ORIGINAL ARTICLE

IMPACT OF RAPID VACCINATION ON SARS-COV-2 GENOMIC DIVERSITY: AN INTERVENTION TO MINIMISE THE PUBLIC BURDEN OF THE PANDEMIC

Amirah Azzeri[1], Shuhaila Mat-Sharani[2], Danish A/L Kumareahsan[3], Ismatul Nurul Asyikin Ismail[4], Muhamad Arif Mohamad Jamali[4] and Liyana Azmi[1]

[1]Faculty of Medicine and Health Sciences, Universiti Sains Islam Malaysia, Nilai, Negeri Sembilan, Malaysia
[2]Biomedicine program, School of Health Sciences, Universiti Sains Malaysia, 16150 Kubang Kerian, Kelantan, Malaysia
[3] Universiti Kebangsaan Malaysia, 43600, Universiti Kebangsaan Malaysia, Bangi, Selangor Malaysia
[4]Faculty of Science and Technology, Universiti Sains Islam Malaysia, Bandar Baru Nilai, 71800 Nilai, Negeri Sembilan, Malaysia

***Corresponding author: Liyana Azmi**
Email: liyana.azmi@usim.edu.my

ABSTRACT

Global sharing of SARS-CoV-2 sequences enabled comprehensive analyses of COVID-19 genomic diversity and demographics. Yet, regional genomic surveillance is often neglected, leading to the possible oversight of novel mutations by public health authorities. Our study used the Global Initiative on Sharing Avian Influenza Data (GISAID) database to analyse infection patterns in the state of Negeri Sembilan, and compare infection patterns to the state of Selangor, in Malaysia. We discuss the impact of rapid vaccination on resulting single nucleotide variants (SNVs) and identified novel sporadic mutations may affect viral fitness and pathogenicity. Four hundred and seventeen SARS-CoV-2 sequences extracted from Negeri Sembilan from July 2021 until June 2022. Infection patterns based on pangolin lineages from Negeri Sembilan was compared to infections of the same period from Selangor. SNVs from the spike protein were sorted by frequency, with the lowest frequency variant submitted for functional prediction using PredictSNP. Negeri Sembilan exhibited a comparable infection pattern to Selangor, but with fewer Omicron sequences which was postulated to occur due to the rapid vaccination strategies by Negeri Sembilan. Omicron outbreaks were associated with eased lockdowns and policy changes in December 2021. From our extracted data, seventy novel SNVs in the spike protein of SARS-CoV-2 were extracted from this study. In silico predictions indicated five of the SNVs (S221L, L226S, V826L, C1240F and C1243F) to may cause functional defects to the spike protein. Rapid sequencing and analysis will aid policymaking for public health controls by detecting potential outbreaks within transient variants.

Keywords: *SARS-CoV-2, Malaysia variant, genetic diversity, spike protein, pangolin replacement*

INTRODUCTION

In December 2019, a novel SARS-CoV-2 was discovered and identified as the coronavirus disease 2019 (COVID-19) pathogen(1). By 5th December 2022, 641 million cases worldwide were recorded, with 6 million deaths reported (2). In Malaysia, the first wave of COVID-19 infection was the first of the country's total four waves of infection (3). Mass gatherings initiated subsequent outbreaks and were further strained by burdened healthcare systems (4). Fortunately, with international collaborations and efforts, vaccinations have been administered worldwide.

As of 5th January 2023, up to 13.18 billion doses were administered (2). In addition, lockdowns and movement restrictions between districts and states have proven effective in containing and controlling the spread of COVID-19 cases (5,6). On top of regional conditions, Malaysia has initiated the SARS-CoV-2 genomic surveillance project to enable the identification of SARS-CoV-2 variants in Malaysia (7). The extracted SARS-CoV-2 sequence from the patients would then be deposited in the Global Initiative on Sharing Avian Influenza Data (GISAID) to enable real-time analysis and surveillance of circulating variants.

Monitoring genomic diversity using advanced sequencing techniques allows public health experts to track the evolution of the virus and its variants over time (8). This surveillance is critical to determine the effectiveness of control measures, develop targeted interventions and establish vaccination strategies. In addition, understanding the genomic landscape of COVID-19 enables the identification of potential drug targets and provides information for therapy development.

The intricate relationship between genomic diversity and mutation, on the other hand, underscores the importance of ongoing research and adaptation of public health interventions. Regular genomic surveillance, combined with comprehensive epidemiological data collection, improves our ability to anticipate and respond to

changes in viral behaviour and helps reduce the impact of the pandemic on global health (9).

Thus, genomic surveillance (10) is crucial to highlight new circulating variants and predicting infection patterns. However, sequence analysis at regional levels needs to be improved. Additionally, sequencing capabilities differ across states and depend highly on the financial support and workforce able to perform sequencing analyses.

