ATLANTIS PRESS
Advances in Biological Sciences Research, volume 22
7th International Conference on Biological Science (ICBS 2021)

A Year of COVID-19 Outbreak in Indonesia #2: Variant Development Based on *Spike* (*S*) Mutations

Nicholas Gerry Andreanto[1] Dwita Novitasari[1] Delia Wahyu Pangesti[1] Rizqi Layli Khusufi[1] Annasa Sabatia[1] Irwansyah Nur Oktafian[1] Nur Alfi Maghfirotus Sa'adah[1] Reni Krisdayana[1] Salsabila Kasta Hygiea Iswara[2] Dwi Listyorini[1, 2,*]

[1]*Department of Biology, Faculty of Mathematics and Natural Sciences, Universitas Negeri Malang Jl. Semarang No.5, Malang, 65145, Indonesia*
[2]*Department of Biotechnology, Faculty of Mathematics and Natural Sciences, Universitas Negeri Malang Jl. Semarang No.5, Malang, 65145, Indonesia*
[*]*Corresponding author. Email: listyorini.aljabari@um.ac.id*

ABSTRACT
SARS-CoV-2 has infected millions of people in Indonesia and taken thousands of lives by bonding Spike (S) protein and Angiotensin-Converting Enzyme 2 (ACE2) human cell receptor. *Spike* gene has a higher mutation rate compared to other genes, which suggested to increase its virulence, transmission and change the virus regulation inside the cell; hence genetic mutation surveillance is needed. This study aimed to determine the dynamics of SARS-CoV-2 *spike* gene mutation to predict the development of Covid-19 in the future. This study was conducted based on Big-data. *Spike* gene sequence samples were retrieved from GISAID EpiCoV^{TM} website database and NCBI from March 2020 to March 2021. Multiple alignment of the sequences was achieved using the ClustalW algorithm from BioEdit 7.2.5 version. Mutation and variant analysis, and phylogenetic tree reconstruction, were performed using MEGA X. A total of 146 mutation sites were discovered within Indonesian samples and 100 from 19 comparison countries (Overseas). As many as 135 variants were exclusively found from Indonesian samples and 77 variants from overseas, and 5 from both. One distinct Indonesian variant is thought to have originated from abroad and underwent further mutations in Indonesia. Based on our results, it can be concluded that the SARS-CoV-2 virus is suspected of continuing to mutate if it is still spreading in the community.

Keywords: *COVID-19, COVID-19 in Indonesia, SARS-CoV-2, Spike gene mutation*

1. INTRODUCTION

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is a new coronavirus variant of the beta coronavirus genus. Coronavirus has a spherical structure with a bulge (Spike protein) on its surface that is shaped like a crown or sun, hence called a coronavirus (corona Latin for "crown"). SARS-CoV-2 is a single-stranded RNA virus with a 60-140 nm [1], [2]. This virus first appeared in Wuhan city, China, which caused the COVID-19 outbreak and then spread rapidly worldwide. The first case was reported in Indonesia by March 2, 2021 [3]. It reached 1.4 million patients and caused deaths of over 3.8 thousand people until March 12, 2021 [4], which shows that the virus spreads rapidly in just a year.

The 26-32 kb SARS-CoV-2 genome consists of 10 open reading frames (ORF) [5], [6] and four main structural proteins, particularly Spike (S), Envelope (E), Membrane (M), and Nucleocapsid (N) [7] which its composition is similar to that of SARS-CoV (79%) and coronaviruses in bats (96.2%) [8]. The difference between SARS-CoV-2 with other coronaviruses is its longer *Spike* gene [9]. Spike protein is the biggest of four other structural proteins, with each monomer

Copyright © 2022 The Authors. Published by Atlantis Press International B.V.
This is an open access article distributed under the CC BY-NC 4.0 license -http://creativecommons.org/licenses/by-nc/4.0/.

measuring around 180 kDa [10]. This protein is responsible for the entry of viruses into host cells [10]–[13]. The membrane protein is one of the functional proteins that play an essential role in maintaining the size and shape of the virion [14] and is the most abundant [7]. In contrast, Envelope protein is relatively small in size, is involved in virion assembly and released [7], [10]. Nucleocapsid protein plays an essential role in packaging viral RNA into ribonucleocapsid. It mediates viral assembly by interacting with the viral genome and Membrane protein which helps in the augmentation of viral RNA transcription and replication [7].

SARS-CoV-2 is thought to originate from the zoonotic transfer of the pangolin betacoronavirus [13], [15]–[18]. It spreads to humans due to the insertion of 12 nucleotides in the binding domain of the Spike protein, which causes optimal binding to the human Angiotensin-Converting Enzyme 2 (ACE2) receptor on human cells [10]–[13], [19], [20], [16]. To date, there is no epidemiological evidence of direct or indirect transmission of SARS-CoV-2 from bats to humans, but based on complete genome analysis, 96.2% of its genome are related to the bat *Rhinolophus affinis* [16], [21]. The Spike protein is the most important SARS-CoV-2 protein for its role in the coronavirus entry into human cells mediated by S1 and S2 subunits to bind the ACE2 cell receptor in humans [11], [22]. Furthermore, the expression *Spike* gene plays a vital role in inhibiting ACE2 expression and promoting the release of IL-6/IL-6R cytokine molecules that cause an imbalance of the renin-angiotensin system, causing an increase in the concentration of pro-inflammatory cytokine molecules in human epithelial cells [23].

The mutation rate in RNA viruses is higher than that of DNA viruses counterparts [24], [25], which may cause viruses to evolve rapidly and, in turn, may increase their virulence [26] and transmissibility [27]. In addition to their rapid spread, RNA viruses such as SARS-CoV-2 are well known for their high mutation rates [24]. It is reported that the mutation rate of coronavirus is ~10^{-6} in every replication [28]. However, each component of the SARS-CoV-2 proteome mutates at a different pace. Some proteins, such as Envelope protein, have a low mutation rate, while Spike proteins and Nucleocapsid show a higher degree of its variability, which means that they mutate relatively fast [29]. Changes or mutations in the Spike protein may affect the rate of transmission and infection or even change the homeostasis regulation in human cells, which still requires further study.

The widely found D614G mutation is associated with more efficient replication and transmission [30] and increased viral infectivity due to decreased shedding of the S1 subunit and increased density of Spike protein in the virion but did not increase Spike protein affinity for ACE2 or make pseudovirus more resistant to neutralization [5], [6], [31]–[33]. Since Spike mutations are fast and could potentially increase its virulence, transmission, and change viral regulation in cells, it is necessary to monitor its development, especially for which spread in Indonesia, in anticipation of the emergence of new local mutants. This data is important to mitigate the possible further outbreak.

