Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Materials Today: Proceedings 48 (2022) 828–836

Contents lists available at ScienceDirect

Materials Today: Proceedings

journal homepage: www.elsevier.com/locate/matpr

COVID-19 outbreak in Malaysia: Decoding D614G mutation of SARS-CoV-2 virus isolated from an asymptomatic case in Pahang

Aini Syahida Mat Yassim [a], Mohd Fazli Farida Asras [a], Ahmad Mahfuz Gazali [a,d], Martin S. Marcial-Coba [b], Ummu Afeera Zainulabid [c], Hajar Fauzan Ahmad [a,d,*]

[a] Faculty Industrial Sciences and Technology, Lebuhraya Tun Razak, 26300 Gambang, Pahang, Malaysia
[b] Pontifical Catholic University of Ecuador, 170143 Quito, Ecuador
[c] Kulliyyah of Medicine, International Islamic University Malaysia, 25200 Kuantan, Pahang, Malaysia
[d] Centre for Research in Advanced Tropical Bioscience (Biotropic Centre), Lebuhraya Tun Razak, 26300 Gambang, Pahang, Malaysia

ARTICLE INFO

Article history:
Available online 27 February 2021

Keywords:
SARS-CoV-2
Mutation
D614G
Spike protein
Malaysia
Pahang

ABSTRACT

SARS-CoV-2 is a very transmissible and pathogenic coronavirus which detected in Malaysia in January 2020. Nevertheless, the sample from Malaysia is still under-sequenced. Hence lacking clarity of the circulating strain in Malaysia leads to a deadlock in understanding the virus infectivity. This study aimed to investigate the genome identity of circulating COVID-19 strains in Pahang and understand disease epidemiology during the pandemic. This study leveraged high-throughput sequencing analysis for the whole genome sequencing and implemented bioinformatic technique for the analysis. Here we reported that the virus with D614G mutation in Spike protein circulates in a few Malaysia states before the Sivagangga cluster announced in Kedah in July 2020. This mutated virus includes our virus sample isolated in April 2020 from an asymptomatic patient in Pahang. Based on the phylogenetic analysis, we discovered the origin of our sample Pahang/IIUM91 was not related to Sivagangga cluster. Here, we have generated 3D structure model of Pahang/IIUM91 Spike protein. D614G mutation in Pahang/IIUM91 Spike protein increases viral stability and flexibility, hence render higher infectivity. Collectively, our results suggest for the establishment of a complete SARS-CoV-2 genome database in Malaysia. Hence, more research should be established to learn the behaviour of this virus.

Selection and peer-review under responsibility of the scientific committee of the 2nd International Conference on Innovative Technology and Sciences (iCITES 2020).

1. Introduction

Globally until January 16, 2021, World Health Organization (WHO) has reported 92,262,621 confirmed cases of COVID-19 with approximately 0.8% of new cases every day, including 1,995,037 deaths, caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) [1]. In Malaysia, WHO has reported that the first wave of COVID-19 infection started on January 24, 2020, identifying 22 cases [2]. As of January 16, 2021, a total of 155,095 positive cases, including 594 deaths, had been reported to the Ministry of Health, Malaysia (MOH) [3]. Whereas until early January 2021, there are 312,896 complete and partial genomes of SARS-CoV-2 deposited to Global Initiative on Sharing All Influenza Data (GISAID) database, of which contributed by clinicians and researchers worldwide [4,5]. This online sharing of SARS-CoV-2 genomes provides insights into the virus's ongoing evolution and epidemiology during the pandemic and will likely play an essential role in the surveillance and eventual mitigation and control [6]. Nevertheless, the number of whole genomes of SARS-CoV-2 Malaysian strains deposited in GISAID is still under-sequenced. As of January 7, 2020, there were only 250 of high coverages of Malaysian strains of the SARS-CoV-2 complete genome has been deposited into the GISAID database by local researchers [4,5]. Hence the lack of data available to assign the current circulating strain corresponding to the significant clusters of COVID-19 reported by MOH.

Earlier studies reported nine different lineages of SARS-CoV-2; A, B, B.1, B.1.1, B.1.1.1, B.1.36, B.2, B.3 and B.6 were circulating from the second wave of infections [7], started from February 27, 2020 [2]. Among these lineages, it reported that lineage B.6, named Indian lineage [6] had become the predominant cause of community transmission in Malaysia, linked to Tablighi Jamaat cluster [7]. Hence suggesting that the lineage B.6 have established community transmission [7,8]. Duchene et al. [9] suggest that circulat-

* Corresponding author.
E-mail address: fauzanahmad@ump.edu.my (H.F. Ahmad).

https://doi.org/10.1016/j.matpr.2021.02.387
2214-7853/
Selection and peer-review under responsibility of the scientific committee of the 2nd International Conference on Innovative Technology and Sciences (iCITES 2020).

ing SARS-CoV-2 lineages accumulate nucleotide mutations at about 1–2 mutations per months, with the pylodynamic threshold attained about two months of the estimated start of the outbreak. A recent announcement by MOH revealed that five clusters in Malaysia, namely Benteng (23 viruses), Sivagangga (4 viruses), Tawar (3 viruses), Sungai (1 virus) and Bukit Tiram (1 virus); were found to display the D614G mutation in Spike protein [10].

