BIODIVERSITAS
Volume 25, Number 4, April 2024
Pages: 1781-1787

ISSN: 1412-033X
E-ISSN: 2085-4722
DOI: 10.13057/biodiv/d250448

Short Communication: Genetic diversity and identification of unique Omicron variant markers of SARS-CoV-2 in Indonesia

IDA B.K. SUARDANA[1,♥], BAYU K. MAHARDIKA[2], ADHIYOGA A.A.N.A. SANTOSA[3], GUSTI N. MAHARDIKA[4]

[1]Virology Laboratory, Faculty of Veterinary Medicine, Universitas Udayana. Jl. P.B. Sudirman, Denpasar 80234, Bali, Indonesia.
Tel./fax.: +62-361-223791, ♥email: idasuardana@unud.ac.id
[2]Animal Biomedical and Molecular Biology Laboratory, Universitas Udayana. Jl. Sesetan-Markisa 6A, Denpasar 80223, Bali, Indonesia
[3]Microbiology Laboratory, Faculty of Medicine, Universitas Udayana. Jl. Sesetan-Markisa 6A, Denpasar 80223, Bali, Indonesia
[4]Animal Biomedical and Molecular Biology Laboratory, Universitas Udayana. Jl. Sesetan-Markisa 6A, Denpasar 80223, Bali, Indonesia

Manuscript received: 7 February 2024. Revision accepted: 27 April 2024.

Abstract. *Suardana IBK, Mahardika BK, Santosa AAANA, Mahardika GN. 2024. Short Communication: Genetic diversity and identification of unique Omicron variant markers of SARS-CoV-2 in Indonesia. Biodiversitas 25: 1781-1787.* The rapid and global spread of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) has caused enormous socioeconomic disruptions, such as local and global travel, country or city lockdown, and manymore. We tested the hypothesis that global introduction and local substitution markers that are unique to Indonesia due to the global travel restrictions could be identified. All complete and high-coverage genomes of all Omicron variant sequences submitted from Indonesia were downloaded from the Global Initiative on Sharing All Influenza Data (GISAID), in October 27, 2022, and aligned with the Wuhan-Hu-1 strain and randomly selected Nextstrain Omicron clades. Genetic diversity was calculated using Tajima's neutrality test, and phylogeny was constructed using the neighbour-joining method in Mega 11. The overall whole genome diversity of the Omicron variant detected in Indonesia is 0.0008, with the three highest coding region diversities being Open Reading Frame 6 (ORF6), Envelope, and Spike protein. The Tajima D value of the whole genome is -2.669, while the three highest Tajima values are ORF6, Envelope and ORF10. The phylogeny shows that there are 168 sequences of Indonesian Omicron variants, which do not share branches with Nextstrain Omicron clades. Unique amino acid substitutions to Wuhan-Hu-1 of the Omicron variants detected in Indonesia occurred in some coding regions, mainly in ORF1AB and Spike. We encourage country-based analysis as a lesson for future pandemic events.

Keywords: Indonesia, Omicron SARS-CoV-2, Tajima test, virus genetic diversity

Abbreviations SARS-CoV-2: Severe Acute Respiratory Syndrome Coronavirus 2, ORF: Open Reading Frame, CFR: Case Fatality Rate

INTRODUCTION

The Coronavirus Disease pandemic 2019 (COVID-19), due to the rapid and global spread of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2), has caused enormous socioeconomic disruptions, including global travel. Various countries have implemented travel restrictions as an "emergency break" to curb virus introduction. Although the end of the epidemic in some countries seems to be on the horizon following mass vaccination, some strains of SARS-CoV-2 are feared to cause new waves of infection due to vaccine breakthrough as it has been observed in many parts of the world (Boekel et al. 2022; Johnson et al. 2022; Kared et al. 2022).

Although travel restrictions have indeed led to important changes in the dynamics of the early phases of the COVID-19 pandemic (Grépin et al. 2021), national borders remain porous. SARS-CoV-2 continues its global circulation, and community transmission is still occurring silently. Fast evolution of the virus has led to the generation of many variants. The most recent and dominant variant is Omicron, which suppressed the circulation of other variants (Meo et al. 2021; Boekel et al. 2022; Mahardika et al. 2022). This variant has rapidly separated into at least eight clades (https://nextstrain.org/ncov/gisaid/global/6m), designated as 21K, 21L, 22A, 22B, 22C, 22D, 22E, and 22F clades as well as various lineages. The WHO's Technical Advisory Group on SARS-CoV-2 Virus Evolution (TAG-VE) concerned that the XBB and BQ was terrified to cause a new wave of infections and mortalities overcoming colossal global Covid-19 immunization (https://www.who.int/). We previously have analyzed the whole genome of the clades and lineages and found the global spread of strains with specific indels and polymorphic amino acids, probable reverted indels/substitutions and indirect evidences of SARS-CoV-2 attenuation in Omicron clades (Suharsono et al. 2023).