2. METHOD

2.1. Sampling

This research is based on big data. The sample consisted of the *Spike* gene retrieved from the whole-genome of SARS-CoV-2 reported from Indonesia, Overseas, and Wuhan. The data was downloaded from the EpiCoV^{TM} GISAID database website [34], [35], and NCBI from available from March 2020 to March 12, 2021. Genomic sequence samples originating from Indonesia were obtained using the "Asia/Indonesia/" location criteria. The number of samples of Indonesian sequences was 548. For comparison, each of the 10 newest samples from 17 countries with the most COVID-19 cases according to WHO (United States, India, Brazil, France, Russia, Turkey, England, Italy, Spain, Germany, Argentina, Poland, Colombia, Iran, Mexico, Ukraine, and Peru) plus 2 countries: South Africa and Australia to complete the entire continent coverage were analyzed. The total sample sequences for 19 comparison countries (Overseas) were 190. The Wuhan sample with sequence ID: NC_045512.2 downloaded from the NCBI database was used as a wild-type reference.

2.2. Sequence processing & multiple alignments

Multiple whole-genome sequence alignment was performed using the ClustalW algorithm through the BioEdit Sequence Alignment Editor software version 7.2.5 [36] to obtain the *Spike* gene sequence. The sequences that have been aligned were cut off from the *Spike* gene section (bases 21563-25384) [37]. Among the 548 genomic samples, there were 12 samples excluded due to the lack of *Spike* gene sequence. The total samples were 536 samples from Indonesia, 190 overseas samples, and one sample from Wuhan. The total sequence was 727. After the sequence of *Spike* gene was obtained, the sequence was translated into a Spike protein sequence, and further analysis for Spike protein mutation was carried out as bellow.

2.3. Mutation analysis and mutation identification

Amino acid mutation analysis was carried out to identify mutations in each sample using the MEGA X

software: Molecular Evolutionary Genetics Analysis across computing platforms [38]. Mutation analysis based on comparison with sequence samples from Wuhan as a wild-type reference. The variable sites of each S protein sequence were highlighted to mark the mutated sites. The highlighted variable areas are then exported and tabulated using Microsoft Excel software.

2.4. Variant identification

Identification of variants was made manually using Microsoft Excel, based on amino acid mutations from the amine-end to the carboxyl end of the Spike polypeptide. Variant naming follows the rule provided by Pangolineage [39], [40].

2.5. Phylogenetic tree reconstruction

Reconstruction of the phylogenetic tree was carried out for group identification and prediction of the origin of the SARS-CoV-2 variants. This reconstruction was carried out based on the amino acid sequence of Spike protein using the MEGA X software: Molecular Evolutionary Genetics Analysis across computing platforms [38]. Protein sequences are used for phylogenetic tree reconstruction because there are more possible characters for amino acids (20) than for nucleotides (4) [41]. This phylogenetic tree reconstruction used the Neighbor-Joining (NJ) method with 1000 bootstrap times. The iTOL: Interactive Tree of Life was used to visualize the tree [42], [43].

3. RESULT

3.1. Types of Mutation and Variant

Mutation analysis of 536 samples of Spike protein of the Indonesian sequences has found 484 samples with mutations and 52 samples without mutations. From 190 samples of comparison countries from 19 countries, it successfully revealed that all samples had mutations. There are 146 mutation sites from all Indonesian samples (Table 1) and 100 mutations from 190 samples of comparison countries (Table 2).

The mutation patterns found in each Spike protein domain were similar between Indonesian and overseas samples. Mutation sites in the N-terminal domain (NTD) in Indonesia were higher (36%) than those from overseas (32%). In other domains, the number of mutation sites is relatively the same (Figure 1).

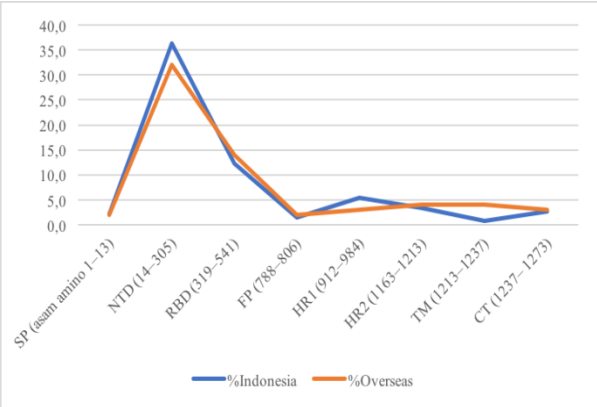

Figure 1. Comparison of the frequency of mutations that occur in each domain of the Spike protein in Indonesia and overseas. SP: signaling peptide; NTD : subunit 1 with N-terminal domain, RBD: receptor binding domain; FP: subunit 2 with fusion peptide, HR1: heptapeptide repeat 1 sequence, HR2: heptapeptide repeat 2 sequence; TM: transmembrane; and CT: cytoplasm.

Identification of variants from all samples revealed as much as 217 variants. A total of 135 variants only exist in Indonesia, exclusively, 77 variants only exist overseas, and five are found both in Indonesia and overseas samples. Most variants from Indonesia only appear in a few samples, for example, variant V31+P561S+D614G (one sample); variant S12F+D614G+P812S (two samples); and the D80Y+D614G+A1078S variant (three samples). Specific variants appeared in many samples, such as the L5F+D614G variant found in 14 samples from Banten, East Java, West Kalimantan, West Java, Yogyakarta, and Jakarta. The S12F+D614G variants were found in 9 samples from Banten, West Java, and Jakarta. The L18F+D614G+S689R variants were found in 4 samples from West Java. Variant V231A+D614G was found in

Table 1. List of mutation sites from the Indonesian samples

L5F	S13I	L18F	T20N/I	Q23R	P26S	A27S	L54F	H69S	G72W	S13I
R78S	D80Y/A	S94F	T95I	S98F	D138Y	W152L	M153T	F157S	N164S	D80Y/A
L176F	M177I	D178N	E180K	F186S	R190S	I210T	V213L	D215G	L216F	M177I
A222V	R246I	L249S	S254F	T307I	Q314R	A348S	V382L	Q414R	K417T/N	R246I
N439K	N440K	L452R	S477N	T478K	E484K	N501Y	L517F	A520S	A522S	N440K
A570D	D574N	Q613H	D614G	T618I	S640F	A653V	E654K	H655Y	A668V	D574N
Q675H	Q677H	N679K	P681H/R	S689I	A701V	T716I	T732A	G769V	T778I	Q677H
E780Q	T791I	F797L	A845S	A846V	T859N/I	A879S	A892S	S939F	D950N	T791I
S982A	T1027I	P1069S	A1078S	V1094I	H1101Q	T1117I	D1118H	Y1155F	G1167A	T1027I
V1176F	N1178D	K1191N	I1216V	G1219C	V1228L	M1237I	C1247F	V1264L	L1265F	N1178D