Analysis of more than 28,000 *S* gene sequence in May 2020 revealed that the variant carrying the D614G Spike mutation became the globally dominant form of SARS-CoV-2 [11]. Zhang et al. [12] reported the mutant virus with glycine at the residue 614 (G614) of Spike protein, replacing aspartic acid (D614) was not detected in January to February 2020, but infrequently observed in March 2020. The frequency of D614G genotype expands by April to May 2020 [12]. Increases in the frequency of this set of mutation during co-circulation within individual regions during outbreaks, suggesting that the increase resulted from a fitness advantage rather than founder effects and/or genetic drift [13]. Plante et al. [13] reported that the G614 virus variant could replicate with a higher viruses titre, hence outcompeting the D614 virus when infecting human airway tissues. They also show that the G614 variant retained higher infectivity at various temperature tested, thus suggesting a D614G mutation increase the stability of SARS-CoV-2 [13].

The necessity of characterising the geographical spread and molecular evolution of SARS-CoV-2, through extensive global sequencing efforts, mainly relies on determining the biological significance of the detected mutations [14,15]. In this context, global tracking data proposed by Korber et al. [11] suggested that the G614 variant in Spike has spread faster than D614 variant. Hence D614G genotype is likely to be more infectious, due to higher viral loads in COVID-19 patients infected with G614 variant [11,16]. Zhang et al. [12] suggest that D614G mutation increased virus infectivity by assembling more functional Spike protein density in the virion, allowing more efficient person-to-person transmission. Nevertheless, it has been observed that the spike D614G substitution increases the susceptibility of G614 virus to neutralisation by antibodies, suggesting that the efficacy of vaccines, designed based on the original D614 spike sequence, could not be reduced [13].

Here, we analysed the dominance lineage of SARS-CoV-2 currently circulating in Malaysia using the whole genome of Malaysian SARS-CoV-2 available in GISAID. Accordingly, we analysed the relative frequency of D614 variant compared to G614 variant in Spike protein of Malaysian SARS-CoV-2 and summarised the G614 variant deposited in the GISAID database. We also investigated the G614 variant Spike protein divergence of Pahang-hCoV19/Malaysia/IIUM91/2020 (later called Pahang/IIUM91) relative to another G614 variant of hCov-19/Malaysia, hence its possible origin. Finally, we presented a possible Spike protein 3D structure of Pahang/IIUM91 for future reference.

2. Materials and method

2.1. Sample processing

The sample used for this study was considered excess diagnostics material, where the leftovers of RNA extract was subjected for whole genome sequencing and reported elsewhere. Briefly, the nasopharyngeal and oropharyngeal swab and sputum samples were collected on April 2, 2020, from an asymptomatic patient. The RNA was extracted before real-time reverse transcriptase (RT) -PCR procedures to detect SARS-CoV-2. The genome details were deposited in public databases such as the National Center of Biotechnology Information (NCBI), and Global Initiative on Sharing All Influenza Data (GISAID).

2.2. Next-generation sequencing of the full-length viral genome

A next-generation sequencing (NGS) library was constructed after amplifying the isolates' full-length genes using the synthesised cDNA using SuperScriptIV (Invitrogen) with some modifications. Briefly, 5 μl of the cDNA was used as the template for multiplex PCR using Q5 polymerase (NEB, Ipwich, MA) and the Artic v3 primer pools during library preparation [17]. The constructed library was sequenced on an iSeq 100 (run configuration of 1 × 300 bp).

2.3. Sequence analysis

The SARS-CoV-2 genome was reconstructed from the raw reads using a combination of a bioinformatic tool as listed in https://github.com/CDCgov/SARS-CoV-2_Sequencing/tree/master/protocols/BFX-UT_ARTIC_Illumina. The genome sequences from other studies related to human and animal coronavirus sequences were mined from the GISAID (https://www.gisaid.org) and NCBI GenBank (https://www.ncbi.nlm.nih.gov/genbank/).

2.4. Public database SARS-CoV-2 genome analysis

Specifically, a total of 292 whole-genome sequences of SARS-CoV-2 of Malaysia uploaded to GISAID were retrieved up to January 7, 2021, for the analysis of dominance lineage and D614G frequency. Only high coverage complete sequences (n = 250) were kept for analysis. Analysis of dominance lineage was done manually by categorising downloaded virus sequences based on their lineages. The frequencies of G614 over D614 virus variants were analysed using Nextstrain SARS-CoV-2 resources database (https://nextstrain.org/). One hundred sixty-nine complete sequences of Malaysian SARS-CoV-2, G614 variant obtained from GISAID was used to retrieve *S* gene and Spike protein sequences. Only high coverage complete sequences (n = 144) were kept for analysis. The *S* gene sequence of 114 Malaysian SARS-CoV-2, G614 variant was identified using multiple sequence alignment against the *S* gene of NCBI reference strain WuHan-Hu-1 genome (NC_045512.2:21563-25384) obtained from GenBank (https://www.ncbi.nlm.nih.gov/sars-cov-2/). The multiple sequence alignment was performed using DECIPHER [18] and SeqinR [19] packages in R version 4.0.2. and finalised using MEGA X 10.1 [20]. The identified *S* gene was translated into the amino acids sequence using MEGA X 10.1. The Spike protein of 114 Malaysian SARS-CoV-2, G614 variant was confirmed through multiple sequence alignment against Spike protein retrieved from NCBI reference sequence: YP_009724390.1 (GenPept). The multiple sequence alignment of amino acids was performed using DECIPHER and SeqinR packages in R version 4.0.2. Here, the Spike protein sequence of Pahang/IIUM91 was used to generate the phylogenetic tree and 3D structural protein.