This modern pandemic is also characterized by high number of complete virus genomic data in public databases. As of December 29th, 2023, the number of complete genome submissions of SARS-CoV-2 in GISAID (Khare et al. 2021) reached more than 16 million. Meaningful analysis should be conducted and interpreted so that huge expenditures, technology, and human resources can be transformed into scientifically sound political decisions such as risk communication, detection strategy, and vaccine development, which will benefit communities for better

pandemic management. To our knowledge, papers covering the diversity of SARS-CoV-2 in certain countries are limited. A study in India covered the distribution of clades in different states (Yadav et al. 2021), but not the genetic diversity of certain clades. In this paper, we started with the data of the Omicron variant detected in Indonesia as an independent and comparative analysis of the different Omicron sublineages as encouraged by the WHO (https://www.who.int/tag-ve-statement). This country is among the most severely affected countries by the COVID-19 pandemic, with a high Case Fatality Rate (CFR) of approximately 2.5%, while the global CFR is approximately 1% (https://www.worldometers.info/coronavirus/). We hypothesized that it must have been global introduction as well as local substitution markers that were unique to Indonesia due to the national implementation of global travel restrictions. There has been a study covering genetic variation of SARS-CoV-2 in Indonesia (Ansori et al. 2020), however the number of sequence data was too small and covering the spike protein only. The paper did not include the Omicron variant. Papers covering large number of full genome data from Thailand and Brazil have been published and focusing on virus pathogenicity and mortality (Aiewsakun et al. 2021; Hahn et al. 2021). Both did not specifically include Omicron variant.

The objective of this study is to compare full genome sequence of a large amount of data of Omicron variant of SARS-CoV-2 from Indonesia to find the genetic diversity and to identify unique Omicron variant markers.

MATERIALS AND METHODS

The full genomes of all Omicron variant sequences submitted from Indonesia, which were annotated as complete and had high coverage, were downloaded from GISAID (Elbe and Buckland-Merrett 2017; Shu and McCauley 2017; Khare et al. 2021) in October 27, 2022. The complete genome, ORF1AB, and spike coding region were aligned online with Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/) in separate files. The outputs were merged in Mega 11 (Tamura et al. 2021), and the alignment was corrected manually. Sequences with a long tract of more than two Ns as well as sequences that contained singleton insertions/deletions were excluded. The other coding regions were aligned using Clustal W in Mega11 (Tamura et al. 2021). Genetic diversity was calculated using Tajima's neutrality test, and phylogeny was constructed using the neighbour-joining protocol in Mega 11. Tajima's neutrality test was applied to the whole genome and all coding regions of SARS-CoV-2.

In the phylogeny analysis, two representatives of Nextstrain Omicron clades (21K, 21L, 22A, 22B, 22C, 22D, 22E, and 22F), as well as original SARS-CoV-2 of Wuhan-Hu-1 strains were included. Sequences with a long tract of more than two Ns as well as sequences that contained singleton insertions/deletions were excluded. Branches with Nextstrain clades as well as those with more than 10 taxa were collapsed. The molecular markers unique to Indonesia were searched in all coding regions by aligning with the Wuhan-Hu-1 sequence and fifteen to twenty-five sequences of the Omicron clades randomly selected from GISAID Nextstrain.

RESULTS AND DISCUSSION

Results

The sequence data of complete and high coverage Omicron variant of SARS-CoV-2 from Indonesia are available at the GISAID with identifier EPI_SET_230109nt, DOI: 10.55876/gis8.230109nt. The sequence data of 15 to 25 representatives of Nexstrain clades are accessible at GISAID with identifier EPI_SET ID: EPI_SET_230327ca and DOI: 10.55876/gis8.230327ca as previously published (Suharsono et al. 2023).

Tajima's neutrality test of the whole genome and all coding regions of Omicron variant sequences detected in Indonesia is presented in Table 1. The data included the number of sequences, total number of sites, number of segregating sites, nucleotide diversity, and the Tajima test statistic. The overall whole genome diversity of the Omicron variant detected in Indonesia is 0.0008. The three highest coding region diversities are the Open Reading Frame 6 (ORF6), Envelope, and Spike protein regions. The Tajima D value of the whole genome is -2.669, while the three highest Tajima values are ORF6, Envelope and ORF10, and the lowest are ORF7A, ORF1AB, and ORF8.

The phylogeny of Omicron variants in Indonesia uploaded in GISAID is presented in Figure 1. There are 168 sequences of Indonesian Omicron variants that do not share branches with Nextstrain Omicron clades.