Table 2. List of mutation sites from overseas samples

V3I	L5F	S12F	Q14K	V16A	L18F	T22P/I	L24V	P26L/S	T29I
H49Y	L54F	N61Y	A67S/V	V70F	T76I	D80Y	V83L	V90F	T95I
E96D	R102I	S116C	V127I	D138Y	F140L	Y144F	Y145D	H146R	K147I
N148T	M153I	E154K	E156D	F157S/L/C	R158S	Y170H	N185Y	F186V	E191K
F192V	V193M	I197V	V213A/L/E	R214H/L	L216F	Q218K/R	A222V	L242F	S254F
W258L/R	T259I	A260V	A262S	P272S	T286I	E309Q	F347L	A348S/Q	A352S
V367F	S371P	L390I	N394K	A397T	F400L	G431V	N439K	L441I	S477I
S494P	N501T/Y	E516Q	A522P	T547I	P561S	A570D/S	T572I	Q613H	D614G
V622F	A623V	A626V	D627H	P631S	S640F	A647S	E654Q	H655Y	A672V
Q675R/K/H	Q677H	N679K	P681H/R	A688V	S689R	Q690H	S691Y	T716I	T719S/I
M731I	T732S	G744S	A771S	Q804P	L806Q	D808E	K811I	P812Q/L/S	S813N
L822F	G832D	I834V	G838D	A845V	N856D	T859I	I870V	A890T	D936H/N
D950Y	V951L	Q954K	A958T	N960K	L962H	S982A	K1073N	A1078S	T1105P
T1117I	D1118H	V1122L	G1124V	I1132V	P1162L	G1167V	I1169L	V1176F	V1177L
Q1208H	M1229I	C1254F	D1259Y	D1260N	P1263L				

13 samples from South Sulawesi, Central Java, Banten, and Yogyakarta. The N439K+D614G variant was found in 23 samples from Jakarta, East Java, Banten, and West Java. The N439K+D614G+P681R variant was found in 16 samples from East Java, Papua, West Java, and Jakarta. The Q613H+D614G variant was found in 6 samples from West Java, Banten, and Jakarta. Variant D614G+Q677H was found in 18 samples from West Java, Jakarta, East Java, Banten, and West Kalimantan. Variant D614G+S689R was found in 7 samples from Banten, East Java, Jakarta, Yogyakarta, and West Java (Figure 2).

One variant of interest and four concerns were found from the Overseas sample. Variant S13I+W152L+L452R+D614G, a B.1.429 VOC, was found in sample from the United States. Variant L18F+T20N+P26S+D138Y+R190S+K417T+E484K+N501Y+D614G+H655Y+T1027I+V1176F, a P.1 VOC, was found in 11 samples from several South American countries, namely Brazil, Colombia, and Peru. Variant samples from European countries: France, Poland, UK, Italy and Germany, Russia, Iran, and Australia. Variant D80G+F157S+L452R+D614G+T859N+D950N, VOI B.1.526.1 was found in one sample from the United States. Variant D80A+D215G+K417N+E484K+N501Y

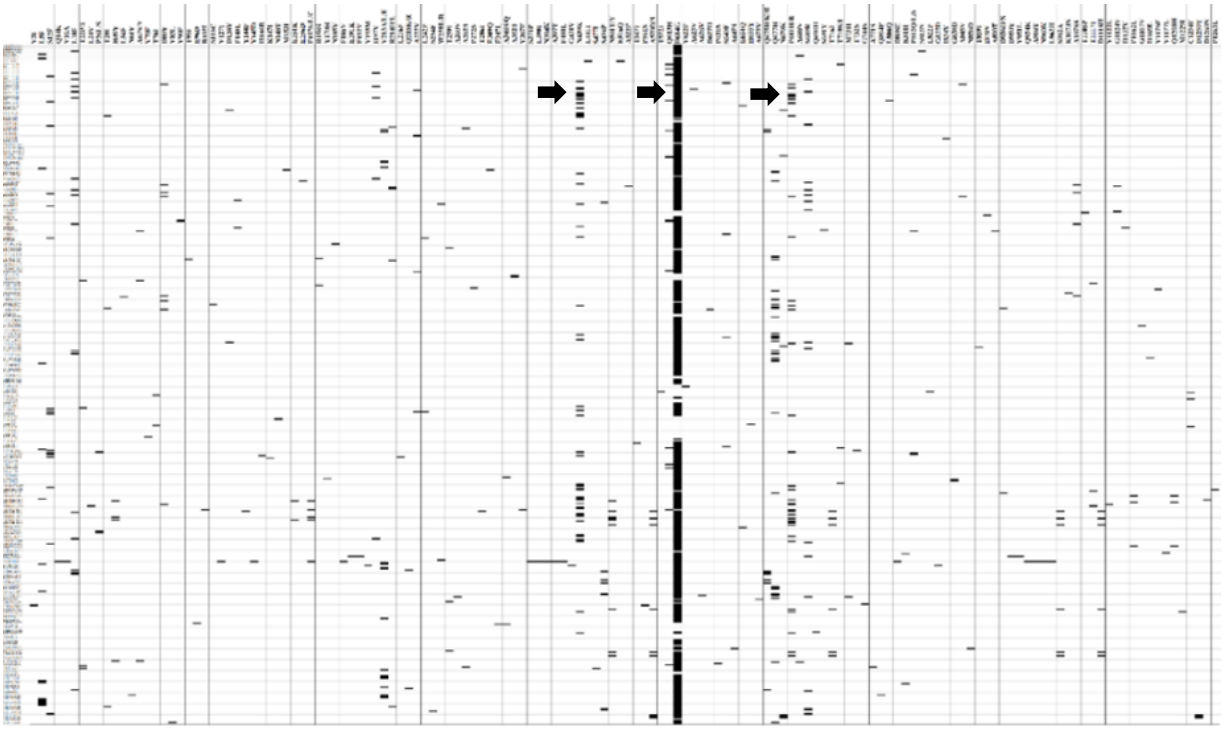

Figure 2. Distribution of mutations from the Indonesian sample. Mutation pattern of variants N439K+D614G+P681R (arrows)