This study compared COVID-19 infection profiles in Selangor and Negeri Sembilan. Selangor had the highest infections in Malaysia (11), while Negeri Sembilan had high vaccination rates (12). We examined demographic differences and investigated new mutations, particularly in the spike protein, which may affect viral fitness and immune system evasion. Our findings aim to predict future infection patterns for better infection control.

METHODS

Sampling

Our study utilised 428 SARS-CoV-2 extracted in Negeri Sembilan from July 2021 until June 2022. Of the 428 sequences, 30 were omitted due to mislabelling of the originating labs which was not located in Negeri Sembilan. Eight sequences were omitted since the nucleotide sequences were < 29,000 nt. A final total of 417 sequences was extracted for analysis. The reference genome of SARS-CoV-2 used in this study is hCoV-19/Wuhan/WIV04/2019 (WIV04). Gene annotation of 417 sequences was the first step of the analysis, which was done using Viral Genome ORF Reader (VIGOR) version 4.1.2 (13) and SNVs was extracted from the GISAID database.

Phylogenomic analysis

Multiple Sequence Alignment (MSA) is a multiple alignment program for amino acid, or nucleotide sequences was performed using MAFFT with FFT-NS-2 methods (14). Once the alignment was completed, the phylogenetic tree was constructed. In this work, MEGA X software (15) is used to construct the phylogenetic tree. The phylogenetic tree was built using the UPGMA method with a Maximum Composite Likelihood (MCL) model, uniform site rates and a 1000-time bootstrap value. After MEGA X, the Interactive Tree of Life (iTOL) (16) was used to display and manage the phylogenetic tree for more accessible and interactive visualisation.

Frequency of mutations and selection for PredictSNP

The single nucleotide polymorphisms for all samples were extracted from the sequences and compared to the GISAID website. The *in-silico* prediction program PredictSNP (17). was used to infer the biological effects of all the samples. PredictSNP is a consensus program which comprises the predictions from APP, PhD-SNP, Polyphen-1, Polyphen-2, SIFT, SNAP, nsSNPAnalyzer, and PANTHER to predict the biological effect of a single nucleotide variation. PredictSNP combines the six best performance tools previously mentioned to improve prediction. The prediction outputs include predictions from the top six prediction tools and outcomes of the mutation residue as “Neutral” or “Deleterious”. It also provides the percentage indicating the expected accuracy of the prediction.

In silico prediction

The functional effects of SNVs were subjected to *in silico* prediction using PredictSNP. PredictSNP combines eight established predictive tools to predict the biological effect of a single nucleotide variant. The output for PredictSNP is confidence scores, with 1 being the highest score. Each submission will report the changes as ’neutral ‘(no biological change) or ’deleterious ‘(possibly causing a functional change in the protein).

RESULTS

The sequence deposited for GISAID from Negeri Sembilan according to districts, were as follows: Seremban (47%, n = 195), Tampin (35%, n = 146), Kuala Pilah (7%, n = 31), Jempol (6%, n = 24), Jelebu (2%, n = 7), Rembau (2%, n = 7), Port Dickson (1%, n = 6) (Figure 1). Notably, the data submitted are less likely to reflect the population densities of the districts and are more dependent on the resources for sequencing and sampling by the health facility. The highest number of sequences submitted was from Seremban and was accrued from Hospital Tunku Jaafar (35%, n = 146) and Health District Office (HDO) Seremban (11%, n = 49). Tampin, the fourth highest population in Seremban, contributed to the second highest sampling of this study and only accrued samples from HDO Tampin. A geometric information of Malaysia in the form of shapefile was acquired from a freely available source (18) through https://earthworks.stanford.edu/catalog/stanford-qg469kj1734.

Phylogenetic analysis of the 417 sequences from Negeri Sembilan genomes revealed 20 lineages of which 9 (BA.1, BA.1.1, BA.1.1.18, BA.2, BA.2.10, BA.23, BA.2.40.1, BA.2.57 and BA.2.9) are still circulating (Figure 2). Three major GISAID clades were identified, namely GRA, GK and O, on a maximum likelihood phylogenetic tree (Figure 2). The two major variants show that the highest number of sequenced genomes were Delta (B.1.617.2) and Omicron (BA.1.1, BA.2.40.1, BA.2.57 and BA.2.9) from Seremban and Tampin. The evolution of variants can clearly be seen from the O clade, followed by the GK clade, which represents Delta variants. Finally, the GRA

clade which represents Omicron variants – which continues to circulate until today.