The complete list of amino acid substitutions and indels of Omicron variants detected in Indonesia as well as a representative of each of Nextstrain Omicron clades compared to Wuhan-Hu-1 is available in Table S1. Unique amino acid substitutions of Omicron variants detected in Indonesia are tabulated in Table 2. Amino acid substitutions unique to Omicron variants detected in Indonesia occurred in all coding regions, except Matrix, Nucleocapsid (NC), ORF6, and ORF8. Six unique sites occurred in ORF1AB and four in spike.

Discussion

The population genetic approach applied in this study is not new. At the beginning of the emergence of the Omicron variant, the Tajima D statistic was applied (Yeh and Contreras 2021). Previously, this protocol has been conducted to predict the spread of pandemic influenza H1N1 2009 (Shiino et al. 2010; Kim et al. 2017), other RNA viruses (Bhatt et al. 2010) and other infectious diseases (Omori and Wu 2017). The problem in applying population genetics to SARS-CoV-2 is that the population is very dynamic. The current pandemic shows that even with restricted travel and the requirement of negative molecular testing prior to boarding, SARS-CoV-2 isconstantly spreading globally (Siwiak et al. 2020; Arora et al. 2021; Chen et al. 2021; Aboura 2023; Satapathy et al. 2024). Very low levels of virus excretion as well as contamination of clothes, vehicles, and transported goods must have occurred.

Table S1. Number of Omicron variants SARS-CoV-2 detected in Indonesia posing substituted amino acids (AA) in all coding regions compared to Wuhan-Hu-1 and the presence of those substitutions in Nexstrains Omicron clades

Position	AA in Wuhan-Hu-1	AA in Omicron in Indonesia	No. of Indonesian viruses bearing the AA	Nexstrain's clades posing the AA
ORF1AB				
47	K	R	13	22F
135	S	R	381	All except: 21K, 21L
631	L	F	14	22C
842	T	I	387	All except: 21K, 21L
1221	S	L	86	22D
1307	G	S	386	All except: 21K, 21L
1640	P	S	85	22D
2235	I	L	393	All
2909	A	V	86	22C
3027	L	F	388	All except: 21K, 21L
3090	T	I	387	All except: 21K, 21L
3201	L	F	350	22C; 22D; 22F
3255	T	I	391	All except: 21L
3395	P	H	392	All except: 21L
3667	L	F	17	None
3675	S	DEL	389	All except: 22D, 22F
3676	G	S	76	None
		DEL	314	All except: 22D, 22F
3677	F	L	78	22D; 22F
		DEL	314	All except: 22D, 22F
3829	L	F	9	22E
3833	K	N	392	ALL
4060	N	S	86	22D
4715	P	L	392	ALL
5086	F	Y	22	None
5451	T	N	16	None
5716	R	C	387	All except: 21K, 21L
5967	I	V	385	All except: 21L
6044	A	V	7	None
6107	V	I	14	22D
6471	Q	H	11	None
6564	T	I	386	All except: 21K, 21L
Spike				
19	T	I	379	All except: 21K, 21L
27	A	S	375	All except: 21K, 21L, 22F
64	W	R	11	22C
83	V	A	19	22F
97	K	E	13	22C
142	G	D	381	All except: 21L, 22C
146	H	Q	8	22F
147	K	E	87	22D
152	W	R	87	22D
157	F	L	87	22D
183	Q	E	18	22F
210	I	V	85	22D
213	V	G	356	All except: 21K, 21L
		E	19	22F
250	Y	S/N/H	9	None
254	G	V	13	22F
256	G	S	87	22D
341	G	H	109	22d, 22fF
		D	276	All except: 21L, 22D, 22F
348	R	T	123	All except: 21L, 22C
		K	5	21K
358	K	T	61	None
370	L	I	18	22F
373	S	L	10	21K
		F	368	All except: 21K, 21L
375	S	P	385	All except: 21L
377	S	F	385	All except: 21L
378	T	A	372	All except: 21K, 21L
407	D	N	374	All except: 21K, 21L
410	R	S	361	22A, 22B, 22C, 22E
419	K	N	372	All except: 21L
		T	14	21L
442	N	K	381	All except: 21L
446	K	T/R	8	22B, 22E
447	V	P	20	22F
448	G	S	109	21K, 22D, 22F
454	L	R	24	22A, 22B, 22E
454	L	Q	20	None
		M	9	None
462	N	K	110	22B, 22D, 22E, 22F
479	S	N	385	All except: 21L
480	T	K	383	All except: 21L
486	E	A	384	All except: 21L
489	F	V	24	22A, 22B, 22E
		S	39	22D, 22F
492	F	S	82	22D, 22F
495	Q	R	251	21K, 22C
500	Q	R	385	All except: 21L
503	N	Y	386	All
507	Y	H	385	All except: 21L
616	D	G	386	All
657	H	Y	385	All except: 21L
681	N	K	385	All except: 21L
683	P	H	386	All
706	S	L	16	22C
766	N	K	383	All except: 21L, 22D
798	D	Y	385	All except: 21L
956	Q	H	385	All except: 21L
971	N	K	385	All except: 21L
1253	G	V	7	22C
1266	V	L	6	None
Envelope				
9	T	I	392	All except: 21L
11	T	A	107	22D, 22F
58	V	F	43	None
Matrix				
3	D	N	19	22B, 22E
		G	5	21K
Nucleocapsid				
19	Q	E	387	All except: 21L, 22F
63	A	T	391	All except: 21L
13	P	L	384	All except: 21L
203	R	K	385	All
204	G	R	385	All
322	M	I	13	22C
413	S	R	378	All except: 21K, 21L
ORF3A				
78	H	Y	7	None
140	L	F	86	22C
223	T	I	386	All except: 21K, 21L
262	P	S	6	None
ORF6				
61	D	L	354	22A, 22C, 22D, 22F
ORF7A				
66	A	S	8	None
ORF7B				
15	A	S	5	None
ORF10				
8	A	V	6	None