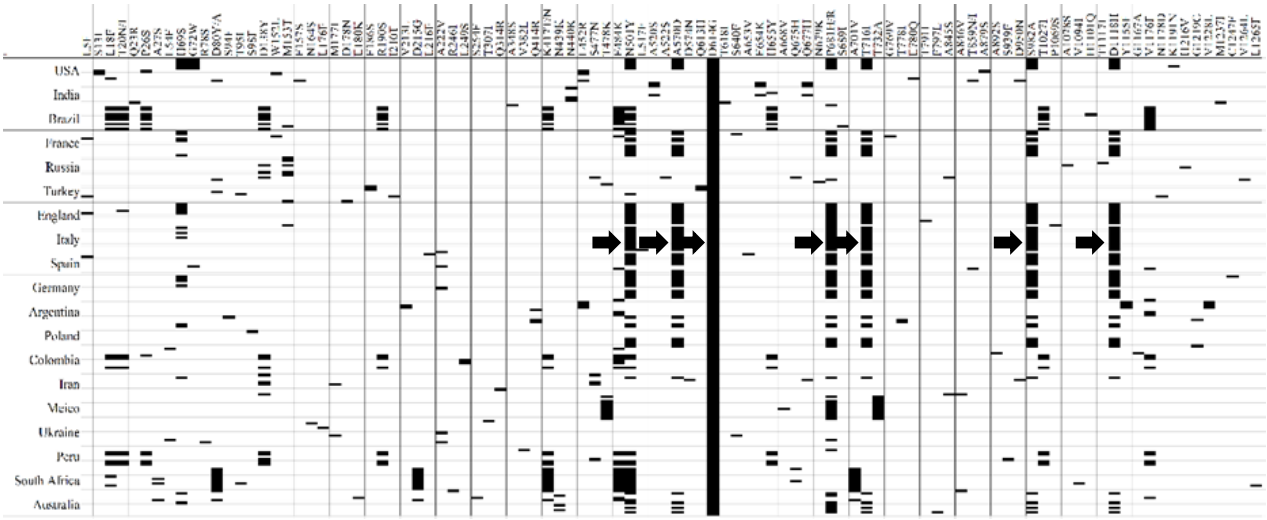

Figure 3. The distribution of mutations from 17 countries with the most COVID-19 cases according to WHO as of 04/24/2021 plus 2 countries, South Africa, and Australia. There is a mutation pattern of the N501Y+A570D+D614G+P681H+T716I+S982A+D1118H (arrows) variant in the United States, France, England, Italy, Spain, Germany, Argentina, Poland, Iran, and Australia.

+D614G+A701V, which is VOC B.1.351, was found in 3 samples from South Africa, and variant L452R+ D614G+Y1155F+V1228L, VOC B.1.427, was found in 1 sample from Argentina (Figure 3).

Variant A222V+D614G was found in 6 samples from Banten, West Java, Germany, Spain, and Ukraine. Variant N501Y+A570D+D614G+P618H+T716I+ S982A+D1118H, VOC B.1.1.7, was found in 32 samples from North and South Sumatra, South Kalimantan, Jakarta, Australia, England, Poland, France, Germany, Spain, and Italy. The D614G variant was found in 217 samples from Bengkulu, Kalimantan, Maluku, West Nusa Tenggara, Aceh, North and South Sumatra, Lampung, Banten, Jakarta, West Java, Central Java, Yogyakarta, East Java, Bali, North and South Sulawesi, Australia, France, India, Russia, Turkey, United States, Colombia, Iran, Poland, Ukraine, and Peru. Variant D614G+N679K was found in 3 samples from Banten, East Java, and Turkey. Variant D614G+T1117I was found in 2 samples from East Java and Russia (Figure 4).

3.2. The occurence of mutations in Indonesia and abroad

The incidence of mutations in the Indonesian sample is less than that in the Overseas sample. Of the 536 Indonesian samples, 462 (86.2%) had D614G mutations, the second most common mutation was N439K (9%), followed by P681H/R (6.5%), and Q677H (5.8%). The most mutations from Overseas samples were D614G with 99.5% of the total sample followed by N501Y

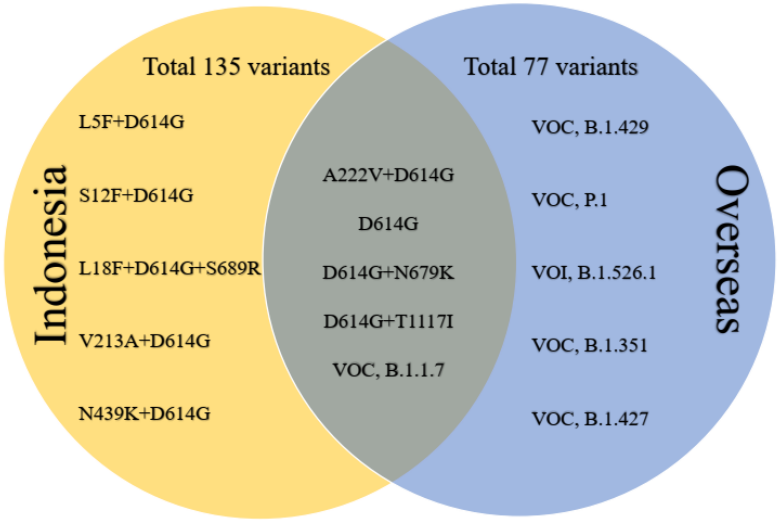

Figure 4. Venn diagram of Indonesia and Overseas Variants. A total of 5 variants are found in both Indonesian and Overseas variant part.