The lineage of SARS-CoV-2 was assigned based on the GISAID clade and Pango lineage assignments. From July 2021 until June 2022, the top circulating variant in Negeri Sembilan is Delta (59.9%, n = 250), followed by Omicron (38.8%, n = 162). Delta lineages AY.79 (30%, n = 128) and AY.59 (23%, n = 96) dominated from July 2021 to January 2022. In January 2022, there was change of circulating variants from Delta to Omicron. In January 2022, BA.1.1 (25%, n = 103) was the dominating lineage, which was then replaced by BA.2 (8%, n = 34) in February 2022. The following months recorded other circulating lineages including BA.2.10 and BA.23. The number of cases has reduced significantly from April 2022 onwards.

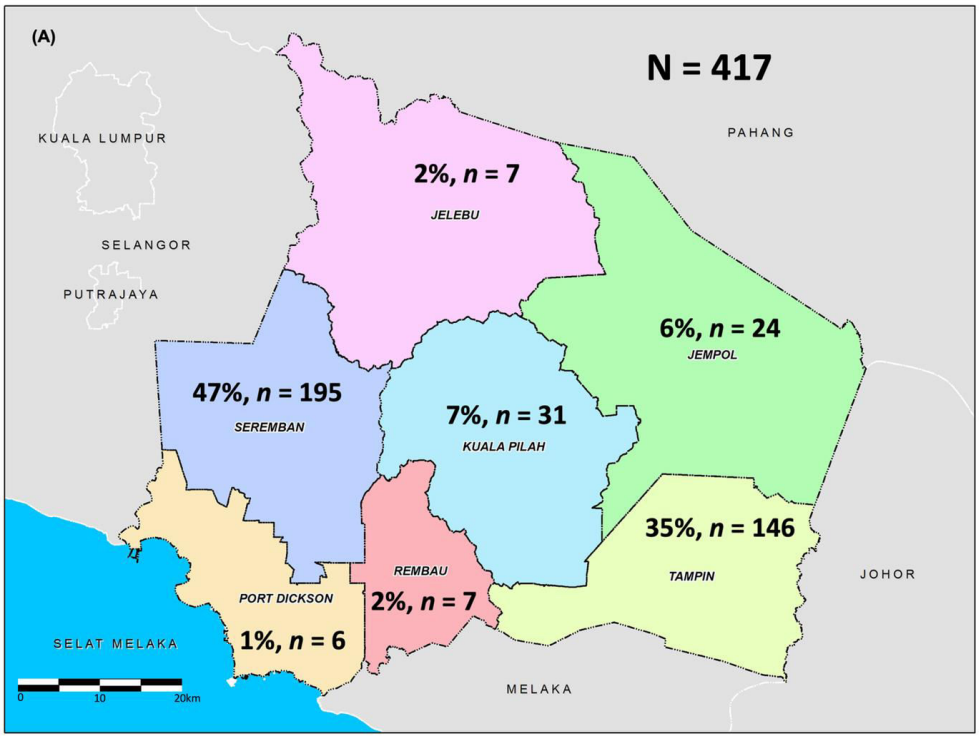

Figure 1: Regional distribution of samples submitted for sequencing and deposition in GISAID database

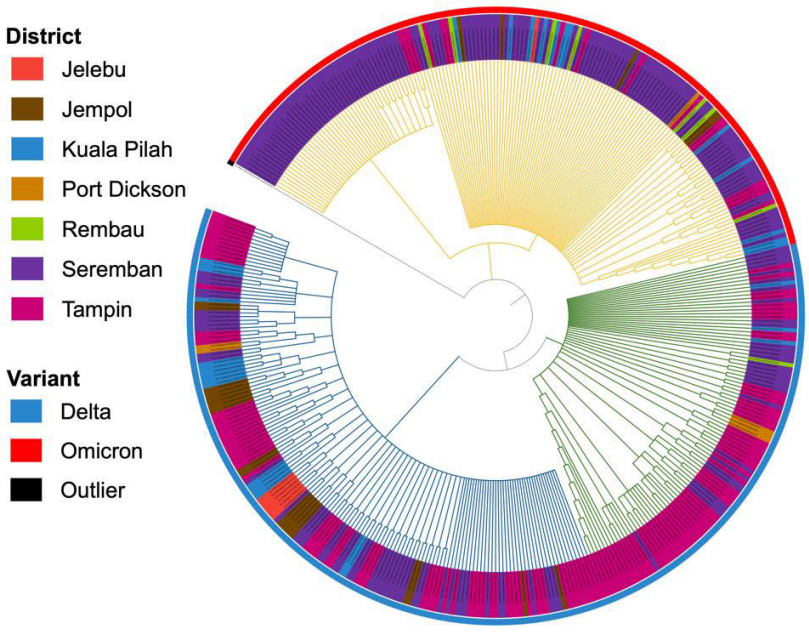

Figure 2: A maximum likelihood phylogenetic tree of sequenced samples in Negeri Sembilan from July 2021 – May 2022. The districts and the GISAID clades GRA (yellow), GK (green) and O (others) are as indicated.