The amount of sequence data and the time frame of sampling in this study should be conclusive, as we could identified unique markers for Indonesian Omicron variant All complete genomes with high coverage of Omicron variant sequences submitted from Indonesia were included. As of October 27, 2022, the number of complete genome sequences annotated as Omicron detected in Indonesia was approximately 30,000, while the number of complete genomes with high coverage was 506 (approximately 17%). We then excluded sequences with a long tract of more than two Ns as well as sequences that contained singleton insertions/ deletions. The total number of Indonesian Omicron sequences in the dataset was 385 (76%). Although the number of COVID-19 tests in Indonesia is relatively low (408,000 tests per 1 M people, compared to Australia with 3 M per 1 M people, or Vietnam 867,000 per 1 M people https://www.worldometers.info/coronavirus/), the number of samples in our study should represent the Omicron population in Indonesia. The time frame seems also appropriate. Due to the global spread and tranmission of SARS-COV-2, most recent data should show that no more unique Indonesian markers could be identified.

Amazingly, the overall whole genome diversity of the Omicron variant detected in Indonesia was 10 times higher than that previously reported, while the Tajima D value was almost equal. We calculated the diversity, and the Tajima statistic of the full genome of Indonesian Omicron was 0.0008 and -2.669. At the beginning of the emergence of the Omicron variant, the Tajima D statistic was applied, and the nucleotide diversity was 0.00008 and the Tajima D value was -2.709 (Yeh and Contreras 2021). The time of sampling contributed to the higher nucleotide diversity in our dataset. An almost equal Tajima D of negative value signifies an excess of low frequency polymorphisms relative to expectation, indicating a population size increase (e.g., after a bottleneck or a selective sweep) (Vásquez-Aguilar et al. 2021).

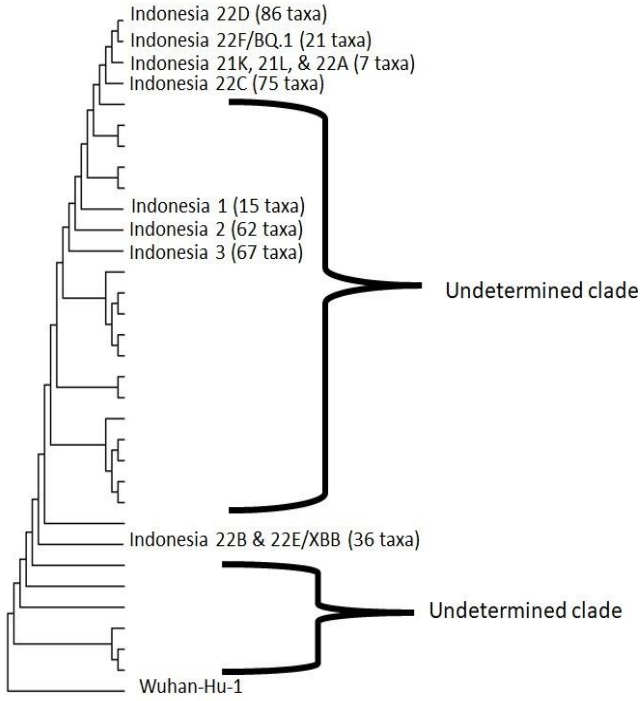

Figure 1. The topology of phylogeny of whole genome of Omicron variants in Indonesia uploaded in GISAID. Data with complete genomes and high coverage were downloaded on October 27, 2022. Two representative sequences of Nextstrain Omicron clades (21K, 21 L, 22A-F, including XBB and BQ.1 lineages) as well as original SARS-CoV-2 of Wuhan-Hu-1 strains were included. Sequences with a long tract of more than two Ns as well as sequences that make singleton insertion/deletion are excluded. Branches with Nextstrain clades as well as with more than 10 taxa collapsed. Numbers of taxa are shown

