustering of the haplotypes of the Alpha, Delta, and Omicron variants on the phylogenetic NJ tree. The evolutionary relationships of Alpha and Delta variants forked from the D614G clade, with 95 and 65 bootstrap values, respectively. Phylogenetic analysis shows that the haplotype group of the Delta variant assembles a separate branch on the tree and is closely related to each other. Furthermore, we found that Omicron is closely related to Delta based on the S gene sequence, with the bootstrap values of the two closest clades between Delta and Omicron being 69 and 95, respectively. In addition, analysis of Omicron clusters in trees revealed intricate genetic relationships.In order to evaluate in detail nucleotide changes in populations, we combined phylogenetic tree and haplotype network methods to determine the relationship between each haplotype in each cluster on the tree (Fig. 5). As a result, we observe mutation in the S gene of the clusters that appear from the first outbreak to the third in Vietnam are limited.

At branches D614 and Alpha, haplotypes differ byonly one nucleotide. The rest are mainly haplotypes that can be imported into Vietnam when they differ by 3 or 5 nucleotides. Especially the Alpha variant, which is mainly recorded by events entering Vietnam.

The rest, in the Delta and Omicron clusters, there are diverse and complex changes in the S gene. In the Delta cluster, record the close relationship between the haplotypes when the haplotype differs only on average 1.989772 ± 1.123948 nucleotides and can be detected from an origin haplotype (Fig. 5). Meanwhile, the Omicron cluster has a huge mean nucleotide difference, respectively $17.951297 \pm 7,970441$, indicating that the haplotypes of the Omicron cluster have different nucleotide similarities.

Discussion

Vietnam has undergone three COVID-19 outbreaks, and the fourth outbreak is and continues to be complicated. Our study provides exhaustive information on the genetic diversity of SARS-CoV-2's S gene variants. We noted a high genetic diversity of the S gene in Vietnam. In 5478 S gene sequences, we identified 1093 haplotypes with 802 polymorphisms. The average nucleotide difference is approximately 27 nucleotides, indicating that this gene has undergone

the accumulation of many different mutations. In evaluating the location distribution of mutations, we found that mainly important mutations, with a high frequency of occurrence, especially proved to be essential for SARS-CoV-2 is commonly found in the region that codes for the S1 subunit (Figure 2), which consists of two major regions, the NTD and the RBD.



Fig. 4. The phylogenetic NJ tree base on the S gene sequences in Vietnam. *Red: D614 cluster; Blue: D614G cluster; Lightblue: Alpha cluster; Yellow: Delta cluster; Pink: Omicron cluster.*

In particular, NTD plays a significant role in the immune response, responding to neutralizing antibodies. Therefore, mutations appearing in the coding region for NTDs are intended to help the virus evade the immune system and reduce the ability of the host to neutralize (Mishra et al., 2022). In this study, nucleotide deletions often appeared in the NTD coding region of the S gene. As reported by McCarthy et al., deletion mutations in the NTD region are considered to evade neutralizing antibodies and reduce vaccine efficacy (McCarthy et al.). Recently, the Omicron variant appeared to insert nine nucleotides (GAGCCAGAA) at position 22204 to help insert three amino acids, EPE, to form a loop specific to the spike protein of Omicron (Campbell et al., 2021). These three amino acid insertions may be responsible for the reduction of the antibodyneutralizing ability by Omicron (Venkatakrishnan, 2021). The RBD region plays a crucial role in SARS-

CoV-2 infection by interacting with the hACE2 receptor on the cell membrane. Based on studies on SARS-CoV and SARS-CoV-2, it is shown that this RBD region has a high mutation rate, and mutations appearing in this region can have a great potential to influence the interaction between the spike protein and the hACE2 receptor. Currently, mutations such as L452R, T478K, E484K/Q, and N501Y appearing in the RBD region of the spike protein have been shown to enhance the infectivity of SARS-CoV-2 and have the potential to reduce association binds to monoclonal antibodies. (Di Giacomo et al., 2021) (Cherian et al., 2021) (Liu et al., 2022). Through the data of four outbreaks in Vietnam, we recorded almost amino acid substitutions with a high frequency appearing in the RBD consisting of G22578A (G339D), C22686T (S375F), G22813T (K417N), T22917G (L452R), C22995A (T478K) and A23063T (N501Y). These are all mutations that previous

studies have shown to enhance the infectivity of SARS-CoV-2. In addition, at the furin cleavage site, we also recorded many mutations appearing with high frequency, namely A23403G (D614G), C23525T (H655Y), T23599G (N679K), and C23604A/G (P681H/R). Among them, mutations D614G and P681H/R have been shown to play an important role in transmitting SARS-CoV-2 through enhanced spike protein cleavage (Liu *et al.*, 2022) [10] (Zhang *et al.*, 2020) or escape the immune system (Lubinski *et al.*,

2022). However, it does not eliminate the entire function of the S2 subunit, which plays an essential role in the viral fusion, host cell entry, and enhanced structural stability of the protein. Because of that, a new variant with beneficial mutations may appear in this region. Therefore, it is essential to track the variables across the S gene. Furthermore, the study of Yang *et al.* suggests that the mutation rate of this gene may be higher as the vaccination rate increases (Yang *et al.*, 2022).



