Molecular and Cellular Biomedical Sciences, Vol.6 No.1, March 2022, p.20-7

Print ISSN: 2527-4384, Online ISSN: 2527-3442
DOI: 10.21705/mcbs.v6i1.221

RESEARCH ARTICLE

MCBS
Mol Cell Biomed Sci. 2022; 6(1): 20-7
DOI: 10.21705/mcbs.v6i1.221

Non-Synonymous Mutation Analysis of SARS-CoV-2 ORF3a in Indonesia

Hartiyowidi Yuliawuri, Jeanne Elvia Christian, Nathanael Steven

Biomedical Science Study Program, Calvin Institute of Technology, Jakarta, Indonesia

Background: The report of mutation sites ORF3a SARS CoV-2 in Indonesia is still limited. Some research showed that mutations in ORF3a protein might alter SARS-CoV-2 pathogenesis. Observation of new variants should be conducted as a risk monitoring framework.

Materials and method: We assessed the impact of mutations in ORF3a protein by analyzing 3,751 SARS-CoV-2 DNA sequences from the GISAID database from March 2020 until July 2021. The whole-genome sequences were aligned using Clustal Omega Multiple Sequence Alignment from EMBL-EBI and analyzed using BioEdit version 7.2.5 software. The reference whole genome sequence was taken from the Genbank database with accession number NC045512. We excluded the samples containing N letters due to inaccurate reading. Effect of point mutations on protein structure was analyzed using PredictProtein (*https://predictprotein.org*) and Protein Variation Effect Analyzer (PROVEAN) v1.1.3. online software.

Results: We identified five most frequent non-synonymous mutations in ORF3a protein of SARS-CoV-2 which were Q57H (58.04%), S26L (27.25%), S220I (10.37%), D155H (8.98%), and P104S (5.47%).

Conclusion: These mutation data showed the phenomenon of amino acid changes in ORF3a SARS-CoV-2 in Indonesia until July 2021. The implication of this mutation needs to be determined in further studies.

Keywords: *Indonesia, mutations, non-synonymous, SARS-CoV-2, whole genome*

Introduction

Corona Virus Disease 19 (COVID-19) was first reported in Wuhan, China on December 31st, 2019. This disease was caused by a novel beta coronavirus believed to be of zoonotic origin, named SARS-CoV-2.[1] In a short time, this virus spread to other countries and created pandemic cases. The first genome sequence of SARS-CoV-2 was published on January 5th, 2020. Since then, genomic sequences have been published from all around the world. In Indonesia, the first case of COVID-19 was announced on March 2nd, 2020. Since then, the virus has been spreading exponentially. Until this study was conducted, there were more than 3.9 million cases, with more than 330,000 active cases in Indonesia. Cases of people being examined are also increasing, with more than 20 million people.[2]

Due to its high infection rate and RNA molecule as the main composition of its genome, SARS-CoV-2 has a high mutation rate. There are many mutations found in the SARS-CoV-2 genome and the effect of these mutations is

Date of submission: September 13, 2021
Last Revised: October 10, 2021
Accepted for publication: October 15, 2021

Corresponding Author:
Hartiyowidi Yuliawuri
Biomedical Science Study Program
Calvin Institute of Technology, Calvin Tower RMCI
Jl. Industri Blok B14, Jakarta 10610, Indonesia
e-mail: hartiyowidi.yuliawuri@calvin.ac.id

Cell and
Biopharmaceutical
Institute

still unknown. SARS-CoV-2 genome contains genes for Spike, Membrane, Envelope, Nucleocapsid proteins and non-structural proteins (NSP1−NSP16). Besides that, there are 6 accessory proteins included in the viral pathogenicity with ORF3a as the largest one containing 247 amino acids.[3] Several studies state that mutations in ORF3a protein of SARS-CoV-2, might alter its pathogenesis and lead the fourth epidemic wave in Hong Kong.[4,5]

Currently, there are limited reports regarding the ORF3a variance of SARS-CoV-2 in Indonesia and especially non-synonymous mutation that makes changes to protein sequences. This research aims to report ORF3a non-synonymous mutations in SARS-CoV-2 from genome sequences submitted to GISAID from March 2020 until July 2021. The result presented here could help for a better understanding of the presence of various genetic varieties of SARS-CoV-2 in Indonesia and as preliminary study to predict the impact of these mutations on the prevention, diagnosis, and treatment of COVID-19 in Indonesia.