Table 1. Tajima's neutrality test of the whole genome and all coding regions of Omicron variant sequences detected in Indonesia

Gene/ORF	n	m	S	p_s	Θ	π	D
Whole genome	29887	385	1096	0.0367	0.0056	0.0008	-2.669
ORF1AB	21291	385	667	0.0313	0.0050	0.0005	-2.768
Spike	3828	378	180	0.0470	0.0072	0.0020	-2.174
Envelope	228	394	7	0.0307	0.0047	0.0030	-0.749
Matrix	669	384	29	0.0434	0.0066	0.0008	-2.336
Nucleocapsid	1260	386	81	0.0643	0.0098	0.0007	-2.701
ORF3A	828	385	60	0.0725	0.0111	0.0016	-2.442
ORF6	186	385	13	0.0699	0.0107	0.0036	-1.529
ORF7A	366	385	68	0.1858	0.0285	0.0018	-2.710
ORF7B	132	385	19	0.1439	0.0221	0.0011	-2.373
ORF8	365	385	79	0.2164	0.0332	0.0015	-2.779
ORF10	119	384	6	0.0504	0.0077	0.0005	-1.749

Note: Calculated following Tajima (Tajima 1989) and Tajima Statistics (Nei and Kumar 2000) in Mega 11 (Tamura et al. 2021), m: Number of sequences, n: Total number of sites, S: Number of segregating sites, ps = S/n, Θ = ps/a1, π = Nucleotide diversity, and D is the Tajima test statistic

Table 2. Number of Omicron variants SARS-CoV-2 detected in Indonesia posing substituted amino acids compared to Wuhan-Hu-1 and the presence of those substitutions in Nextstrain Omicron clades

Coding Region-Position	AA* in Wuhan-Hu-1	AA in Omicron in Indonesia	Number of Indonesian viruses bearing the AA	Nextstrain's clades posing the AA
ORF1AB-3667	L	F	17	None
ORF1AB-3676	G	S	76	None
ORF1AB-5086	F	Y	22	None
ORF1AB-5451	T	N	16	None
ORF1AB-6044	A	V	7	None
ORF1AB-6471	Q	H	11	None
Spike-250	Y	S/N/H	9	None
Spike-358	K	T	61	None
Spike-454	L	Q/M	29	None
Spike-1266	V	L	6	None
Envelope-58	V	F	43	None
ORF3A-78	H	Y	7	None
ORF3A-262	P	S	6	None
ORF7A-66	A	S	8	None
ORF7B-15	A	S	5	None
ORF10-8	A	V	6	None

Note: AA: amino acid

Interestingly, the three highest coding region diversities are the ORF6, Envelope, and Spike coding regions, while the three highest Tajima values are ORF6, Envelope and ORF10. Except for the Spike coding region, which has high diversity but lower Tajima values, the ORF6, Envelope and ORF10 coding regions seem to mutate close to as expected. In contrast, the spike coding region belongs to the most divergent genome gene fragment, which could be even more diverse. Located at the outer part of the virion, more changes can be accommodated without disturbing virus integrity. It is generally believed that the spike protein is a major pathogenic coronavirus determinant (Millet and Whittaker 2015), which is highly glycosylated (Tian et al. 2021) and carries protease activation sites (Masters 2006) and various antibody binding sites (Zhang et al. 2020; Harvey et al. 2021; He et al. 2021).

A negative Tajima's index can be interpreted to indicate that the diversity is lower than expected, indicating population expansion after a bottleneck. Travel restriction could be assumed to generate a bottleneck. However, the lower-than-expected diversity can contribute to virus integrity. Mutations that cause amino acid substitutions are restricted to the function of certain genes and might be lethal.

To draw a rapid overview and estimate of the whole genome of Indonesian Omicron variants, a phylogeny was constructed with the original SARS-CoV-2 strain of Wuhan-Hu-1 and representatives of eight Nextstrain Omicron clades. The phylogeny shows that some Indonesian Omicron strains are clustered with all Nextstrain clades. However, there are sequences of Indonesian Omicron variants that do not share branches with any clade. As quantitative statistics were not conducted, the phylogeny could not explain the ancestor-descendant relationship between branches. Nonetheless, it is plausible to state that all Nextstrain clade descendants have been detected in Indonesia, which then mutated to be probably unique Indonesian strains, as explained in other findings below.