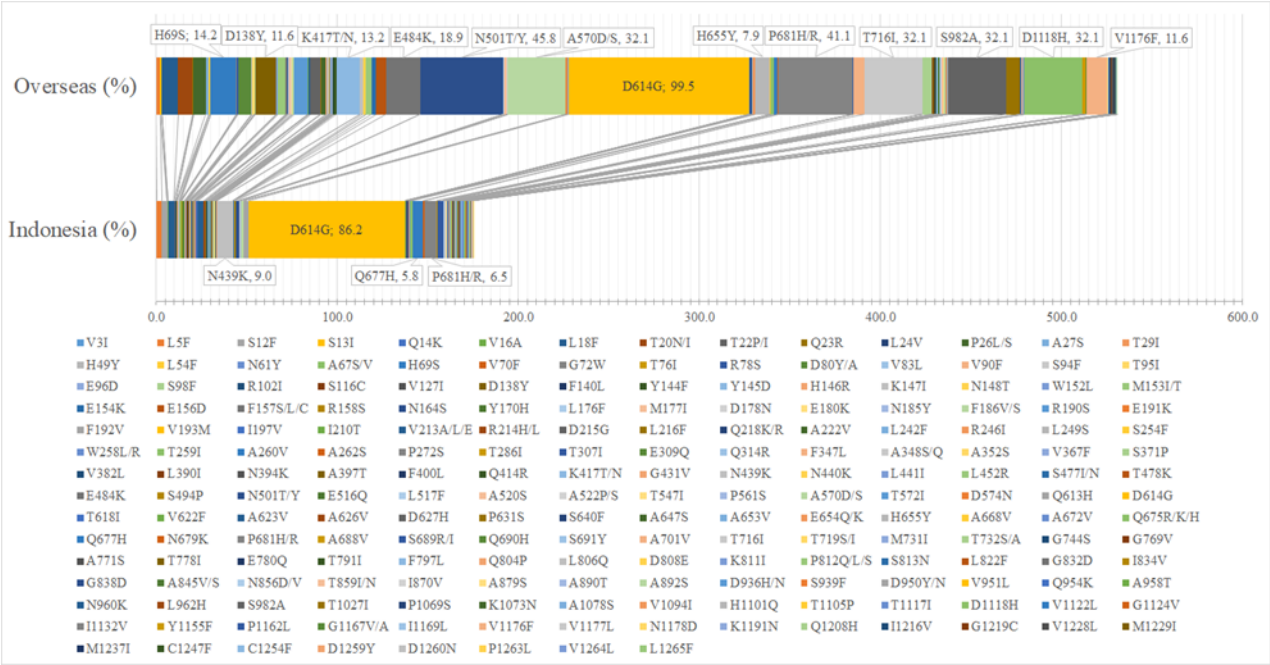

Figure 5 Frequency of mutation sites (%) from Indonesian and Overseas samples.

(45.8%), P681H/R (41.1%), and A570D (32.1%). In the comparison of 536 samples from Indonesia and 190 samples from Overseas, shows that D614G is the dominant mutation both in Indonesia and Overseas. Another dominant mutation is different between Indonesia and Overseas. This also shows that the mutation pattern in Indonesia is not the same as Overseas (Figure 5).

3.3. Results of phylogenetic tree analysis

Based on the results of the reconstruction of the phylogenetic tree, it is known that within one year of the epidemic in Indonesia, several Indonesian SARS-CoV-2 variants were in a different clade from Overseas and the others were in the same clade (Figure 6).

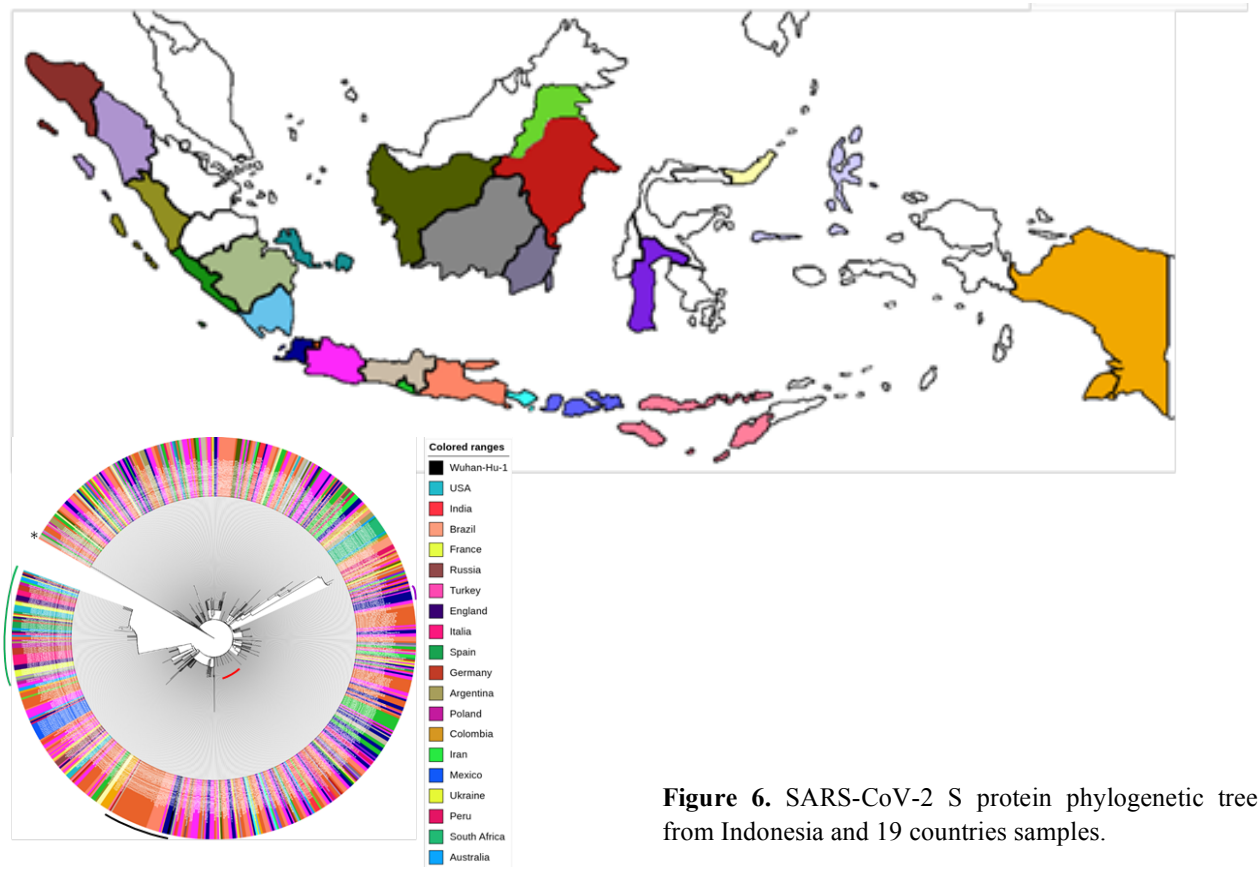

Figure 6. SARS-CoV-2 S protein phylogenetic tree from Indonesia and 19 countries samples.

A total of 52 Indonesian samples are in the same clade as samples from Wuhan (Figure 6. Black curve). This shows that the 52 samples are wild-type. Based on mutation analysis, no mutation was found in the 52 samples. From the Wuhan clade (SARS-CoV wild-type), new clades were formed. Based on mutation analysis and identification of variants, the clades are composed of variants with 1-17 mutation points.