Materials and methods

ORF3a DNA Sequences Retrieval from Database

A total of 3,916 SARS-CoV-2 sequences originated from Indonesia, from March 2nd, 2020 until July 31st, 2021 were acquired from GISAID (*www.gisaid.org*).[6] These DNA sequences consisted of 3,883 whole-genomes and 33 partial genomes.

Multiple Sequences Alignment and Analysis of Frequent Non-synonymous Mutations of ORF3a

The DNA sequences that contained N letters were removed from the analysis, because it would cause undetermined or inaccurate amino acid reading when translated to protein sequences. The 33 partial genomes were excluded, because these were partial sequences that do not contain ORF3a sequence. Therefore, a total of 3,751 SARS-CoV-2 sequences were analyzed further. These DNA sequences were aligned using Clustal Omega Multiple Sequence Alignment from EMBL-EBI (*www.ebi.ac.uk/Tools/msa/clustalo/*). The DNA alignment sequences were analyzed in BioEdit version 7.2.5 software to make ORF3a protein. We compare the amino acid changes with isolate hCoV-19/Wuhan/WIV04/2019 as reference sequence from GISAID (EPI_ISL_402124). Furthermore, non-synonymous mutations of ORF3a protein were recorded to know the position and the changes of amino acid compared to reference sequences.

We focused on the top five most frequent non-synonymous mutations of ORF3a protein. We were mapping the distribution of these frequent non-synonymous mutations throughout 34 provinces in Indonesia using Indonesia Map by Vemaps.com as a blank template map.[7] We calculated the percentages of carrying mutated variants of each mutation position in a single province. These data only reflect the percentages of the mutated variant, not the number of samples, because one sample may contain more than one non-synonymous mutation.[8] We also reported variants of concern (VOC) using PANGOLIN annotations available in metadata from the GISAID database of the respective genome until this study was conducted. We calculated the number of samples and percentages of carrying mutations in five most frequent non-synonymous mutations of ORF3a protein monthly, from March 2nd, 2020 until July 31st, 2021 in Indonesia.

Simulation of ORF3a Protein Structure

We predicted the structure of protein sequences which contained a single or combination of five most frequent non-synonymous mutations using PredictProtein Open (*https://predictprotein.org*). Subsequently, we compared this prediction with wild type ORF3a as the reference sequence protein structure. We predicted the amino acid changes to check whether it has an impact on the biological function of protein. By submitting the protein sequences to Protein Variation Effect Analyzer (PROVEAN) v1.1.3. software from the website, we get the PROVEAN score with -2.5 as a cut off. More positive score or above -2.5 indicates that the variation effect on protein was neutral and vice versa.[9]

Results

We analyzed the 3,751 ORF3a DNA sequences of SARS-CoV-2 after removing sequences containing N letters. Based on protein data alignment, we recorded 203 positions that showed the changes of amino acid, called non-synonymous mutations. Furthermore, we found five most frequent non-synonymous mutations in our sample were Q57H, S26L, S220I, D155H, and P104S (Figure 1). All these five mutations were scattered in Indonesia with the percentages of 58.04%, 27.24%, 10.37%, 8.98%, and 5.46%, respectively. We also reported variant of concern (VOC) in our data from GISAID metadata. We observed that the highest percentage was Delta variant (B.1.617.2) in 25.73% (965 out of 3,751 sequences), Alpha variant