As expected, the reintroduction of SARS-CoV-2 strains despite strict travel restriction and border closure caused all Omicron clades to be identified in Indonesia, while unique strains probably emerged due to local adaptation. The markers of global introduction and local substitutions could be identified at the amino acid level. The signatures of global Omicron clades occurred in all coding regions except ORF8. Meanwhile, unique amino acid substitutions to Wuhan-Hu-1 of the Omicron variants detected in Indonesia occurred in many coding regions, six sites occurred in ORF1AB and four in Spike. No Indonesian Omicron signature was identified in the Matrix, Nucleocapsid, ORF6, and ORF8 coding regions.

Additionally, we identified the pattern of global Omicron variants in Indonesia, namely, deletion in Nucleocapsid, 3'-untranslated region (3'-UTR), and truncation of ORF8, probable indirect evidence of decreased virulence of this variant (Suharsono et al. 2023). Deletion in nucleocapsid (nucleotide no 28346-28354 in our dataset) occurred in 375 of 385 Indonesian Omicron variants, while deletion of the 3'-UTR occurred in 11 of 45 valid sequences (not shown). In the 3'-UTR analysis, we excluded sequences with a tract of Ns at the deletion site. Additionally, truncation of ORF8 at position 8 or 18 of the Indonesian Omicron variants occurred in 15 sequences.

For further research, the same analysis should be conducted in each country to draw a global picture of the effect of border closure at the molecular level as a lesson for future pandemic events. Unique markers of any country could be identified for a certain time frame prior global spread. The risk of porous borders even with rigid closure protocols should be communicated appropriately. In our study, strains that bear many national markers should be identified to be used to predict the introduction of any strain from Indonesia. Moreover, those who have access to the patient status of Omicron infections as well as those of other variants should present the data in international scientific platforms, so we will have strong indirect evidence

of their pathogenicity without conducting laboratory experiments, which bring a risk of unintentional leaks.

Data availability

The sequence data of complete and high coverage Omicron variants of SARS-CoV-2 from Indonesia are at the GISAID with identifier EPI_SET_230109nt, DOI: 10.55876/gis8.230109nt. The sequence data of 15 to 20 representatives of Nexstrain clades are accessible at GISAID with identifier EPI_SET ID: EPI_SET_230327ca and DOI: 10.55876/gis8.230327ca. All genome sequences and associated metadata in this dataset are published in GISAID's EpiCov database. To view the contributors and each individual sequence with details such as accession number, virus name, collection date, originating lab and submitting lab and the list of authors, visit 10.55876/gis8.230109nt and 10.55876/gis8.230327ca. All polymorphic amino acids of all proteins of Indonesian and Nextstrain Omicron clades are listed in Table S1.

We conclude that the overall whole genome diversity of the Omicron variants detected in Indonesia is 0.0008, with the three highest coding region diversities being the ORF6, Envelope, and Spike protein. The Tajima D value of the whole genome is -2.669, while the three highest Tajima values are ORF6, Envelope and ORF10. The phylogeny shows that there are 168 sequences of Indonesian Omicron variants, which do not share branches with Nextstrain Omicron clades. Most likely, unique amino acid substitutions to Wuhan-Hu-1 of the Omicron variants detected in Indonesia occurred in many coding regions, six sites occurred in ORF1AB and four in Spike.

ACKNOWLEDGEMENTS

This study was supported by Research, Technology and Higher Education (RISTEKDIKTI) of Indonesia through World Class Research. We gratefully acknowledge all data contributors, i.e., the authors and their originating laboratories responsible for obtaining the specimens and their submitting laboratories for generating the genetic sequence and metadata and sharing via the GISAID Initiative, on which this research is based. The English language of the manuscript has been edited by Springer Nature Author Services. Author Contributions Ida B.K. Suardana and Gusti N. Mahardika contributed to the conception. Ida B.K. Suardana, Bayu K. Mahardika, Adhiyoga A.A.N.A. Santosa contributed to the acquisition, analysis, and interpretation of data; Ida B.K. Suardana and Gusti N. Mahardika drafted the manuscript. All authors reviewed the manuscript and approved the version to be published.

REFERENCES

Aboura S. 2023. The role of climate on Covid-19 spread in France. Intl J Environ Health Res 33: 809-822. DOI: 10.1080/09603123.2022.2055747.

Aiewsakun P, Nilplub P, Wongtrakoongate P, Hongeng S, Thitithanyanont A. 2021. SARS-CoV-2 genetic variations associated with COVID-19 pathogenicity. Microb Genom 7: 000734. DOI: 10.1099/mgen.0.000734.