Fig. 5. The haplotype network of the S gene in Vietnam.

In addition, we also noted that genetic diversity in the S gene varied across disease outbreaks. In the early stages of COVID-19 in Vietnam, from January 2020 to April 2021, nucleotide changes in the S gene are minimal. In contrast, the United States and the United Kingdom recorded the rapid evolution of the S gene in October 2020 and an intense outbreak of the Alpha variant. Until the appearance of the Delta variant, similar to the world, Vietnam recorded the dominance of the Delta variant, which forced a terrible epidemic situation. In Vietnam, the Delta variant is closely related and differs only on average 1.989772 \pm 1.123948 nucleotides, indicating that this variant has changed relatively in Vietnam through

transmission in the community. The scenario for the Omicron variant appearing in Vietnam could be very complicated because the nucleotide difference of this variant in Vietnam is enormous, with an average of 17 nucleotides. Therefore, it can be said that Omicron in Vietnam has been imported from many countries worldwide. Especially recently, Vietnam has also recently introduced highly infectious sub-Omicron variants such as BA.4 and BA.5; this could warn of a new danger to the country.

We especially recorded the split of different VOC variants based on the S gene sequence. The phylogenetic NJ tree defined the relationship between

three groups of variants appearing mainly in Vietnam, namely Alpha, Delta, and Omicron. Our results are similar to the study of O'Toole *et al.* suggested that the S gene sequence region could be used to identify different groups of VOC variants (O'Toole *et al.*, 2022). Our analysis again showed the importance of the S gene region in terms of the biology and genetic relationships between different variants. Therefore, we highlight the critical work evaluating nucleotide variations and the rate of evolution in the S gene region of SARS-CoV-2.

Acknowledgements

This study is financially supported by Saigon University - Vietnam (Project no. CSB2021-01).

References

Berrio A, Gartner V, Wray GA. 2020. Positive selection within the genomes of SARS-CoV-2 and other Coronaviruses independent of impact on protein function. Peer J **8**, e10234.

http://dx.doi.org/10.7717/peerj.10234.

Bolze A, Luo S, White S, Cirulli ET, Wyman D, Dei Rossi A, Machado H, Cassens T, Jacobs S, Schiabor Barrett KM, Tanudjaja F, Tsan K, Nguyen J, Ramirez J 3rd M, Sandoval E, Wang X, Wong D, Becker D, Laurent M, Lu JT, Isaksson M, Washington NL, Lee W. 2022. SARS-CoV-2 variant Delta rapidly displaced variant Alpha in the United States and led to higher viral loads. Cell Rep Med **3(3)**, 100564.

http://dx.doi.org/10.1016/j.xcrm.2022.100564.

Campbell F, Archer B, Laurenson-Schafer H, Jinnai Y, Konings F, Batra N, Pavlin B, Vandemaele K, Van Kerkhove MD, Jombart T, Morgan O, le Polain de Waroux O. 2021. Increased transmissibility and global spread of SARS-CoV-2 variants of concern as at June 2021. Euro surveillance : bulletin Europeen sur les maladies transmissibles = European communicable disease bulletin **26(24)**, 2100509.

http://dx.doi.org/10.2807/15607917.ES.2021.26.24.2 100509. Chaguza C, Coppi A, Earnest R, Ferguson D, Kerantzas N, Warner F, Young HP, Breban MI, Billig K, Koch RT, Pham K, Kalinich CC, Ott IM, Fauver JR, Hahn AM, Tikhonova IR, Castaldi C, De Kumar B, Pettker CM, Warren JL, Weinberger DM, Landry ML, Peaper DR, Schulz W, Vogels CBF, Grubaugh ND. 2022. Rapid emergence of SARS-CoV-2 Omicron variant is associated with an infection advantage over Delta in vaccinated persons. Med (New York, N.Y.). http://dx.doi.org/10.1016/j.medj.2022.03.010.

Chau NVV, Hong NTT, Ngoc NM, Anh NT, Trieu HT, Nhu LNT, Yen LM, Minh NN, Phong QNT, Truong NT, Huong LTT, Tu TNH, Hung LM, Thanh TT, Dung NT, Dung NT, Thwaites G, Van Tan L, For OCRG. 2021. Rapid wholegenome sequencing to inform COVID-19 outbreak response in Vietnam. The Journal of infection **82(6)**, 276-316.

http://dx.doi.org/10.1016/j.jinf.2021.03.017.