2.5. Phylogenetic tree analysis

The evolutionary analysis was inferred using the Neighbour-Joining method [21]. The bootstrap consensus tree inferred from 1000 replicates was taken to represent the evolutionary history of the taxa analysed [22]. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates were collapsed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) were shown next to the branches. The evolutionary distances were computed using the Jones-Taylor-Thornton matrix-based method [23] and were in the units of the number of amino acid substitutions per site. The rate variation among sites was modelled with a gamma distribution (shape parameter = 2). All positions containing gaps

Table 1
The number of complete and high coverage SARS-CoV-2 genome (Malaysian isolate) for each lineage, and the summary of the linage.

Lineage	Number of complete and high coverage hCoV-19/Malaysia genome	Most common countries	Description
A	3	United_Arab_Emirates 21.0%, China 14.0%, USA 9.0%, Japan 5.0%, UK 5.0%	The root of the pandemic lies within lineage A. Many sequences originating from China and many global exports; including to South East Asia Japan South Korea Australia the USA and Europe represented in this lineage
B	18	UK 40.0%, USA 15.0%, China 13.0%, Spain 3.0%, Singapore 3.0%	Base of this lineage also lies in China, with many global exports, two distinct SNPs '8782TC' and '28144CT' define this lineage
B.1	9	USA 48.0%, UK 20.0%, Denmark 4.0%, France 3.0%, Australia 3.0%	A large European lineage that corresponds to the Italian outbreak.
B.1.1	16	UK 49.0%, USA 11.0%, Russia 4.0%, Portugal 3.0%, Denmark 3.0%	European lineage with 3 clear SNPs '28881GA', '28882GA', '28883GC'
B.1.1.1	2	UK 84.0%, Belgium 2.0%, Denmark 2.0%, Switzerland 2.0%, Peru 1.0%	UK/ Europe lineage
B.1.1.282	1	USA 96.0%, Mexico 2.0%, UK 2.0%	USA lineage
B.1.1.312	1	USA 100.0%	Gambian lineage
B.1.1.63	2	Hong_Kong 69.0%, Australia 15.0%, Singapore 3.0%, South_Korea 3.0%, UK 2.0%	Hong Kong lineage
B.1.160.3	2	Indonesia 61.0%, Singapore 17.0%, Malaysia 9.0%, Hong_Kong 4.0%, Taiwan 4.0%	Indonesian lineage
B.1.246	1	Saudi_Arabia 55.0%, India 15.0%, UK 10.0%, Nigeria 6.0%, Australia 4.0%	Saudi Arabian lineage, previously some assigned B.1.160
B.1.247	2	Denmark 24.0%, India 22.0%, UK 20.0%, Saudi_Arabia 13.0%, Australia 4.0%	Indian/ Saudi Arabian diversity, now European and Australian diversity too, previously some assigned B.1.160 and B.1.36
B.1.255	1	USA 70.0%, UK 5.0%, Canada 4.0%, Australia 3.0%, Colombia 2.0%	North American lineage, with other global diversity
B.1.36	2	UK 46.0%, India 25.0%, Denmark 9.0%, Australia 3.0%, Saudi_Arabia 3.0%	Global lineage with lots of representation of sequences from India and Saudi Arabia. Sequences also from Europe and the UK. Now includes some sequences that previously had been assigned B.1.113.
B.1.36.16	6	UK 40.0%, Thailand 30.0%, Bangladesh 13.0%, Singapore 7.0%, Malaysia 5.0%	Bangladesh/UK lineage
B.1.5	**97**	**UK 25.0%, USA 20.0%, Spain 12.0%, Switzerland 7.0%, France 3.0%**	**Spanish base, European lineage/ lots of Spanish sequences towards the basal end of the subtree and exports worldwide.**
B.1.78	1	Netherlands 93.0%, UK 4.0%, USA 1.0%, Malaysia 1.0%	Netherlands lineage
B.1.98	1	UK 74.0%, USA 16.0%, Australia 4.0%, Canada 1.0%, Chile 1.0%	Some reassigned to B.1.5.33, UK lineage
B.12	1	Japan 97.0%, Malaysia 3.0%	Japanese lineage
B.28	1	UK 78.0%, Australia 6.0%, USA 5.0%, Jordan 2.0%, Canada 1.0%	UK lineage
B.3	1	UK 61.0%, Germany 9.0%, Belgium 4.0%, Australia 4.0%, Denmark 4.0%	A European lineage
B.6	45	India 56.0%, Australia 8.0%, Malaysia 6.0%, Singapore 5.0%, USA 5.0%	Indian lineage
B.6.1	18	Malaysia 91.0%, India 5.0%, Brunei 5.0%	Malaysian lineage
B.6.2	12	Malaysia 100.0%	Malaysian lineage
B.6.6	7	Singapore 92.0%, India 7.0%, Malaysia 1.0%, Australia 1.0%	Singapore lineage

Notes: The analysis presented here was based on January 7, 2021. The current major lineage circulated in Malaysia was highlighted in bold.
The lineage description was described as in lineage database (https://cov-lineages.org/lineages.html).