Ansori ANM, Kharishma VD, Muttaqin SS, Antonius Y, Parikesit AA. 2020. Genetic variant of SARS-CoV-2 isolates in Indonesia: Spike Glycoprotein gene. J Pure Appl Microbiol 14 (Suppl 1): 971-978. DOI: 10.22207/jpam.14.spl1.35.

Arora P, Mrig S, Goldust Y, Kroumpouzos G, Karadağ AS, Rudnicka L, Galadari H, Szepietowski JC, Di Lernia V, Goren A, Kassir M, Goldust M. 2021. New Coronavirus (SARS-CoV-2) crossing borders beyond cities, nations, and continents: Impact of international travel. Balkan Med J 38 (4): 205-211. DOI: 10.5152/balkanmedj.2021.21074.

Bhatt S, Katzourakis A, Pybus OG. 2010. Detecting natural selection in RNA virus populations using sequence summary statistics. Infect Genet Evol 10 (3): 421-430. DOI: 10.1016/j.meegid.2009.06.001.

Boekel L, Besten YR, Hooijberg F et al. 2022. SARS-CoV-2 breakthrough infections in patients with immune-mediated inflammatory diseases during the omicron dominant period. Lancet Rheumatol 4 (11): e747-e750. DOI: 10.1016/S2665-9913(22)00221-1.

Chen S, Prettner K, Kuhn M, Geldsetzer P, Wang C, Bärnighausen T, Bloom DE. 2021. Climate and the spread of COVID-19. Sci Rep 11 (1): 9042. DOI: 10.1038/s41598-021-87692-z.

Elbe S, Buckland-Merrett G. 2017. Data, disease and diplomacy: GISAID's innovative contribution to global health. Glob Chall 1 (1): 33-46. DOI: 10.1002/gch2.1018.

Grépin KA, Ho T-L, Liu Z, Marion S, Piper J, Worsnop CZ, Lee K. 2021. Evidence of the effectiveness of travel-related measures during the early phase of the COVID-19 pandemic: A rapid systematic review. BMJ Glob Health 6 (3): e004537. DOI: 10.1136/bmjgh-2020-004537.

Hahn G, Wu CM, Lee S et al. 2021. Genome-wide association analysis of COVID-19 mortality risk in SARS-CoV-2 genomes identifies mutation in the SARS-CoV-2 spike protein that colocalizes with P.1 of the Brazilian strain. Genet Epidemiol 45 (7): 685-693. DOI: 10.1002/gepi.22421.

Harvey WT, Carabelli AM, Jackson B, Gupta RK, Thomson EC, Harrison EM, Ludden C, Reeve R, Rambaut A, Consortium COG-UK, Peacock SJ, Robertson DL. 2021. SARS-CoV-2 variants, spike mutations and immune escape. Nat Rev Microbiol 19 (7): 409-424. DOI: 10.1038/s41579-021-00573-0.

He J, Huang F, Zhang J, Chen Q, Zheng Z, Zhou Q, Chen D, Li J, Chen J. 2021. Vaccine design based on 16 epitopes of SARS-CoV-2 spike protein. J Med Virol 93 (4): 2115-2131. DOI: 10.1002/jmv.26596.

Johnson AG, Amin AB, Ali AR et al. 2022. COVID-19 incidence and death rates among unvaccinated and fully vaccinated adults with and without booster doses during periods of Delta and Omicron Variant emergence - 25 U.S. Jurisdictions, April 4-December 25, 2021. MMWR Morb Mortal Wkly Rep 71 (4): 132-138. DOI: 10.15585/mmwr.mm7104e2.

Kared H, Wolf A-S, Alirezaylavasani A, Ravussin A, Solum G, Tran TT, Lund-Johansen F, Vaage JT, Nissen-Meyer LS, Nygaard UC, Hungnes O, Robertson AH, Næss LM, Trogstad L, Magnus P, Munthe LA, Mjaaland S. 2022. Immune responses in Omicron SARS-CoV-2 breakthrough infection in vaccinated adults. Nat Commun 13 (1): 4165. DOI: 10.1038/s41467-022-31888-y.

Khare S, Gurry C, Freitas L, Schultz MB, Bach G, Diallo A, Akite N, Ho J, Lee RT, Yeo W, Curation Team GC, Maurer-Stroh S. 2021. GISAID's role in pandemic response. China CDC Wkly 3 (49): 1049-1051. DOI: 10.46234/ccdcw2021.255.

Kim K, Omori R, Ito K. 2017. Inferring epidemiological dynamics of infectious diseases using Tajima's D statistic on nucleotide sequences of pathogens. Epidemics 21: 21-29. DOI: 10.1016/j.epidem.2017.04.004.

Mahardika GN, Mahendra NB, Mahardika BK, Suardana IBK, Pharmawati M. 2022. Annotating spike protein polymorphic amino acids of variants of SARS-CoV-2, including Omicron. Biochem Res Intl 2022: 2164749. DOI: 10.1155/2022/2164749.