We compared the circulating lineages to the neighboring state Selangor, which was the state with the highest number of COVID-19 cases from July 2021 until June 2022. Selangor deposited a total of 5287 SARS-CoV-2 sequences, which was 12.6-fold higher than the number of sequences submitted by Negeri Sembilan. To enable a direct comparison, we only compared the lineages detected in both Negeri Sembilan and Selangor. Compared to Negeri Sembilan, Selangor recorded a much higher dominance for Omicron variants (85.5%, n = 4192), with the leading lineage of BA.2 (41.3%, n = 2027) (Figure 3). Both states reported a surge of BA.2 cases to occur in March 2022 and saw the waning of total Omicron sequenced genomes towards May 2022 Selangor also had a much diverse range of circulating lineages compared to Negeri Sembilan. On top of the linages circulating in both states, Selangor recorded an added 69 circulating lineages (data not shown), which accounted for 7% of total Selangor sequences (n = 387), The additional 69 circulating lineages included both Delta and Omicron variants.

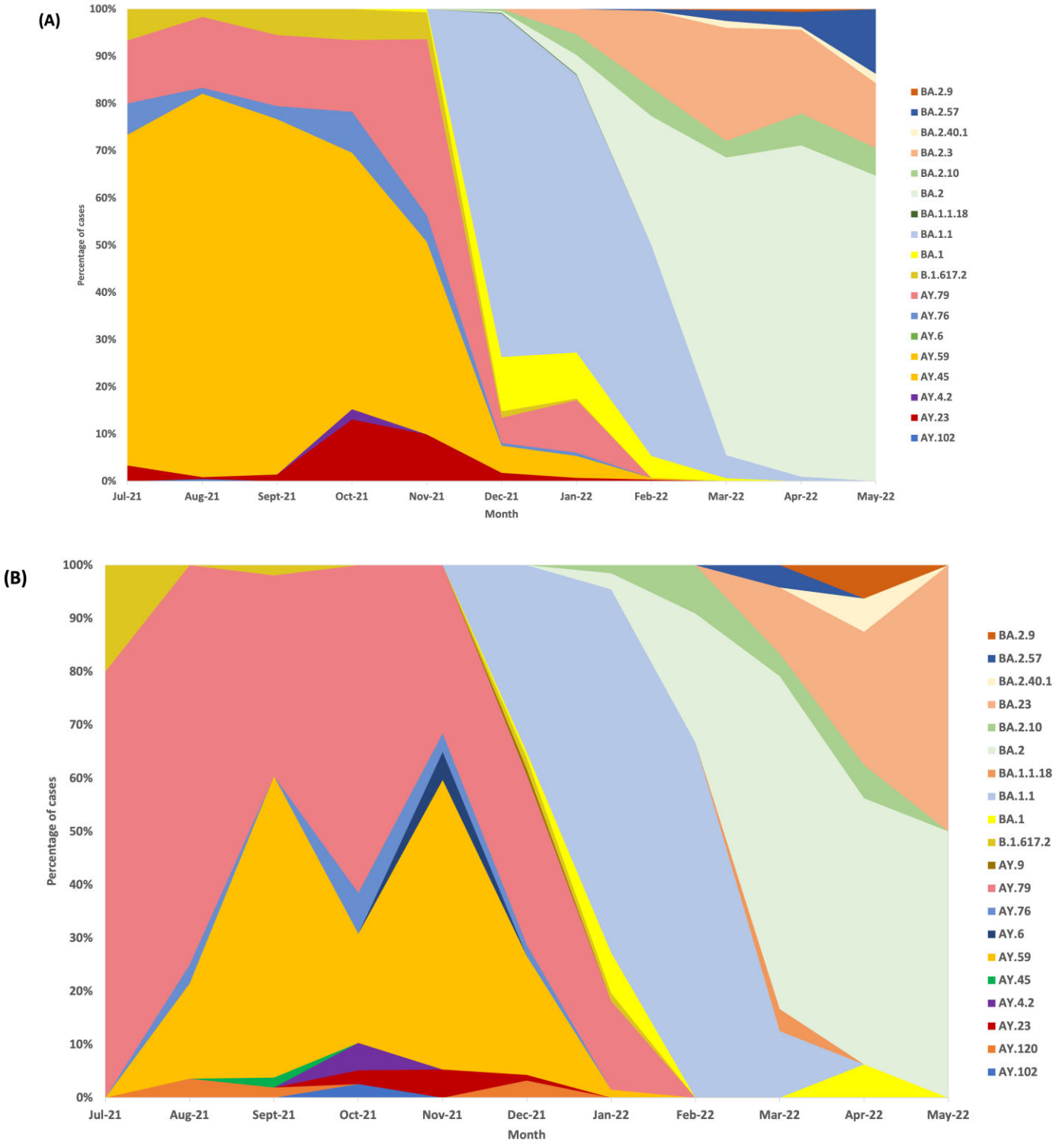

Figure 3: SARS-CoV-2 variants from July 2021 – May 2022 in (A) Negeri Sembilan (n = 417), (B) and Selangor (n = 5287). Circulating SARS-CoV-2 variants in Negeri Sembilan by month. The 'Others' variants include AY.102, AY.42, AY.45, AY.9, BA.2.40.1, BA.2.57 and BA.2.9.

Reports of immune evasion and diagnostic escape of SARS-CoV-2 following the increasing mutations particularly within the S protein of the virus was notable. Based on these findings, we focused on the mutations within the S protein (19). From our utilised dataset, ninety-nine-point five percent (414/417) of the sequences showed the D614G mutation – distinctive for the increased infectivity phenotype in the Delta variant. Other mutations with high frequencies (>50% of the samples), which are Y505H (57%), E156G (57%), F157del (57%), R158del (57%), T19R (59%), D950N (61%), L452R (61%), G142D (73%) and T478K (98%) mutations (Table 1). Our dataset also showed subvariants carrying rare mutations, including C1240F, D1084A, D574Y, E661D, G1251R, I1210V, I693V, L216del, L226S, N925S, P26L, Q613H, R214L, S256L, S704L, T1117I, V62I, V826L (all at 0.24%), V915I, A93S, C1243F, E1262D, L18F, N81D, V1176F, V171I, S221L (all at 0.48%), G446V, Q787K, and T573I at 1.2%, respectively.