Print ISSN: 2527-4384, Online ISSN: 2527-3442
DOI: 10.21705/mcbs.v6i1.221

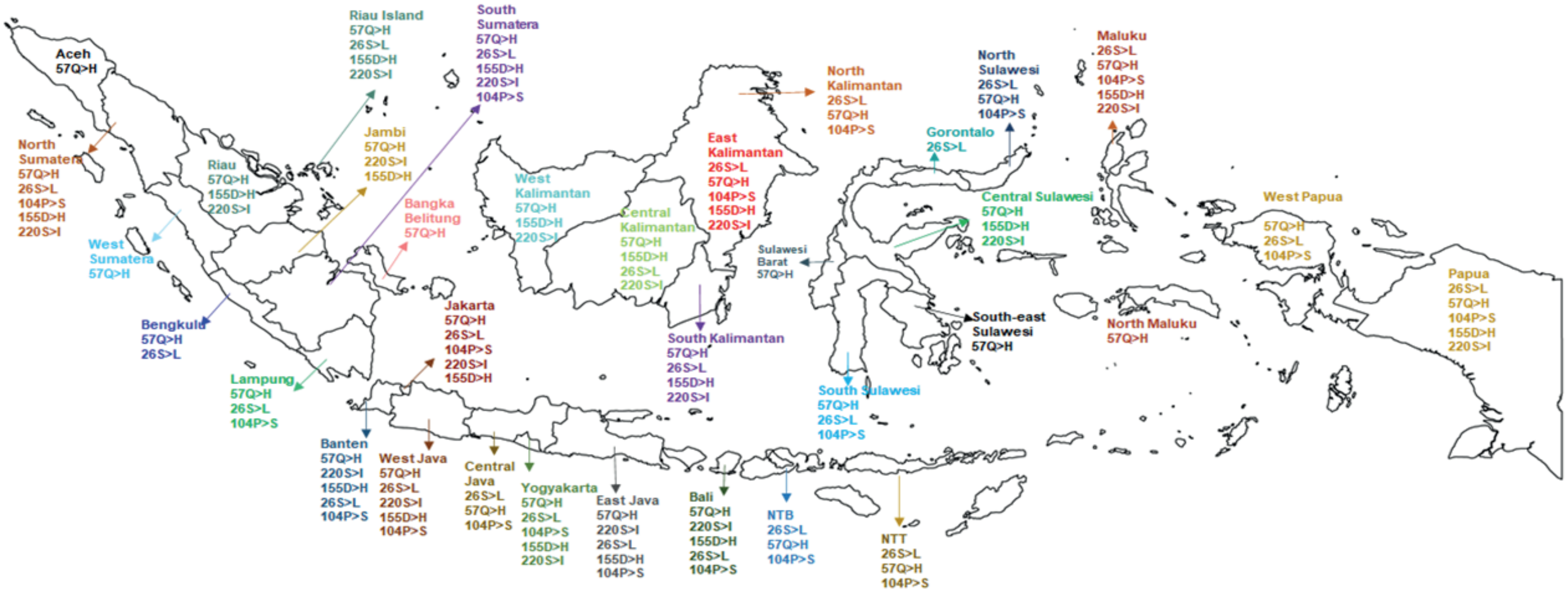

Figure 1. Five most frequent non-synonymous mutation in different provinces of Indonesia.

Table 1. List of provincial origin of ORF3a protein non-synonymous mutations in Indonesia (part I).

Province	Number of Samples	Sequence Location	Amino Acid Changes	Individuals with Mutation	Percentages of Mutation-carrying Samples
Aceh	13	57	Q-->H	13	100
North Sumatera	98	26	S-->L	22	22.45
		57	Q-->H	65	66.33
		104	P-->S	2	2.04
		155	D-->H	2	2.04
		220	S-->I	2	2.04
West Sumatera	87	57	Q-->H	60	68.97
Riau	12	57	Q-->H	10	83.33
		155	D-->H	3	25
		220	S-->I	3	25
Riau Island	131	26	S-->L	2	1.53
		57	Q-->H	121	92.37
		155	D-->H	2	1.53
		220	S-->I	2	1.53
Bengkulu	8	26	S-->L	3	37.5
		57	Q-->H	5	62.5
Jambi	102	57	Q-->H	89	87.25
		155	D-->H	7	6.86
		220	S-->I	15	14.71
South Sumatera	42	26	S-->L	11	26.19
		57	Q-->H	27	64.29
		104	P-->S	1	2.38
		155	D-->H	2	4.76
		220	S-->I	2	4.76
Lampung	10	26	S-->L	3	30
		57	Q-->H	6	60
		104	P-->S	1	10
Bangka Belitung	1	57	Q-->H	1	100
Banten	150	26	S-->L	17	11.33
		57	Q-->H	102	68
		104	P-->S	12	8
		155	D-->H	23	15.33
		220	S-->I	26	17.33
Jakarta	900	26	S-->L	310	34.44
		57	Q-->H	378	42
		57	Q-->N	1	0.11
		57	Q-->R	3	0.33
		104	P-->S	92	10.22
		155	D-->H	72	8
		220	S-->I	80	8.89
West Java	569	26	S-->L	231	40.6
		57	Q-->H	267	46.92
		57	Q-->R	1	0.18
		104	P-->S	35	6.15
		155	D-->H	55	9.67
		220	S-->I	66	11.6
Yogyakarta	71	26	S-->L	17	23.94
		57	Q-->H	45	63.38
		104	P-->S	3	4.23
		155	D-->H	1	1.41
		220	S-->I	1	1.41
Central Java	213	26	S-->L	151	70.89
		57	Q-->H	57	26.76
		104	P-->S	3	1.41
East Java	180	26	S-->L	18	10
		57	Q-->H	141	78.33
		104	P-->S	12	6.67
		155	D-->H	16	8.89
		220	S-->I	30	16.67