The variant of concern B.1.1.7 forms a separate clade. Some samples from North and South Sumatra, South Kalimantan, and Jakarta are in clade B.1.1.7 and SARS-CoV-2 Europe and Australia (Figure 6. Green curve). One sample of Indonesian VOC B.1.1.7 is separate from the other 5. This indicates a further mutation of variant B.1.1.7. In addition, there is the D614G variant clade which consists of many Indonesian samples and all countries from Overseas (Figure 6. Red curve), and the N439K+D614G variant clade with 20 samples from Indonesia without any Overseas samples (Figure 6. Purple curve). One sample from Central Java with a total of 17 mutations formed a separate clade that was far apart from the other clades (Figure 6, an asterisk (*)).

4. DISCUSSION

Within one year of the COVID-19 outbreak, more mutation cases were found in Indonesia than Overseas. From the aspect of number, samples from Overseas were taken from the last 10 events, so, likely, samples carrying the mutation were not taken and identified in this study. Another possibility is that this mutation occurred in the SARS-CoV-2 virus spread in Indonesia. This is because RNA viruses are easier to mutate than DNA viruses [24], [25], [28]. In addition, the Spike gene has a higher mutation rate than the other SARS-CoV-2 genes [29].

The type of mutation in the Indonesian SARS-CoV-2 Spike protein is not the same as the Overseas SARS-CoV-2. Some mutations are found in Indonesia but not Overseas and vice versa. Mutations were also found at the same site but with different mutated amino acid constitutions. It is suspected that this mutation is a substitution that impacts amino acid changes [44]–[46]. This condition is reinforced by data on mutation patterns per domain in Spike. In several domains, Indonesian SARS-CoV-2 has more mutations per domain than Overseas SARS-CoV-2, which strengthens the notion that mutations occurred after the virus entered Indonesia. A variant has one or more mutations that distinguish it from other circulating variants [47]. Like the mutation sites, there are more variants in Indonesia rather than in overseas samples. Meanwhile, as much as 52 samples of wild-type SARS-CoV-2 were found in Indonesian samples indicating that the patients were in contact with those who had been in or contacted with people from the city of Wuhan, China, where the SARS-CoV-2 virus was first detected or emerged [48]. This is evidenced by the results of the reconstruction of the phylogenetic tree based on the Spike protein. Samples with no mutation (wild type) reported from Indonesia occupy one clade with samples from Wuhan.

As well as the situation overseas, the D614G variant was dominant in Indonesia. A case of the D614G mutation was initially detected in China on January 24, 2020 [48]. A non-synonymous mutation that causes aspartate at codon 614 to become glycine gives rise to this variant [27] which has more infectious and transmission efficiency and can spread more rapidly [5], [6], [27], [30]–[33], [49]. In developing the SARS-CoV-2 genome, the D614G mutation was also accompanied by other mutations that formed various variants. Identification of local variant unveiled that mutation in spreading SARS-CoV-2 in Indonesia running in high speed resulted in many distinct variants from those recorded in WHO. Other mutations were the same found in other countries, one of them is the L5F + D614G variant. The variant with a single D614G mutation and the L5F+D614G variant had a higher infection ability than the wild type variant, but no differences were found between the two variants and other variants [5], [6]. This shows that the increase in the ability of infection comes from the D614G mutation itself. The higher the ability of virus infection increases its transmission capability [50]. So far, there have been no reports on the ability of infection or transmission of the D614G variant accompanied by other mutations.

Reconstruction of a phylogenetic tree shows the origin and distribution of SARS-CoV-2 in Indonesia. The Indonesian D614G variant forms one clade with all samples from comparison (19) countries. VOC B.1.1.7 Indonesia formed a separate clade from local variants and VOC B.1.1.7 from European countries and Australia. In addition, one variant with 17 mutation points forms a separate clade far from other clades. This shows that SARS-CoV-2 in Indonesia came from various countries and underwent further mutation unpredicted in Indonesia.

This study has several limitations; specifically, the sample size is limited to those submitted to GISAID^{TM} until February 27, 2021. Several things cause these limitations. First, the release of data reported to the GISAID^{TM} database takes quite a long time, sometimes it takes more than one month from sampling time until data release. Thus, by the sampling deadline of this study (12 March 2021), the available samples were obtained until one was released on February 27, 2021, while cases continue to occur in Indonesia based on data from WHO and the *Kemenkes* [51]. As long as the virus continues to spread, we believe that mutations would continue to develop in Indonesia, since mutations appear as natural by products of viral replication [52]. Second, in Indonesia, not all samples are reported for

their genome sequences. Out of the 1,410,134 million cases as of March 12, 2021 [51] reported in Indonesia, only 548 SARS-CoV-2 genome sequences were reported based on data released in the GISAID EpiCoV^{TM}. This means that not many samples were released so that the real genomic condition, including mutations and variants, can be identified.

Although the sample size is small, this is the first report on the variation of the Spike SARS-CoV-2 protein mutation in Indonesia during the year of the outbreak. The valuable data and findings can be used as a basis for molecular studies regarding the types of mutations. Also, types of variants that have been successfully mapped can be used for the development of vaccines, immunotherapy, and diagnostic tools. In addition, further research is needed to correlate mutation data and patient clinical data (if available for access), infectiousness and virulence, or viral factors in the future. Changes in molecular characteristics and morphology to viral biological functions caused by mutations were not examined in detail, because this study focused on the types of mutations and their variants.

During one year of the COVID-19 outbreak in Indonesia, many mutations and variants were found. The variants are thought to have originated from overseas which later developed and became new variants in Indonesia. Some variants are also suspected to be local variants developed in Indonesia, but further analysis is required. Changes in molecular characteristics and morphology which changes in viral biological functions caused by mutations were not examined in detail since this study focused on the types of mutations and their variants formed. Further investigations are also needed to trace the changes in the characteristics, morphology, molecular, and biological function of those variants.

AUTHORS' CONTRIBUTIONS

This research was a part of D.L. research project and carried out in collaboration. Conceptualization, mostly completed by D.L. and N.G.A.; methodology was developed by N.G.A., and D.L.; resources by A.S., S.K.H.I.; data curation and analysis by N.G.A. and I.N.O.; N.G.A and D.L prepared and wrote the original draft; D.W.P., D.N., and R.L.K. provided a massive critical revision of manuscript; visualization, N.G.A. and N.A.M.; supervision was conducted by D.L; All authors have read and agreed to the published version of the manuscript.

ACKNOWLEDGMENTS

The author would like to thanks to Dwi Mutia Prajawati and Vina Rizkiana for their support on bioinformatics technical guidance to N.G.A.

REFERENCES

[1] S.R. Weiss and S. Navas-Martin, Coronavirus Pathogenesis and the Emerging Pathogen Severe Acute Respiratory Syndrome Coronavirus. vol. 69, Microbiol. Mol. Biol. Rev., 2005. pp. 635–664. DOI: 10.1128/MMBR.69.4.635-664.2005.