To understand the biological effects of these mutations, the SNVs were subjected to phenotypical predictions using PredictSNP (17). Based on the computational analysis by PredictSNP, 11 out of 417 mutations were predicted to be 'deleterious' (Table 2).

Table 1: Frequencies of isolated spike mutations from patients in Negeri Sembilan from July 2021 until June 2022

Isolated spike mutations from sequenced samples	Frequency (N, %)
D614G	414/417, 99.5
T478K	401/417, 96.4
G142D	295/417, 71
L452R, D950N	249/417, 59.9
T19R	243/417, 58.4
E156G	232/417, 55.7
N501Y, P681H, H655Y, E484A	162/417,39
C1240F, D1084A, D574Y, E661D, G1251R, I1210V, I693V, L216del, S221L, L226S, N925S, P26L, Q613H, R214L, S256L, S704L, T1117I, V62I, V826L	1/417, 0.24
V915I, A93S, C1243F, E1262D, L18F, N81D, V1176F, V171I, S221L	2/417, 0.48
G446V, Q787K, and T573I	5/417, 1.2

Six mutations (T95I, N856K, N764K, N969K, Y505H, and L452R) have previously been characterised and were noted to be natural SNVs in variants of Omicron, Iota, and Mu. The other mutations were uncharacterised and showed the lowest frequency with harmful effects, including C1240F, V826L, L226S, C1243F and S221L. C1240F, V826L and L226S. The SNVs mutations only occurred once, while C1243F and S221L mutations occurred twice. We mapped the mutations on the spike protein of SARS-CoV-2 and determined the location for the harmful mutation as follows: L226S and S221L (N-terminal domain), V826L, C1240F and C1243 (S2 domain), C1240F and C1243F (cytoplasm domain).

Table 2a: List of PredictSNP predictions in Negeri Sembilan from July 2021 – May 2022

Wild residue	Position	Mutated residue	Mutation frequency (n, %)	PredictSNP prediction	PredictSNP accuracy score
C	1240	F	1, 0.24	Deleterious	0.86908365
V	826	L	1, 0.24	Deleterious	0.7556615
L	226	S	1, 0.24	Deleterious	0.52145411
C	1243	F	2, 0.48	Deleterious	0.7556615
S	221	L	2, 0.48	Deleterious	0.50595948
T	95	I	98, 23.56	Deleterious	0.60697259
N	856	K	106, 25.48	Deleterious	0.7556615
N	764	K	151, 36.30	Deleterious	0.71871275
N	969	K	156, 37.50	Deleterious	0.54946365
Y	505	H	162, 38.94	Deleterious	0.60697259
L	452	R	249, 59.86	Deleterious	0.65494636
Q	613	H	1, 0.24	Neutral	0.62831343
S	704	L	1, 0.24	Neutral	0.73834499
P	26	L	1, 0.24	Neutral	0.82622462
D	574	Y	1, 0.24	Neutral	0.82622462
E	661	D	1, 0.24	Neutral	0.82622462