Cherian S, Potdar V, Jadhav S, Yadav P. 2021. SARS-CoV-2 Spike Mutations, L452R, T478K, E484Q and P681R, in the Second Wave of COVID-19 in Maharashtra, India. 9(7).

http://dx.doi.org/10.3390/microorganisms9071542.

Di Giacomo S, Mercatelli D, Rakhimov A, Giorgi FM. 2021. Preliminary report on severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) Spike mutation T478K. Journal of Medical Virology **93(9)**, 5638-5643.

http://dx.doi.org/10.1002/jmv.27062.

Durmaz B, Abdulmajed O, Durmaz R. 2020. Mutations Observed in the SARS-CoV-2 Spike Glycoprotein and Their Effects in the Interaction of Virus with ACE-2 Receptor. Medeniyet Medical Journa **35(3)**, 253-260.

http://dx.doi.org/10.5222/MMJ.2020.98048.

Excoffier L, Lischer HE. 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol Ecol Resour **10(3)**, 564-567. http://dx.doi.org/10.1111/j.1755-0998.2010.02847.x.

Fan LQ, Hu XY, Chen YY, Peng XL, Fu YH, Zheng YP, Yu JM, He JS. 2021. Biological Significance of the Genomic Variation and Structural Dynamics of SARS-CoV-2 B.1.617. Frontiers in microbiology **12**, 750725-750725.

http://dx.doi.org/10.3389/fmicb.2021.750725.

Hoang VT, Pham TD, Nguyen QT, Nguyen DC, Nguyen DT, Nguyen TB, Tran TKT, Phan TL, Vo PLN, Dao TL, Fenollar F, Gautret P. 2022. Seroprevalence of SARS-CoV-2 among high-density communities and hyper-endemicity of COVID-19 in Vietnam. Trop Med Int Health.

http://dx.doi.org/10.1111/tmi.13744.

Katoh K, Standley DM. 2013. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. Molecular Biology and Evolution **30(4)**, 772-780. http://dx.doi.org/10.1093/molbev/mst010.

Kumar S, Thambiraja TS, Karuppanan K, Subramaniam G. 2022. Omicron and Delta variant of SARS-CoV-2: A comparative computational study of spike protein. Journal of Medical Virology **94(4)**, 1641-1649.

http://dx.doi.org/10.1002/jmv.27526.

Kumar V, Singh J, Hasnain SE, Sundar D. 2021. Possible Link between Higher Transmissibility of Alpha, Kappa and Delta Variants of SARS-CoV-2 and Increased Structural Stability of Its Spike Protein and hACE2 Affinity. International journal of molecular sciences **22(17)**, 9131.

http://dx.doi.org/10.3390/ijms22179131.

Leigh JW, Bryant D. 2015. popart: full-feature software for haplotype network construction. Methods in Ecology and Evolution **6(9)**, 1110-1116. https://doi.org/10.1111/2041-210X.12410.

Liu Y, Liu J, Johnson BA, Xia H, Ku Z,

Schindewolf C, Widen SG, An Z, Weaver SC, Menachery VD, Xie X, Shi PY. 2022. Delta spike P681R mutation enhances SARS-CoV-2 fitness over Alpha variant. Cell Reports **39(7)**, 110829. https://doi.org/10.1016/j.celrep.2022.110829. Liu Y, Liu J, Plante KS, Plante JA, Xie X, Zhang X, Ku Z, An Z, Scharton D, Schindewolf C, Widen SG, Menachery VD, Shi PY, Weaver SC. 2022. The N501Y spike substitution enhances SARS-CoV-2 infection and transmission. Nature **602(7896)**, 294-299.

http://dx.doi.org/10.1038/s41586-021-04245-0.

Lubinski B, Fernandes MHV, Frazier L, Tang T, Daniel S, Diel DG, Jaimes JA, Whittaker GR. 2022. Functional evaluation of the P681H mutation on the proteolytic activation of the SARS-CoV-2 variant B.1.1.7 (Alpha) spike. iScience **25(1)**, 103589.

https://doi.org/10.1016/j.isci.2021.103589.

Lyngse FP, Molbak K, Skov RL, Christiansen LE, Mortensen LH, Albertsen M, Moller CH, Krause TG, Rasmussen M, Michaelsen TY, Voldstedlund M, Fonager J, Steenhard NC. Danish Covid-19 Genome, Kirkeby CT. 2021. Increased transmissibility of SARS-CoV-2 lineage B.1.1.7 by age and viral load. Nat Commun **12(1)**, 7251.

http://dx.doi.org/10.1038/s41467-021-27202-x.