[2] S.M. Haque, O. Ashwaq, A. Sarief, and A.K. Azad John Mohamed, A comprehensive review about SARS-CoV-2, vol. 15, Future Virol., 2020, pp. 625–648. DOI: 10.2217/fvl-2020-0124.

[3] Kemenkes RI, *Pedoman Pencegahan dan Pengendalian Coronavirus Disease (COVID-19)*. 2020.

[4] WHO, WHO Coronavirus (COVID-19) Dashboard, 2021. https://covid19.who.int (accessed Apr. 19, 2021).

[5] Q. Li, J. Wu, J. Nie, L. Zhang, H. Hao, S. Liu, C. Zhao, Q. Zhang, H. Liu, L. Nie, *et al.*, The Impact of Mutations in SARS-CoV-2 Spike on Viral Infectivity and Antigenicity, Cell, vol. 182, 2020, pp. 1284-1294. DOI: 10.1016/j.cell.2020.07.012.

[6] X. Li, M. Geng, Y. Peng, L. Meng, and S. Lu, Molecular immune pathogenesis and diagnosis of COVID-19, J. Pharm. Anal., Vol. 10, 2020, pp. 102–108. DOI: 10.1016/j.jpha.2020.03.001.

[7] A.A.T. Naqvi, K. Fatima, T. Mohammad, U. Fatima, I.K. Singh, A. Singh, S.M. Atif, G. Hariprasad, G.M. Hasan, and M.I. Hassan, Insights into SARS-CoV-2 genome, structure, evolution, pathogenesis and therapies: Structural genomics approach, Vol. 1866, Biochim. Biophys. Acta BBA - Mol. Basis Dis., 2020, pp. 165878. DOI: 10.1016/j.bbadis.2020.165878.

[8] Z. Abdelrahman, M.Li, and X. Wang, Comparative Review of SARS-CoV-2, SARS-CoV, MERS-CoV, and Influenza A Respiratory Viruses, Front Immunol, 2020.DOI: 10.3389/fimmu.2020.552909.

[9] R. Lu, X. Zhao, J. Li, P. Niu, B. Yang, H. Wu, W. Wang, H. Song, B. Huang, N.Zhu, *et al.*, Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding, The Lancet, vol. 395, 2020, pp. 565–574. DOI: 10.1016/S0140-6736(20)30251-8.

[10] T.S. Pillay, Gene of the month: the 2019-nCoV/SARS-CoV-2 novel coronavirus spike protein, J. Clin. Pathol., vol. 73, 2020, pp. 366–369. DOI: 10.1136/jclinpath-2020-206658.

[11] J. Lan, J. Ge, J. Yu, S. Shan, H. Zhou, S. Fan, Q. Zhang, X. Shi, Q. Wang, L. Zhang, and X. Wang, Structure of the SARS-CoV-2 spike receptor-

binding domain bound to the ACE2 receptor, Nature, vol. 581, 2020, pp. 215–220. DOI: 10.1038/s41586-020-2180-5.

[12] Z. Liu, X. Xiao, X. Wei, J. Li, J Yang, H. Tan, J. Zhu, Q. Zhang, J. Wu, and L. Liu, Composition and divergence of coronavirus spike proteins and host ACE2 receptors predict potential intermediate hosts of SARS-CoV-2, J. Med. Virol., vol. 92, 2020, pp. 595–601. DOI: 10.1002/jmv.25726.

[13] P. Liu, J.Z. Jiang, X.F. Wan, Y. Hua, L. Li, J. Zhou, X. Wang, F. Hou, J. Chen, J. Zou, and J. Chen, Are pangolins the intermediate host of the 2019 novel coronavirus (SARS-CoV-2)?, PLOS Pathogens, vol. 16, 2020. DOI: 10.1371/journal.ppat.1008421.

[14] R. Mahtarin, S. Islam, Md. J. Islam, M. O. Ullah, M. A. Ali, and M. A. Halim, Structure and dynamics of membrane protein in SARS-CoV-2, J. Biomol. Struct. Dyn., vol. 0, 2020, pp. 1–14. DOI: 10.1080/07391102.2020.1861983.

[15] K. G. Andersen, A. Rambaut, W. I. Lipkin, E. C. Holmes, and R. F. Garry, The proximal origin of SARS-CoV-2, Nat Med, vol. 26, 2020, pp. 450–452. DOI: 10.1038/s41591-020-0820-9.

[16] J. S. Mackenzie and D. W. Smith, COVID-19: a novel zoonotic disease caused by a coronavirus from China: what we know and what we don't, Microbiol. Aust., vol. 41, 2020, p. 45. DOI: 10.1071/MA20013.

[17] T.T.Y. Lam, N. Jia, Y.W. Zhang, M.H.H. Shum, J.F. Jiang, H.C. Zhu, Y.G. Tong, Y.X. Shi, X.B. Ni, Y.S. Liao, *et al.*, Identifying SARS-CoV-2-related coronaviruses in Malayan pangolins, Nature, vol. 583, 2020, pp. 282–285. DOI: 10.1038/s41586-020-2169-0.

[18] K. Dhama, S.K. Patel, K. Sharun, M. Pathak, R. Tiwari, M.I. Yatoo, Y.S. Malik, R. Sah, *et al.*, SARS-CoV-2 jumping the species barrier: Zoonotic lessons from SARS, MERS and recent advances to combat this pandemic virus, Travel Med. Infect, 2020, pp: 101830. DOI: 10.1016/j.tmaid.2020.101830.

[19] A. Bernardi *et al.*, Development and simulation of fully glycosylated molecular models of ACE2-Fc fusion proteins and their interaction with the SARS-CoV-2 spike protein binding domain, PLoS One, 2020, pp: e0237295. DOI: 10.1371/journal.pone.0237295.

[20] J.T. Ortega, M.L. Serrano, F.H. Pujol, and H.R. Rangel, Role of changes in SARS-CoV-2 spike protein in the interaction with the human ACE2 receptor: An in silico analysis, EXCLI J, 2020, pp: 410–417. DOI: 10.17179/excli2020-1167.

[21] P. Zhou *et al.*, A pneumonia outbreak associated with a new coronavirus of probable bat origin, Nature, 2020, pp: 270–273. DOI: 10.1038/s41586-020-2012-7.

[22] S.K. Saxena, S. Kumar, P. Baxi, N. Srivastava, B. Puri, and R.K. Ratho, Chasing COVID-19 through SARS-CoV-2 spike glycoprotein, VirusDis., 2020, pp: 399–407. DOI: 10.1007/s13337-020-00642-7.