Table 2b: List of PredictSNP predictions in Negeri Sembilan from July 2021 – May 2022

I	**693**	**V**	**1, 0.24**	**Neutral**	**0.82622462**
N	925	S	1, 0.24	Neutral	0.82622462
R	214	L	1, 0.24	Neutral	0.75291375
V	62	I	1, 0.24	Neutral	0.73834499
I	1210	V	1, 0.24	Neutral	0.73834499
T	1117	I	1, 0.24	Neutral	0.73688811
G	1251	R	1, 0.24	Neutral	0.62831343
S	256	L	1, 0.24	Neutral	0.6025641
V	915	I	1, 0.24	Neutral	0.6025641
D	1084	A	1, 0.24	Neutral	0.6025641
L	18	F	2, 0.48	Neutral	0.68365861
A	701	V	2, 0.48	Neutral	0.82622462
V	1176	F	2, 0.48	Neutral	0.82622462
A	93	S	2, 0.48	Neutral	0.82622462
V	171	I	2, 0.48	Neutral	0.82622462
E	1262	D	2, 0.48	Neutral	0.65307311
N	81	D	2, 0.48	Neutral	0.62831343
T	573	I	5, 1.20	Neutral	0.82622462
G	446	V	5, 1.20	Neutral	0.75291375
Q	787	K	5, 1.20	Neutral	0.73834499
K	1205	N	6, 1.44	Neutral	0.73834499
T	19	R	46, 11.06	Neutral	0.63151762
V	213	G	50, 12.02	Neutral	0.62831343
T	376	A	51, 12.26	Neutral	0.65307311
R	408	S	51, 12.26	Neutral	0.73688811
S	371	F	51, 12.26	Neutral	0.73834499
L	212	I	83, 19.95	Neutral	0.82622462
G	446	S	89, 21.39	Neutral	0.82622462
A	222	V	90, 21.63	Neutral	0.82622462
R	346	K	104, 25.00	Neutral	0.82622462
G	496	S	111, 26.68	Neutral	0.63151762
T	547	K	111, 26.68	Neutral	0.82622462
S	371	L	111, 26.68	Neutral	0.82622462
K	417	N	137, 32.93	Neutral	0.73834499
N	440	K	138, 33.17	Neutral	0.73688811
S	477	N	154, 37.02	Neutral	0.82622462
D	796	Y	156, 37.50	Neutral	0.75203963
Q	954	H	156, 37.50	Neutral	0.73688811
Q	493	R	161, 38.70	Neutral	0.75203963
N	501	Y	162, 38.94	Neutral	0.6025641
P	681	H	162, 38.94	Neutral	0.75291375
H	655	Y	162, 38.94	Neutral	0.73688811
E	484	A	162, 38.94	Neutral	0.82622462
G	339	D	162, 38.94	Neutral	0.73834499
S	373	P	162, 38.94	Neutral	0.82622462
S	375	F	162, 38.94	Neutral	0.75203963
Q	498	R	162, 38.94	Neutral	0.68365861
N	679	K	162, 38.94	Neutral	0.82622462
E	156	G	232, 55.77	Neutral	0.73834499
T	19	R	243, 58.41	Neutral	0.65307311
D	950	N	249, 59.86	Neutral	0.82622462
G	142	D	295, 70.91	Neutral	0.74796037
T	478	K	401, 96.39	Neutral	0.63151762
D	614	G	414, 99.52	Neutral	0.82622462

DISCUSSION

During the vaccination campaign in Malaysia, Negeri Sembilan achieved the highest COVID-19 vaccination rates. This study aimed to identify any circulating mutations in Negeri Sembilan following the rapid vaccinations. Access to sequenced genomes was made possible through the Ministry of Science, Technology, and innovation sequencing consortium. Previous research highlighted variant trends in Malaysia, which differed across states (20). By comparing Negeri Sembilan and neighboring Selangor, the state with the highest infection rates, we aimed to understand the impact of vaccinations, policy changes, and geographical factors on SARS-CoV-2 conditions.

Our observations for Negeri Sembilan were approximately similar to the trends at the national levels but with a higher dominance for the AY.79 lineage. The highest reports of AY.79 infection were recorded by Terengganu and Perlis, followed by Negeri Sembilan, suggesting a route of infection that could have stemmed from Terengganu and Perlis. Additionally, since the relaxing of lockdowns in October 2021, interstate travel allowed infection to spread between states (20). During this period, efforts were concentrated on preventing infection and thus focused more on providing treatment and vaccination than on genomic surveillance (21). The rapid and intensive vaccination strategy reduced Delta cases in October, to which SARS-COV-2 mutated into the Omicron variant.