Table 1. List of provincial origin of ORF3a protein non-synonymous mutations in Indonesia (part II).

Province	Number of Samples	Sequence Location	Amino Acid Changes	Individuals with Mutation	Percentages of Mutation-carrying Samples
Bali	457	26	S-->L	24	5.25
		57	Q-->H	364	79.65
		104	P-->S	7	1.53
		155	D-->H	132	28.88
		220	S-->I	140	30.63
NTB	28	26	S-->L	16	57.14
		57	Q-->H	11	39.29
		104	P-->S	4	14.29
NTT	66	26	S-->L	39	59.09
		57	Q-->H	24	36.36
		104	P-->S	3	4.55
Central Kalimantan	37	26	S-->L	3	8.11
		57	Q-->H	33	89.19
		155	D-->H	4	10.81
		220	S-->I	3	8.11
East Kalimantan	61	26	S-->L	31	50.82
		57	Q-->H	30	49.18
		104	P-->S	6	9.84
		155	D-->H	1	1.64
		220	S-->I	1	1.64
South Kalimantan	49	26	S-->L	1	2.04
		57	Q-->H	45	91.84
		155	D-->H	1	2.04
		220	S-->I	1	2.04
North Kalimantan	30	26	S-->L	16	53.33
		57	Q-->H	12	40
		104	P-->S	3	10
West Kalimantan	23	57	Q-->H	19	82.61
		155	D-->H	1	4.35
		220	S-->I	1	4.35
Central Sulawesi	4	57	Q-->H	4	100
		155	D-->H	1	25
		220	S-->I	1	25
Gorontalo	3	26	S-->L	1	33.33
North Sulawesi	20	26	S-->L	7	35
		57	Q-->H	5	25
		104	P-->S	2	10
Southeast Sulawesi	3	57	Q-->H	2	66.67
West Sulawesi	23	57	Q-->H	22	95.65
South Sulawesi	65	26	S-->L	14	21.54
		57	Q-->H	49	75.38
		57	Q-->R	1	1.54
		104	P-->S	8	12.31
Maluku	13	26	S-->L	9	69.23
		57	Q-->H	1	7.69
		104	P-->S	1	7.69
		155	D-->H	1	7.69
		220	S-->I	1	7.69
North Maluku	1	57	Q-->H	1	100
Papua Barat	27	26	S-->L	12	44.44
		57	Q-->H	15	55.56
		104	P-->S	3	11.11
Papua	23	26	S-->L	10	43.48
		57	Q-->H	8	34.78
		104	P-->S	3	13.04
		155	D-->H	1	4.35
		220	S-->I	1	4.35

(B.1.1.7) in 1.63% (61 out of 3,751 sequences), and Beta variant (B.1.351) in 0.35% (13 out of 3,751 sequences).

We calculated the percentage of carrying mutations per province in Indonesia. There were 34 provinces with variable sequence numbers and mutations in ORF3a protein (Table 1). The highest number of samples was from Jakarta with 900 samples. We mapped the presence of five most frequent mutations in each province, with the highest mutation at the 1st order. According to our mutation analysis data until July 2021, the mutation Q57H was found in April 2020 and reached the highest percentage almost one year later in March 2021. The mutation S220I and D155H was discovered for the first time in August 2020 and their highest percentage was in February 2021. While the S26L mutation was first found in January 2021 and P104S in April 2021 and both of them had their highest percentage in July 2021. Overall, the highest total percentage for all types of mutations first occurred in January 2021 and the second peak in June 2021.