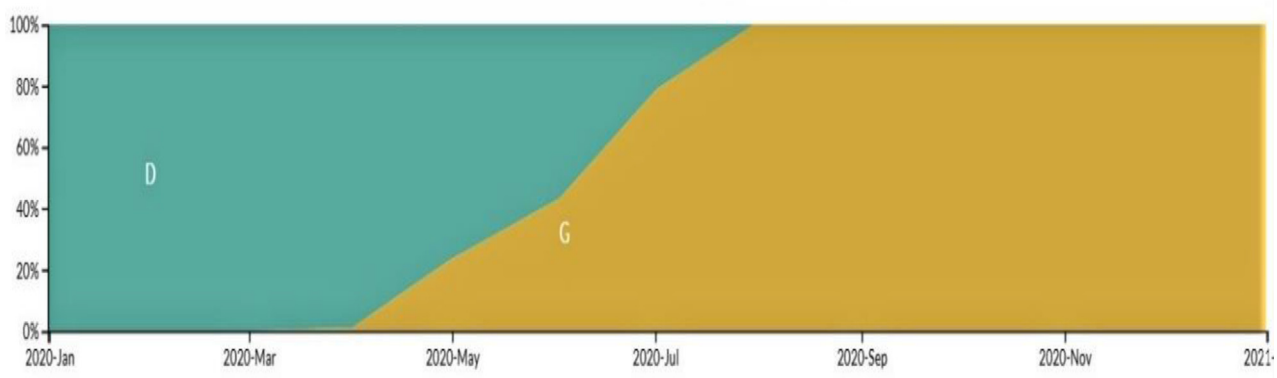

Fig. 1. Frequencies of D614 variant and G614 variant since first detected in Malaysia (colored by Genotype at Spike protein, positioned 614 and normalized to 100% at each time point for 56 out of total of 4017 tips). The figure was generated using Nextstrain SARS-CoV-2 resources database (https://nextstrain.org/ncov/asia?c=gt-S_614&f_division=Malaysia).

and missing data were eliminated (complete deletion option). Evolutionary analyses were conducted in MEGA X [20].

2.6. 3D structure of protein molecular modelling of Spike protein

The 3D structure of G614 Spike protein of Pahang/IIUM91 was modelled using the SWISS-MODEL server [24] using the most fitted protein template available from Protein Database Bank (PDB). Model quality was evaluated by Qualitative Model Energy ANalysis (QMEAN) [25,26], while the structure of the model was visualised using the PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC. The 3D structure of D614 Spike protein of YP_009724390.1 was uploaded onto the DynaMut web server [27] to examine the effect of D614G mutation on Pahang/IIUM91 Spike protein.

3. Results and discussion

Targeting for a complete and high coverage SARS-CoV-2, the 250 SARS-Cov-2 Malaysian isolates were categorised accordingly

Table 2
The hCoV-19/Malaysia harbouring D614G at Spike protein (March to May 2020).