In October, the number of vaccinations increased in Malaysia, which led to the decision to relax national lockdowns. During the time, vaccinations were thought to protect from Delta infections. Consequentially, the release of lockdowns led to a sharp increase in Omicron cases in Negeri Sembilan from October 2021. The sharp spike in cases in December 2021 correlated to the high volume of travel post-lockdowns, school holidays and mass gatherings. In Negeri Sembilan, the two significant lineages from the phylogenetic analysis show that the predominating variants are Delta (B.1.617.2) and Omicron (BA.1.1, BA.2.40.1, BA.2.57 and BA.2.9), which was highly found in Seremban and Tampin in Negeri Sembilan.

The collected data for sequencing and deposited in the GIASID database have limitations and biases due to selected samples (22,23). These samples included patients under close surveillance, international point of entry cases, pediatric cases, severe cases of older adults without comorbidities, patients with unusual manifestations, suspected reinfection cases, breakthrough infections in healthcare workers, clinical cases with negative RT-PCR results, and sporadic cases. However, some categories like patients with comorbidities, children, and asymptomatic patients need to be included for a more comprehensive analysis.

Our results also show several novel mutations circulating within a sub-rural region of Malaysia. In our dataset, 50 SNVs occurred in 42-417 samples and were categorised as high-frequency mutations. Thus, it was interesting to note that the five novel mutations emerged as harmful mutations and were captured in only one or two sequenced samples. Although SNVs at high frequencies have previously been shown to increase viral transmissibility and virulence, detecting short-lived S protein mutations is crucial to predict future mutational sites in emerging variants (24). Based on the location of the modifications, it was difficult to deduce how the predictions would change the protein function. However, molecular dynamics of the S proteins of the mutants could determine the structural changes caused by the mutation.

There is a need to analyse new circulating mutations, which could increase viral fitness. Genomic surveillance at regional levels is vital as the results will support clinical decisions and epidemiological management during outbreaks. Furthermore, newer variants pose new challenges due to their ability to cause reinfection, even in vaccinated people. Reports have also shown that new variants can now escape therapeutics and diagnostics. As SARS-CoV-2 mutates over time, more unique variants may have increased fitness and infectivity and induce different symptoms based on the mutations.

Increased genomic surveillance and data sharing worldwide are needed to support the timely and responsive delivery of public health interventions. Data-driven public health interventions have previously been shown to effectively hamper disease transmission, allow rapid vaccine-design improvements and curation of appropriate action plans and policy updates. However, as of January 2022, only 68% of the global countries have enough support and capacity for rapid genomic surveillance (WHO). Therefore, international collaborations, partnering and sharing resources, particularly to countries with limited resources, are imperative to increase the frequency of genomic surveillance.

CONCLUSION

Our work analyzed spatiotemporal SARS-CoV-2 lineages in Negeri Sembilan. The observed lineages exhibited rapid evolution across age groups and districts, highlighting the need for real-time surveillance to monitor COVID-19 infections. The data can aid in understanding SARS-CoV-2 pathogenesis and inform policymaking for Malaysia.

ACKNOWLEDGEMENT

The authors acknowledge all data contributors, i.e., laboratories responsible for obtaining the specimens, and their sequencing laboratories for generating the genetic sequence and metadata and sharing via the GISAID Initiative, on which this research is based. This research was funded by USIM Internal Grant, grant number PPPI/FPSK/0122/USIM/14322 and PPPI/FPSK/0122/USIM/14622. Publicly available datasets were analysed in this study. This data can be found here: https://gisaid.org/.

Conflict of interests
The authors declare no potential conflict of interest.

REFERENCES

1. A. Muhammad Azami et al., "SARS-CoV-2 genomic surveillance in Malaysia: displacement of B.1.617.2 with AY lineages as the dominant Delta variants and the introduction of Omicron during the fourth epidemic wave," *Int J Infect Dis* 2022, 125:216-226.

2. Bendl et al., "PredictSNP: Robust and Accurate Consensus Classifier for Prediction of Disease-Related Mutations," *PLoS Comput Biol* 2014, 10(1): e1003440.

3. Department of Statistics Malaysia Official Portal. Available online: https://www.dosm.gov.my/v1/index.php?r=column/cthemeByCat&cat=401&bul_id=QkxLckg3WjlzcEZyVzRIajllenBIQT09&menu_id=L0pheU43NWJwRWVSZklWdzQ4TlhUUT09. (Accessed Dec. 12, 2022).

4. First-level Administrative Divisions, Malaysia, 2015 | Stanford Digital Repository. Available online: https://purl.stanford.edu/qg469kj1734. (Accessed Jan. 06, 2023).