Table 2 and Figure 2 showed the Percentages of mutation-carrying samples in five most frequent amino acid changes in SARS-CoV-2 ORF3a, meanwhile Table 3 showed the incidence and PROVEAN score of five most frequent non-synonymous mutation.

Discussion

We recorded five most frequent non-synonymous mutations in ORF3a protein that circulate in Indonesia's samples. Based on the exclusion requirement mentioned previously, 3,751 complete sequences of ORF3a SARS-CoV-2 from 34 provinces data samples (March 2nd, 2020 to July 31st, 2021) were retrieved. Non-synonymous mutations at amino acid site 57 (Q→H) of ORF3a is detected in 2,177 sequences (58.04%) and scattered in 34 provinces in Indonesia. Moreover, there were another amino acid changes in site 57 of ORF3a, Glutamine to Asparagine (Q→N) and Glutamine to Arginine (Q→R). They were found in Jakarta and South Sulawesi provinces (Q→R only) in <0.05% of the samples. The protein coded by ORF3a in SARS-CoV-2 interacts with immune system signaling pathways, such as the TLR pathway and T cell pathway. This was predicted to be an immune evasion mechanism and deleterious to the immune response against SARS-CoV-2 infection.[10] There are six functional domains owned by ORF3a SARS-CoV2, that are based on its role in the host cell. The TRAF3-binding motif that was found to activate NF-kB and the NLRP3 inflammasome was located in domain II (PLQAS, position 36-40), the K^+ ion channel and cysteine-rich domain was

Table 2. Monthly percentages of mutation-carrying samples.

Month	Number of Samples	Percentages of Mutation-carrying Samples in 5 Frequent Amino Acid Changes in ORF3a SARS-Cov-2				
		Q57H	**S26L**	**S220I**	**D155H**	**P104S**
March 2020	30	0.00*	0.00*	0.00*	0.00*	0.00*
April 2020	35	31.43	0.00*	0.00*	0.00*	0.00*
May 2020	18	50	0.00*	0.00*	0.00*	0.00*
June 2020	44	52.27	0.00*	0.00*	0.00*	0.00*
July 2020	35	54.29	0.00*	0.00*	0.00*	0.00*
August 2020	46	73.91	0.00*	2.17	2.17	0.00*
September 2020	48	79.17	0.00*	0.00*	0.00*	0.00*
October 2020	58	70.69	0.00*	5.17	1.72	0.00*
November 2020	79	56.96	0.00*	3.8	2.53	0.00*
December 2020	195	67.69	0.00*	14.87	10.26	0.00*
January 2021	608	78.78	0.99	20.39	16.45	0.00*
February 2021	401	83.04	0.00*	26.93	25.94	0.00*
March 2021	416	88.7	0.96	20.43	18.03	0.00*
April 2021	244	68.44	13.93	9.02	8.61	1.23
May 2021	524	62.21	29.2	2.29	2.29	2.67
June 2021	910	15.93	85.16	0.11	0.00*	18.46
July 2021	58	8.62	86.21	0.00*	0.00*	34.48

*Non-synonymous mutations were not found in samples.

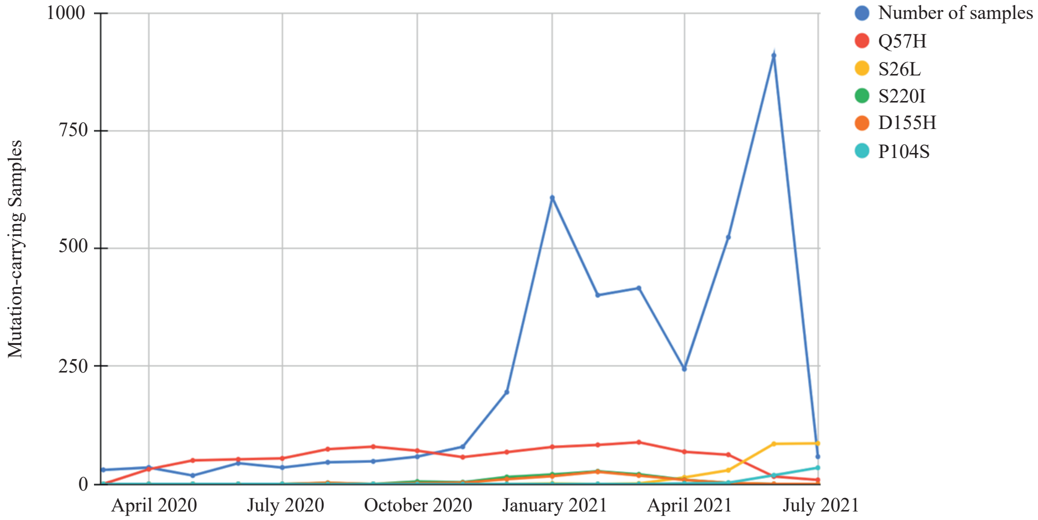

Figure 2. Percentages of mutation-carrying samples in five most frequent amino acid changes in SARS-CoV-2 ORF3a.