Virus name	Accession ID	Collection date	Location information	Lineage
hCoV-19/Malaysia/IIUM91/2020	EPI_ISL_455313	2/4/2020	Pahang	B.1.247
hCoV-19/Malaysia/IMR_WC55122/2020	EPI_ISL_490089	10/5/2020	Selangor	B.1.1.1
hCoV-19/Malaysia/IMR_WC90685/2020	EPI_ISL_490101	29/5/2020	Not Specified	B.1.1.1
hCoV-19/Malaysia/IMR_WC94764/2020	EPI_ISL_490103	29/5/2020	Not Specified	B.1.255
hCoV-19/Malaysia/0121/2020	EPI_ISL_501176	21/3/2020	Kuala Lumpur	B.1.1
hCoV-19/Malaysia/0309/2020	EPI_ISL_501177	22/3/2020	Kuala Lumpur	B.1.1
hCoV-19/Malaysia/1497/2020	EPI_ISL_501185	25/3/2020	Kuala Lumpur	B.1.1
hCoV-19/Malaysia/3479/2020	EPI_ISL_501204	30/3/2020	Kuala Lumpur	B.1.1
hCoV-19/Malaysia/3611/2020	EPI_ISL_501207	31/3/2020	Kuala Lumpur	B.1.1.282
hCoV-19/Malaysia/8451/2020	EPI_ISL_501222	12/4/2020	Kuala Lumpur	B.1.98
hCoV-19/Malaysia/MGI-G873/2020	EPI_ISL_528738	7/4/2020	Selangor	B.1.246
hCoV-19/Malaysia/UNIMAS-0729/2020	EPI_ISL_718132	7/4/2020	Sarawak	B.1
hCoV-19/Malaysia/UNIMAS-0803/2020	EPI_ISL_718136	8/4/2020	Sarawak	B.1
hCoV-19/Malaysia/UNIMAS-0298/2020	EPI_ISL_718138	2/4/2020	Sarawak	B.6
hCoV-19/Malaysia/UNIMAS-3325/2020	EPI_ISL_718139	21/4/2020	Sarawak	B.1
hCoV-19/Malaysia/UNIMAS-0179/2020	EPI_ISL_718140	1/4/2020	Sarawak	B.1
hCoV-19/Malaysia/UNIMAS-M0710/2020	EPI_ISL_718145	13/4/2020	Sarawak	B.1.5
hCoV-19/Malaysia/UNIMAS-0358/2020	EPI_ISL_718148	3/4/2020	Sarawak	B.1
hCoV-19/Malaysia/UNIMAS-1020/2020	EPI_ISL_718149	9/4/2020	Sarawak	B.1.5
hCoV-19/Malaysia/UNIMAS-0172/2020	EPI_ISL_718150	31/3/2020	Sarawak	B.1.5
hCoV-19/Malaysia/UNIMAS-1741/2020	EPI_ISL_718151	13/4/2020	Sarawak	B.1.5
hCoV-19/Malaysia/UNIMAS-0136/2020	EPI_ISL_718152	31/3/2020	Sarawak	B.1.5
hCoV-19/Malaysia/UNIMAS-1589/2020	EPI_ISL_718153	13/4/2020	Sarawak	B.1.5
hCoV-19/Malaysia/UNIMAS-M3002/2020	EPI_ISL_718154	7/5/2020	Sarawak	B.1
hCoV-19/Malaysia/UNIMAS-2304/2020	EPI_ISL_718155	16/4/2020	Sarawak	B.1.5
hCoV-19/Malaysia/UNIMAS-0243/2020	EPI_ISL_718157	1/4/2020	Sarawak	B.1.5
hCoV-19/Malaysia/UNIMAS-4248R/2020	EPI_ISL_718158	25/4/2020	Sarawak	B.1.5
hCoV-19/Malaysia/UNIMAS-1197/2020	EPI_ISL_718161	10/4/2020	Sarawak	B.1.5
hCoV-19/Malaysia/UNIMAS-0242/2020	EPI_ISL_718162	1/4/2020	Sarawak	B.1.5
hCoV-19/Malaysia/UNIMAS-0099/2020	EPI_ISL_718163	31/3/2020	Sarawak	B.1
hCoV-19/Malaysia/UNIMAS-0217/2020	EPI_ISL_718164	1/4/2020	Sarawak	B.1.5

*Pahang/IIUM91 was highlighted in grey. The data presented here were analysed using dataset available in the GISAID database and was based on January 7, 2021.

Table 3
Summary of G614 hCoV-19/Malaysia variant.

States	Lineage	Number of the virus with D614G at Spike protein	First virus identified with D614G mutation in Spike protein	Earliest collection date
Pahang	B.1.247	1	hCoV-19/Malaysia/IIUM91/2020	2/4/2020
Selangor	B.1.1.1	1	hCoV-19/Malaysia/IMR_WC55122/2020	10/5/2020
	B.1.246	1	hCoV-19/Malaysia/MGI-G873/2020	7/4/2020
Kuala Lumpur	B.1.1	4	hCoV-19/Malaysia/0121/2020	21/3/2020
	B.1.1.282	1	hCoV-19/Malaysia/3611/2020	31/3/2020
	B.1.98	1	hCoV-19/Malaysia/8451/2020	12/4/2020
Sarawak	B.1	8	hCoV-19/Malaysia/UNIMAS-0099/2020	31/3/2020
	B.1.5	60	hCoV-19/Malaysia/UNIMAS-0172/2020	31/3/2020
	B.1.160.3	2	hCoV-19/Malaysia/UNIMAS-15695/2020	14/10/2020
	B.1.36	2	hCoV-19/Malaysia/UNIMAS-15723/2020	28/10/2020
	B.1.1.312	1	hCoV-19/Malaysia/UNIMAS-15784/2020	12/11/2020
	B.6	1	hCoV-19/Malaysia/UNIMAS-0298/2020	2/4/2020
Not specified	B.1	1	hCoV-19/Malaysia/IMR-CV136859/2020	1/9/2020
	B.1.1	12	hCoV-19/Malaysia/IMR-WI085/2020	27/7/2020
	B.1.1.1	1	hCoV-19/Malaysia/IMR_WC90685/2020	29/5/2020
	B.1.1.63	2	hCoV-19/Malaysia/IMR-WI109/2020	25/7/2020
	B.1.36.16	6	hCoV-19/Malaysia/IMR-WI194/2020	16/10/2020
	B.1.255	1	hCoV-19/Malaysia/IMR_WC94764/2020	29/5/2020
	B.1.5	37	hCoV-19/Malaysia/IMR-CV138548/2020	1/9/2020
	B.1.247	1	hCoV-19/Malaysia/IMR-WC206980/2020	12/10/2020
Total		144		

*The data presented here were analysed using dataset available in GISAID database and was based on January 7, 2021.