5. H. Hashim, M. A. Adman, Z. Hashim, M. F. Mohd Radi, and S. C. Kwan, "COVID-19 Epidemic in Malaysia: Epidemic Progression, Challenges, and Response," *Front Public Health* 2021, 9:560792.

6. H. M. Mengist, A. J. Kombe Kombe, D. Mekonnen, A. Abebaw, M. Getachew, and T. Jin, Mutations of SARS-CoV-2 spike protein: Implications on immune evasion and vaccine-induced immunity, *Semin Immunol* 2021, 55:101533.

7. J. Rozewicki, S. Li, K. M. Amada, D. M. Standley, and K. Katoh, "MAFFT-DASH: integrated protein sequence and structural alignment," *Nucleic Acids Res* 2019, 47(W1): W5-W10.

8. Verity Hill, George Githinji et al., "Toward a global virus genomic surveillance network", *Cell Host & Microbe* 2023, 31:6:6

9. Caroline Dushoff, David JD Dushhoff et al., "First Nations Health: The need for linked genomic surveillance of SARS-CoV-2", *Canada Communicable Disease Report* 2022, 48:4:4

10. Letunic and P. Bork, "Interactive Tree of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation," *Nucleic Acids Res* 2021, 49(W1): W293-W296.

11. Mohamad et al., Compliance to Infection Prevention and Control Practices Among Healthcare Workers During COVID-19 Pandemic in Malaysia, *Front Public Health* 2022, 10:878396.

12. Negeri Sembilan, Selangor, Perak Record Biggest COVID-19 Vaccination Surges - CodeBlue. Available online: https://codeblue.galencentre.org/2021/07/15/negeri-sembilan-selangor-perak-record-biggest-covid-19-vaccination-surges/. (Accessed Jan. 18, 2023).

13. Pan health Organization (PAHO), Guidance on SARS-CoV-2 sampling for genomic surveillance Guidance for SARS-CoV-2 samples selection for genomic characterization and surveillance, no. February, pp. 1–3, 2021, Available: https://www.paho.org/en/documents/guidance-sars-cov-2-samples-selection-genomic-characterization-and-surveillance. (Accessed Apr. 13, 2023).

14. R. Isnain, N. C. Dom, S. Abdullah, N. Precha, and H. Salim, Efficiency of Malaysian states in managing the COVID-19 outbreak in 2020 and 2021, *PLoS One* 2022, 17(10).

15. RM15 million in additional funds for third phase of SARS-CoV-2 genomic surveillance project. Available online: https://www.nst.com.my/news/nation/2022/05/794802/rm15-million-additional-funds-third-phase-sars-cov-2-genomic-surveillance. (Accessed Dec. 06, 2022).

16. S. M. Ghazali et al., COVID-19 in Malaysia: Descriptive Epidemiologic Characteristics of the First Wave, *International Journal of Environmental Research and Public Health 2022* 2022, 19(328).

17. S. Shrivastava and P. Shrivastava, 2019-nCoV Outbreak Declared as Public Health Emergency of

International Concern: What Next? *Int J Prev Med* 2020, 11(1).'

18. Tamura, G. Stecher, and S. Kumar, MEGA11: Molecular Evolutionary Genetics Analysis Version 11, *Mol Biol Evol* 2021, 25;38(7):3022-3027.

19. V. J. Jayaraj, S. Rampal, C. W. Ng, and D. W. Q. Chong, The Epidemiology of COVID-19 in Malaysia, *Lancet Reg Health West Pac* 2021, 17:100295.
20. Wang S, Sundaram JP, Spiro D. VIGOR, an annotation program for small viral genomes. *BMC Bioinformatics* 2010, 11(451).

21. WHO and ECDC, Methods for The Detection and Characterisation of SARS-CoV-2 Variants-Second Update, *European Centre for Disease Prevention and Control/World Health Organization Regional Office for Europe*, no. August, pp. 1–14, 2022, [Online]. [Retrieved Apr. 13, 2023]. Available online: https://www.ecdc.europa.eu/en/publications-data/methods-detection-and-characterisation-sars-cov-2-variants-second-update

22. WHO Coronavirus (COVID-19) Dashboard | WHO Coronavirus (COVID-19) Dashboard with Vaccination Data. Available online: https://covid19.who.int/. (Accessed Dec. 06, 2022).

23. Y. Yu *et al.*, Whole genome sequencing analysis of SARS-CoV-2 from Malaysia: From alpha to Omicron, *Front Med (Lausanne)* 2022, 9:1001022.

24. Yin, "Genotyping coronavirus SARS-CoV-2: methods and implications," *Genomics* 2022, 112(5):3588-3596.