Masters PS. 2006. The molecular biology of coronaviruses. Adv Virus Res 66: 193-292. DOI: 10.1016/S0065-3527(06)66005-3.

Meo SA, Meo AS, Al-Jassir FF, Klonoff DC. 2021. Omicron SARS-CoV-2 new variant: Global prevalence and biological and clinical characteristics. Eur Rev Med Pharmacol Sci 25 (24): 8012-8018. DOI: 10.26355/eurrev_202112_27652.

Millet JK, Whittaker GR. 2015. Host cell proteases: Critical determinants of coronavirus tropism and pathogenesis. Virus Res 202: 120-134. DOI: 10.1016/j.virusres.2014.11.021.

Nei M, Kumar S. 2000. Molecular Evolution and Phylogenetics. Oxford University Press, New York.

Omori R, Wu J. 2017. Tajima's D and site-specific nucleotide frequency in a population during an infectious disease outbreak. SIAM J Appl Math 77 (6): 2156-2171. DOI: 10.1137/17M1114946.

Satapathy P, Kumar P, Mehta V, Suresh V, Khare A, Rustagi S, Daulati MN, Neyazi M, Najafi E, Neyazi A. 2024. Global spread of COVID-19's JN.1 variant: Implications and public health responses. New Microbes New Infect 57: 101225. DOI: 10.1016/j.nmni.2024.101225.

Shiino T, Okabe N, Yasui Y, Sunagawa T, Ujike M, Obuchi M, Kishida N, Xu H, Takashita E, Anraku A, Ito R, Doi T, Ejima M, Sugawara H, Horikawa H, Yamazaki S, Kato Y, Oguchi A, Fujita N, Odagiri T, Tashiro M, Watanabe H. 2010. Molecular evolutionary analysis of the influenza A(H1N1)pdm, May-September, 2009: Temporal and spatial spreading profile of the viruses in Japan. PLoS One 5 (6): e11057. DOI: 10.1371/journal.pone.0011057.

Shu Y, McCauley J, 2017. GISAID: Global initiative on sharing all influenza data - from vision to reality. Euro Surveill 22 (13): 30494. DOI: 10.2807/1560-7917.ES.2017.22.13.30494.

Siwiak M, Szczesny P, Siwiak M. 2020. From the index case to global spread: The global mobility based modelling of the COVID-19 pandemic implies higher infection rate and lower detection ratio than current estimates. PeerJ 8: e9548. DOI: 10.7717/peerj.9548.

Suharsono H, Mahardika BK, Sudipa PH, Sari TK, Suardana IBK, Mahardika GN. 2023. Consensus insertion/deletions and amino acid variations of all coding and noncoding regions of the SARS-CoV-2 Omicron clades, including the XBB and BQ.1 lineages. Arch Virol 168 (6): 156. DOI: 10.1007/s00705-023-05787-6.

Tajima F. 1989. Statistical methods to test for nucleotide mutation hypothesis by DNA polymorphism. Genetics 123 (3): 585-595. DOI: 10.1093/genetics/123.3.585.

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

Tian W, Li D, Zhang N, Bai G, Yuan K, Xiao H, Gao F, Chen Y, Wong CCL, Gao GF. 2021. O-glycosylation pattern of the SARS-CoV-2 spike protein reveals an "O-Follow-N" rule. Cell Res 31 (10): 1123-1125. DOI: 10.1038/s41422-021-00545-2.

Vásquez-Aguilar AA, Barbachano-Guerrero A, Angulo DF, Jarquín-Díaz VH. 2021. Phylogeography and population differentiation in *Hepatozoon canis* (Apicomplexa: Hepatozoidae) reveal expansion and gene flow in world populations. Parasit Vectors 14 (1): 467. DOI: 10.1186/s13071-021-04924-x.

Yadav PD, Nyayanit DA, Majumdar T et al. 2021. An epidemiological analysis of SARS-CoV-2 genomic sequences from different regions of India. Viruses 13 (5): 925. DOI: 10.3390/v13050925.

Yeh T-Y, Contreras GP. 2021. Tajima D test accurately forecasts Omicron medRxiv. DOI: 10.1101/2021.12.02.21267185.

Zhang B-Z, Hu Y-F, Chen L-L, Yau T, Tong Y-G, Hu J-C, Cai J-P, Chan K-H, Dou Y, Deng J, Wang X-L, Hung IF-N, To KK-W, Yuen KY, Huang J-D. 2020. Mining of epitopes on spike protein of SARS-CoV-2 from COVID-19 patients. Cell Res 30 (8): 702-704. DOI: 10.1038/s41422-020-0366-x.