to its lineages. As of the analysis presented here is based on January 7, 2021, there were 24 lineages of SARS-CoV-2 circulating in Malaysia, with the major lineage reported here was B.1.5 (Table 1, highlighted in bold). Further analysis using lineage database (https://cov-lineages.org/lineages.html) shows that the lineage of B.1.5 was found to be common in the UK (25.0%), USA (20.0%), Spain (12.0%), Switzerland (7.0%), France (3.0%), hence described as Spanish base, European lineage/lots of Spanish sequences towards the basal end of the subtree and exports around the globe (Table 1, highlighted in bold). The lack of available clinical metadata deposited in GISAID prevented our investigation of the association between viral lineage and severity of circulating virus in Malaysia. Besides, the number of whole-genome sequence of hCoV-19/Malaysia strains deposited in GISAID is still not representing the real patterns of the SARS-CoV-2 outbreak in Malaysia. Our analysis found that, of 250 whole-genome sequences of hCoV-19/Malaysia, 77 complete sequences were not specified to the states where the samples were isolated. On top of that, there were only four states of Malaysia had contributed to the complete and high coverage whole genome sequence of hCoV-19/Malaysia; Selangor (25), Kuala Lumpur (57), Pahang (2) and Sarawak (89). This significant limitation faced in our present study is likely to be a significant hurdle to similar studies. The issues could be resolved by establishing a complete 2019 Novel Coronavirus (SARS-CoV-2) Strain Genome Database in Malaysia assigning a current circulating strain to the corresponding cluster and better clarity of strain and mutation identification.

As of January 16, 2021, MOH has announced the number of new cases of positive COVID-19 soar to 4,029, the highest record of daily positive cases reported in Malaysia [3]. In this study, we proposed that the vast majority of all new cases of COVID-19 in Malaysia may be contributed by an increase in normalising frequency of