McCarthy KAO, Rennick LAO, Nambulli SAO, Robinson-McCarthy LAO, Bain WAO, Haidar G, Duprex WAO. Recurrent deletions in the SARS-CoV-2 spike glycoprotein drive antibody escape. (1095-9203 (Electronic)).

Minh LHN, Khoi Quan N, Le TN, Khanh PNQ, Huy NT. 2021. COVID-19 Timeline of Vietnam: Important Milestones Through Four Waves of the Pandemic and Lesson Learned. Front Public Health 9, 709067.

http://dx.doi.org/10.3389/fpubh.2021.709067.

Mishra T, Dalavi R, Joshi G, Kumar A, Pandey P, Shukla S, Mishra RK, Chande A. 2022. SARS-CoV-2 spike E156G/Delta157-158 mutations contribute to increased infectivity and immune escape. Life Sci Alliance 5(7).

http://dx.doi.org/10.26508/lsa.202201415.

O'Toole A, Pybus OG, Abram ME, Kelly EJ. Rambaut A. 2022. Pango lineage designation and assignment using SARS-CoV-2 spike gene nucleotide sequences. BMC Genomics **23(1)**, 121.

http://dx.doi.org/10.1186/s12864-022-08358-2.

Ovsyannikova IG, Haralambieva IH, Crooke S. N, Poland GA, Kennedy RB. 2020. The role of host genetics in the immune response to SARS-CoV-2 and COVID-19 susceptibility and severity. Immunol Rev **296(1)**, 205-219.

http://dx.doi.org/10.1111/imr.12897.

Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, Sánchez-Gracia A. 2017. DnaSP 6: DNA Sequence Polymorphism Analysis of Large Data Sets. Molecular Biology and Evolution **34(12)**, 3299-3302. http://dx.doi.org/10.1093/molbev/msx248.

Tamura K, Stecher G, Kumar S. 2021. MEGA11: Molecular Evolutionary Genetics Analysis Version 11. Mol Biol Evol **38(7)**, 3022-3027. http://dx.doi.org/10.1093/molbev/msab120.

Venkatakrishnan A, Praveen Anand, Patrick Lenehan, Rohit Suratekar, Bharathwaj Raghunathan, Michiel Niesen J, Venky Soundararajan. 2021. Omicron Variant of Sarscov-2 Harbors a Unique Insertion Mutation of Putative Viral or Human Genomic Origin. OSF Preprints December 3.

http://dx.doi.org/10.31219/osf.io/f7txy.

Walker AS, Vihta KD, Gethings O, Pritchard E, Jones J, House T, Bell I, Bell JI, Newton JN, Farrar J, Diamond I, Studley R, Rourke E, Hay J, Hopkins S, Crook D, Peto T, Matthews DW, Eyre N, Stoesser, Pouwels KB. 2021. Tracking the Emergence of SARS-CoV-2 Alpha Variant in the United Kingdom. New England Journal of Medicine **385(27)**, 2582-2585.

http://dx.doi.org/10.1056/NEJMc2103227.

Wei C, Shan KJ, Wang W, Zhang S, Huan Q, Qian W. 2021. Evidence for a mouse origin of the SARS-CoV-2 Omicron variant. J Genet Genomics **48(12)**, 1111-1121.

http://dx.doi.org/10.1016/j.jgg.2021.12.003.

Yang J, Han M, Wang L, Wang L, Xu T, Wu L, Ma J, Wong G, Liu W, Gao GF, Bi Y. 2022. Relatively rapid evolution rates of SARS-CoV-2 spike gene at the primary stage of massive vaccination. Biosaf Health 4(4), 228-233.

http://dx.doi.org/10.1016/j.bsheal.2022.07.001.

Yang W, Shaman J. 2022. COVID-19 pandemic dynamics in India, the SARS-CoV-2 Delta variant and implications for vaccination. J R Soc Interface **19(191)**, 20210900.

http://dx.doi.org/10.1098/rsif.2021.0900.

Zhang L, Jackson CB, Mou H, Ojha A, Peng H, Quinlan BD, Rangarajan ES, Pan A, Vanderheiden A, Suthar MS, Li W, Izard T, Rader C, Farzan M, Choe H. 2020. SARS-CoV-2 spike-protein D614G mutation increases virion spike density and infectivity. Nature Communications 11(1), 6013.

http://dx.doi.org/10.1038/s41467-020-19808-4.

Zhou H, Yang J, Zhou C, Chen B, Fang H, Chen S, Zhang X, Wang L, Zhang L. 2021. A Review of SARS-CoV2: Compared With SARS-CoV and MERS-CoV. Front Med (Lausanne) **8**, 628370. http://dx.doi.org/10.3389/fmed.2021.628370