located in domain III (position 91-133), a caveolin-binding motif was located in domain IV (position 141-149), and the conserved motif in domain V (position 160-163) for ORF3a protein to be transported from Golgi apparatus to plasma membrane.[11]

These top five most frequent mutations are observed throughout the 34 provinces in Indonesia, and we found some mutations that are dominant to specific provinces. Aceh, Bangka Belitung, West Sumatera, and South-east Sulawesi carry the mutation at the position 57 (Q→H) only. Another province, Gorontalo show the mutation at the position 26 (L→S) only. Based on data, L26S mutation detected in 1,022 (27.25%) in Indonesia. There were 9 provinces that were dominant with this mutation (Figure 1). The amino acid changes in position 26 Leucine to Serine. Moreover, there is a stop codon in this position which belongs to Bali samples (<0.03%).

We observed two mutations appeared together with Q57H, those were S220I and D155H. These combinations occurred in 335 samples (9.4%). The side-chain in position 220 serine was polar amino acid and became hydrophobic (Isoleucine) whereas in position 155 Aspartic acid was negative turned to positive amino acid (Histidine), but their secondary structure did not show any changes. The double mutant Q57H+A99V in Europe emerged during the early pandemic. In August 2020, double mutant Q57H+ S58N emerged in the Netherlands. Another variant Q57H+ G172V was found in the top five most relatively frequent mutations observed in the USA. The presence of single mutation of A99V, S58N and G172V cause an effect of either destabilizing (A99V, S58N) or stabilizing (G172V) ORF3a protein.[12]

Q57H mutation was found to have deleterious effects based on PROVEAN score.[9,10] Even though their secondary structure did not show any changes. The protein coded by ORF3a in SARS-CoV-2 could interact with spike protein, resulting in delayed premature viral RNA release. Interaction between ORF3a protein and spike protein can also enhance

Table 3. The incidence and PROVEAN score of five most frequent non-synonymous mutation.

Amino Acids Substitution in ORF3a	Incidence	PROVEAN Score*	Variation Effect on Protein
S26L	27.24%	-2.314	Neutral
Q57H	58.04%	-3.286	Deleterious
P104S	5.46%	-3.343	Deleterious
D155H	8.98%	-4.533	Deleterious
S220I	10.37%	-3.029	Deleterious

*The cut-off value was -2.5.

viral packaging.[13] This interaction benefits the survival and infectivity of SARS-CoV-2. With a similar sequence and structure, it can be predicted that ORF3a protein and spike protein in SARS-CoV-2 will also interact in the same manner. Q57H mutation in ORF3a was predicted to increase ORF3a-spike protein interaction. Q57H mutation increased binding affinity against spike protein by 4.2 kcal/mol compared to wild type.[14] Around 58.04% of the samples have Q57H mutation, this shows that Q57H mutation could increase survivability and benefit the virus in an infection. However, Q57H mutation was not located in the six functional domains of ORF3a SARS-CoV-2. Similar results with the other four most frequent non-synonymous mutations that we found in this study.[11]

Based on PROVEAN score prediction, the S220I, D155H, and P104S mutations had deleterious effects, but not for S26L mutation (Table 3). The biological effect of S26L, S220I, D155H, and P104S mutations for the pathogenicity of the virus are still unknown but may have some effect towards the survival and spread of the virus.

Based on lineage data downloaded from GISAID on July 31st, 2021, we reported the existence of Alpha, Beta, and Delta as variants of concern (VOC) in Indonesia. The Delta variant circulates rapidly, increasing in number during 6 months after first detected in January 2021. There were 965 of 3,883 (24.85%) samples that belong to the Delta variant. Three out of five most frequent mutations of ORF3a gene in our study belong to combination variants of concern (Q57H, S26L, and P104S). This new VOC data could be important to monitor the variants that circulate in specific regions and study the transmissibility, pathogenicity, and severity of SARS-CoV-2.