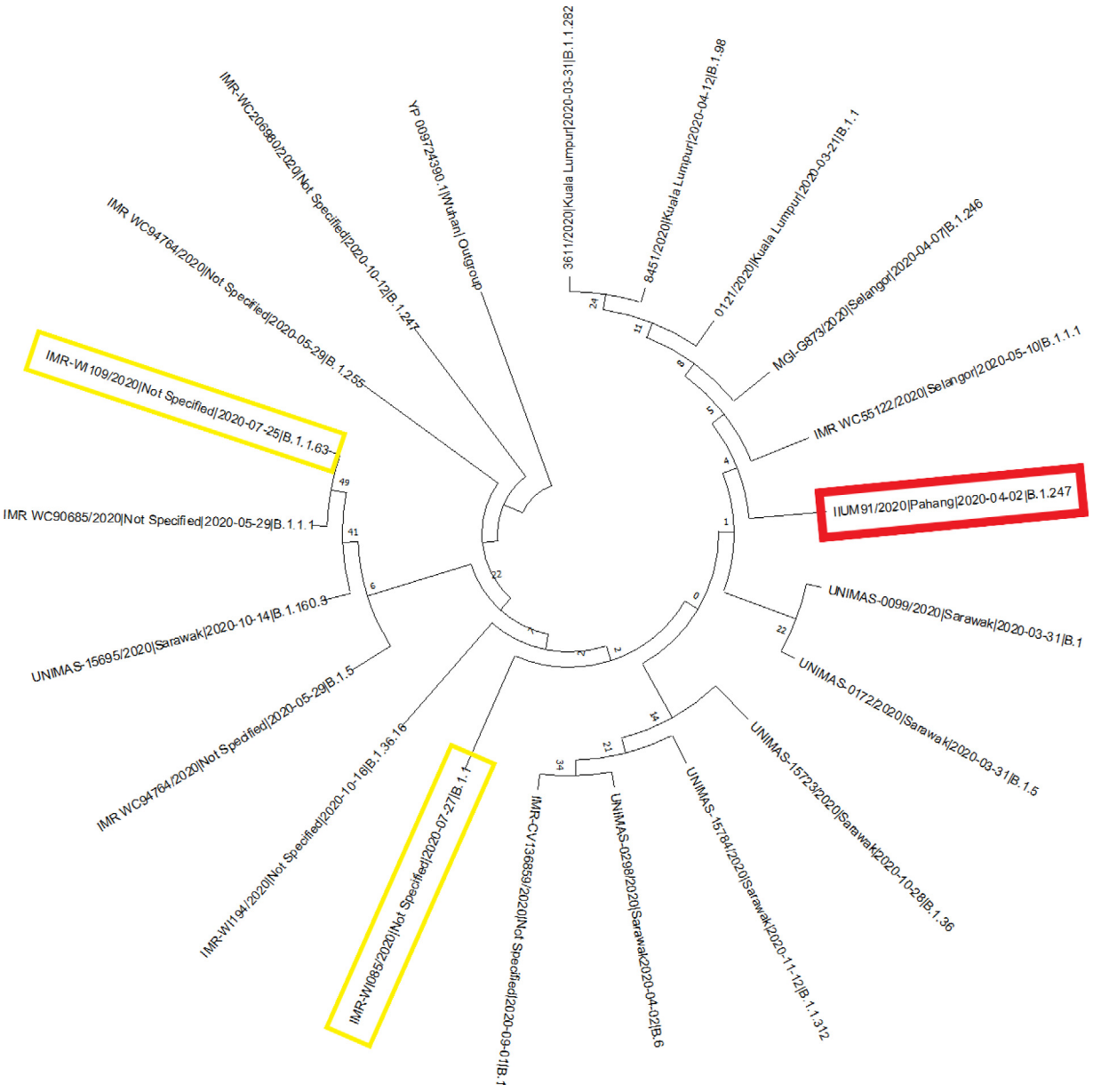

Fig. 2. Neighbor-Joining phylogenetic tree of G614 variant Spike protein built from 20 virus sequences first identified in Malaysia (Table 3). NCBI reference sequence: YP_009724390.1, D614 variant Spike protein from reference strain WuHan-Hu-1 genome (NC_045512.2) was used as outgroup. Pahang/IIUM91 was labelled in a red box. Possible virus isolates from Sivagangga cluster based on the sample collection date (July 2020) were labelled in a yellow box. The tree was rooted to the YP_009724390.1. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

the G614 variant over D614 variant circulating in this country since first detected in Malaysia until January 2021 (Fig. 1). While MOH announced D614G mutation was first identified in Malaysia on July 13, 2020, which belongs to Sivagangga cluster in Kedah [28], our analysis revealed 31 strains of hCoV-19/Malaysia harbouring D614G mutation were detected as early as in March to May 2020 (Table 2). Moreover, the data stated here was supported by the data reported in the Nextstrain SARS-CoV-2 resources database (https://nextstrain.org/) shown in Fig. 1. The data presented in Table 2 also suggest that G614 variant of SARS-CoV-2 has been circulating earlier in Pahang, Selangor, Kuala Lumpur and Sarawak. This study also reported that of 250 high coverage complete genome hCoV-19/Malaysia strains deposited in the GISAID database, 114 of the virus strains harboured D614G mutations in Spike protein (Table 3). Altogether, these results suggest the SARS-CoV-2 Malaysia isolate was subjected to intense positive selection pressure [29] and a persistent D614G mutation identified may be responsible for the quick spread of SARS-CoV-2 in Malaysia.

In this study, we reported that the D614G mutated SARS-CoV-2 of Pahang isolate-Pahang/IIUM91earlier collected from an asymptomatic patient on April 2, 2020 (highlighted in grey (Table 2)) has been deposited in GenBank and GISAID database with the accession number MW079428 and the GISAID EpiCoV EPI_ISL_455313, respectively. The accession numbers for the Illumina iSeq 100 sequence raw reads in the NCBI Sequence Read Archive (SRA) were PRJNA667798 (BioProject), SRX9252202 (SRA), & SAMN16383835 (BioSample).

Next, we investigated our Pahang/IIUM91 G614 variant Spike protein's phylogenetic analysis against other G614 variants of Spike protein of Malaysia strains (Fig. 2). To do this, of the 114 of