[23] T. Patra *et al.*, SARS-CoV-2 spike protein promotes IL-6 trans-signaling by activation of angiotensin II receptor signaling in epithelial cells, PLoS Pathog, 2020, pp: e1009128. DOI: 10.1371/journal.ppat.1009128.

[24] S. Duffy, Why are RNA virus mutation rates so damn high?, PLoS Biol, 2018, pp: e3000003. DOI: 10.1371/journal.pbio.3000003.

[25] J.W. Drake and J.J. Holland, Mutation rates among RNA viruses, PNAS, 1999, pp: 13910–13913. DOI: 10.1073/pnas.96.24.13910.

[26] D. Haddad *et al.*, SARS-CoV-2: Possible recombination and emergence of potentially more virulent strains, PLoS One, 2021, pp: e0251368. DOI: 10.1371/journal.pone.0251368.

[27] E. Volz *et al.*, Evaluating the Effects of SARS-CoV-2 Spike Mutation D614G on Transmissibility and Pathogenicity, Cell, 2021, pp. 64-75.e11. DOI: 10.1016/j.cell.2020.11.020.

[28] Y.M. Bar-On, A. Flamholz, R. Phillips, and R. Milo, SARS-CoV-2 (COVID-19) by the numbers, eLife, 2020, pp: e57309. DOI: 10.7554/eLife.57309.

[29] S. Vilar and D.G. Isom, One Year of SARS-CoV-2: How Much Has the Virus Changed?, Biology, 2021, pp: 91. DOI: 10.3390/biology10020091.

[30] Y.J. Hou *et al.*, SARS-CoV-2 D614G variant exhibits efficient replication ex vivo and transmission in vivo, Science, 2020, pp: 1464–1468. DOI: 10.1126/science.abe8499.

[31] B. Korber *et al.*, Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus, Cell, 2020, pp: 812-827.e19. DOI: 10.1016/j.cell.2020.06.043.

[32] L. Zhang *et al.*, The D614G mutation in the SARS-CoV-2 spike protein reduces S1 shedding and increases infectivity, bioRxiv, 2020, pp: 148726. DOI: 10.1101/2020.06.12.148726.

[33] L. Zhang *et al.*, SARS-CoV-2 spike-protein D614G mutation increases virion spike density and

infectivity, Nat Commun, 2020, pp: 6013. DOI: 10.1038/s41467-020-19808-4.

[34] S. Elbe and G. Buckland-Merrett, Data, disease and diplomacy: GISAID's innovative contribution to global health, Glob Chall, 2017, pp: 33–46. DOI: 10.1002/gch2.1018.

[35] Y. Shu and J. McCauley, GISAID: Global initiative on sharing all influenza data – from vision to reality, Euro Surveill, 2017, pp: 30494. DOI: 10.2807/1560-7917.ES.2017.22.13.30494.

[36] T. A. Hall, BioEdit : a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT, Nucleic Acids Symp Ser, 1999, pp: 95–98.

[37] M. I. Khan *et al.*, Comparative genome analysis of novel coronavirus (SARS-CoV-2) from different geographical locations and the effect of mutations on major target proteins: An in silico insight, PLoS One, 2020, pp: e0238344. DOI: 10.1371/journal.pone.0238344.

[38] S. Kumar, G. Stecher, M. Li, C. Knyaz, and K. Tamura, MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms, Mol Biol Evol, 2018, pp: 1547–1549. DOI: 10.1093/molbev/msy096.

[39] A. Rambaut *et al.*, A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology, Nat Microbiol, 2020, pp: 1403–1407. DOI: 10.1038/s41564-020-0770-5.

[40] Á. O'Toole *et al.*, Tracking the international spread of SARS-CoV-2 lineages B.1.1.7 and B.1.351/501Y-V2, Wellcome Open Res, 2021, pp: 121. DOI: 10.12688/wellcomeopenres.16661.1.

[41] C. J. Harrison and J. A. Langdale, A step by step guide to phylogeny reconstruction, Plant J., 2006, pp: 561–572. DOI: 10.1111/j.1365-313X.2005.02611.x.

[42] F. D. Ciccarelli, Toward Automatic Reconstruction of a Highly Resolved Tree of Life, Science, 2006, pp. 1283–1287. DOI: 10.1126/science.1123061.

[43] I. Letunic and P. Bork, Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation, Bioinformatics, 2007, pp. 127–128. DOI: 10.1093/bioinformatics/btl529.

[44] S. J. Freeland and L. D. Hurst, The Genetic Code Is One in a Million, J Mol Evol, 1998, pp: 238–248. DOI: 10.1007/PL00006381.

[45] N. Goldman, Further results on error minimization in the genetic code, J Mol Evol, 1993, pp: 662–664. DOI: 10.1007/BF00182752.

[46] N. Goldman and Z. Yang, A codon-based model of nucleotide substitution for protein-coding DNA sequences, Mol. Biol. Evol., 1994, pp: 725–736. DOI: 10.1093/oxfordjournals.molbev.a040153.

[47] CDC, Coronavirus Disease 2019 (COVID-19), Feb. 11, 2020. https://www.cdc.gov/coronavirus/2019-ncov/variants/variant-info.html (accessed Aug. 02, 2021)

[48] R. Wang, J. Chen, K. Gao, Y. Hozumi, C. Yin, and G.-W. Wei, Characterizing SARS-CoV-2 mutations in the United States, Res Sq, 2020, pp: 49671. DOI: 10.21203/rs.3.rs-49671/v1.

[49] L. Yurkovetskiy *et al.*, Structural and Functional Analysis of the D614G SARS-CoV-2 Spike Protein Variant, Cell, 2020, pp. 739-751.e8. DOI: 10.1016/j.cell.2020.09.032.

[50] N. H. L. Leung, Transmissibility and transmission of respiratory viruses, Nat Rev Microbiol, 2021, pp: 528–545. DOI: 10.1038/s41579-021-00535-6.

[51] Kemenkes RI, Situasi Terkini Perkembangan Coronavirus Disease (COVID-19) 7 Juli 2021, 2021. https://covid19.kemkes.go.id/situasi-infeksi-emerging/situasi-terkini-perkembangan-coronavirus-disease-covid-19-7-juli-2021 (accessed Aug. 02, 2021)

[52] N.D. Grubaugh, M.E. Petrone, and E.C. Holmes, We shouldn't worry when a virus mutates during disease outbreaks, Nat. Microbiol., 2020, pp: 529–530. DOI: 10.1038/s41564-020-0690-4.