We highlight the most frequent mutations of ORF3a protein at the peak of the pandemic in Indonesia in January, June, and July 2021 (Table 2). The highest number of COVID-19 confirmed cases in Indonesia leading national regulation to limited mobilization, called Community Activities Restrictions Enforcement. In January, June, and July 2021, the most frequent mutations were Q57H, S220I, and S26L respectively. The most frequent mutation Q57H was not found in samples from June 2021.

Conclusion

This study identified five most frequent non-synonymous mutations of SARS-CoV-2 ORF3a sequences. We need further studies to understand the implication of this mutation on the function of ORF3a and for the circulation of virus in Indonesia or globally. With the integrated network from the government, clinician, researcher, and other institutions, hopefully we can understand the mutation pattern and make predictions to prevent the occurrence of another wave of new SARS-CoV-2 cases in Indonesia.

References

1. van Dorp L, Acman M, Richard D, Shaw LP, Ford CE, Ormond L, *et al.* Emergence of genomic diversity and recurrent mutations in SARS-CoV-2. Infect Genet Evol. 2020; 83: 104351. doi: 10.1016/j.meegid.2020.104351.
2. Satuan Tugas Penanganan COVID-19 [Internet]. 2021. Peta Sebaran [updated 2021 Agt 20; cited 2021 Agt 20]. Available from: https://covid19.go.id/peta-sebaran.
3. Majumdar P, Niyogi S. ORF3a mutation associated with higher mortality rate in SARS-CoV-2 infection. Epidemiol Infect. 2020; 148: e262. doi: 10.1017/S0950268820002599.
4. Azad GK, Khan PK. Variations in Orf3a protein of SARS-CoV-2 alter its structure and function. Biochem Biophys Rep. 2021 ;26: 100933. doi: 10.1016/j.bbrep.2021.100933.
5. Chu DKW, Hui KPY, Gu H, Ronald LW, Krishnan P, Daisy, YMN, *et al.* Introduction of ORF3a-Q57H SARS-CoV-2 variant causing fourth epidemic wave of COVID-19, Hong Kong, China. Emerging Infectious Diseases. Emerg Infect Dis. 2021; 27(5): 15. https://doi.org/10.3201/eid2705.210015.
6. GISAID - Initiative [Internet]. Freunde von GISAID e.V; 2008-2021 [cited 2021 July 31]. Available from: https://www.gisaid.org/.
7. Vemaps [Image on internet]. Outline Map of Indonesia with Provinces [cited 2021 Agt 20]. Available from: https://vemaps.com/indonesia/id-02.
8. Banerkee S, Seal S, Dey R, Mondal KK, Bhattacharjee P. Mutational spectra of SARS-CoV-2 orf1ab polyprotein and signature mutations in the United States of America. J Med Virol. 2021; 93(3): 1428-35.
9. J Craig Venter Institute [Internet]. PROVEAN Score [cited 2021 Aug 31]. Available from: http://provean.jcvi.org/index.php.
10. Rehman S, Mahmood T, Aziz E, Batool R. Identification of novel mutations in SARS-CoV-2 isolates from Turkey. Arch Virol. 2020; 165(12): 2937-44.
11. Issa E, Merhi G, Panossian B, Salloum T, Tokajian S. SARS-CoV-2 and ORF3a: Nonsynonymous Mutations, Functional Domains, and Viral Pathogenesis. mSystems. 2020; 5(3): e00266-20. https://doi.org/10.1128/mSystems.00266-20.
12. Bianchi M, Borsetti A, Ciccozzi M, Pascarella S. SARS-CoV-2 ORF3a: Mutability and function. Int J Biol Macromol. 2021; 170: 820-6.
13. Tan, YJ. The Severe Acute Respiratory Syndrome (SARS)-coronavirus 3a protein may function as a modulator of the trafficking properties of the spike protein. Virol J. 2005; 2: 5. doi: 10.1186/1743-422X-2-5
14. Wu S, Tian C, Liu P, Guo D, Zheng W, Huang X, *et al.* Effects of SARS-CoV-2 mutations on protein structures and intraviral protein–protein interactions. J Med Virol. 2020; 93(4): 2132-40.