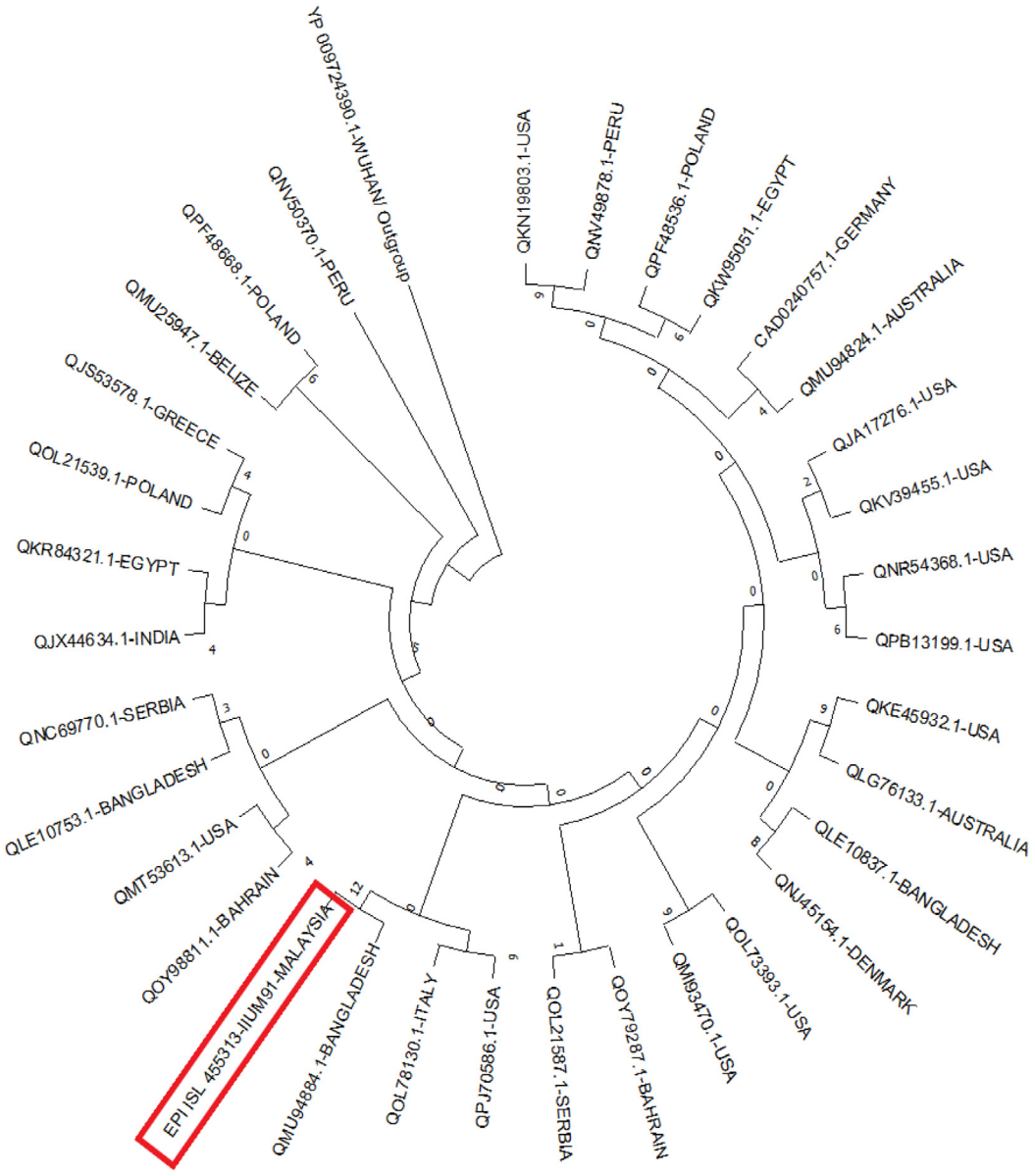

Fig. 3. Neighbor-Joining phylogenetic tree of G614 variant Spike protein built from Protein BLAST searches of Pahang/IIUM91 (EPI ISL 455313-IIUM91-MALAYSIA) Spike amino acid sequence. A maximum of ten amino acid sequences of Spike protein was selected as representative for each country. NCBI reference sequence: YP_009724390.1, D614 variant Spike protein from WuHan-Hu-1 genome (NC_045512.2) was used as outgroup. Pahang/IIUM91 (EPI ISL 455313-IIUM91-MALAYSIA) was labelled in a red box. The tree was rooted to the YP_009724390.1. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

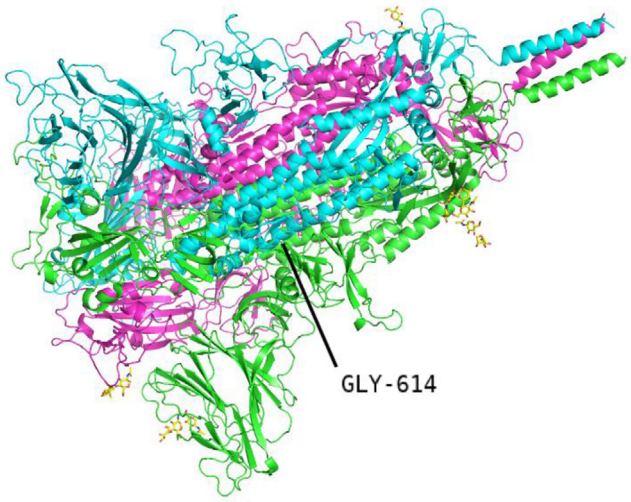

Fig. 4. 3D structure model of G614 variant Pahang/IIUM91Spike protein. The model was coloured by the chain and built with ProMod3 3.2.0 using SWISS-MODEL server. Chain A = Green, Chain B = Light blue, Chain C = Purple. The location of G614 residues was shown as Gly-614. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

G614 variants, 20 earliest virus strains from each lineage found in each state (Table 3) were selected for analysis. Our phylogenetic analysis of Spike protein of G614 variant shows Pahang/IIUM91 (a red box) sampled on April 2, 2020, from lineage B.1.247 was closely related to Selangor /IMR_WC55122 sampled on May 10, 2020, from lineage B.1.1.1 (Fig. 2). While the G614 variant from Sivagangga cluster occurred in Kedah was reported to emerge in Malaysia in July 2020 [28,30], the data deposited in the GISAID suggest that the possible virus responsible for spreading from this cluster could be either; IMR-WI109, sampled on July 25, 2020 (a yellow box) from lineage B.1.1.63 or IMR-WI085, sampled on July 27, 2020 (a yellow box) from lineage B.1.1. Hence, suggesting that Pahang/IIUM91, G614 variant Spike protein, was distantly correlated to these possible Sivagangga cluster's virus strains.

This result raised a question of where this Pahang/IIUM91, G614 variant Spike protein possibly originated from. To do this, we uploaded the Spike protein sequence of Pahang/IIUM91 onto the Protein BLAST (blastP) database, and the protein sequence was blasted using default parameters. This analysis result 100 sequences of Spike protein of SARS-CoV-2, with a percentage identity range of 99.90% to 100%, and a query coverage score of 99%. Of this, 100 Spike protein sequences, we select a maximum of 10 Spike protein sequences as representatives from each country. Our phylogenetic analysis of G614 Spike protein indicates that a target of Pahang/IIUM91 (EPI ISL 455313-IIUM91-MALAYSIA, in a red box), was closely related to QMU 94884.1-BANGLADESH (Fig. 3). Therefore, suggesting Pahang/IIUM91 was not correlated to Sivagangga cluster, which originated from India [29].

3D structure model of Pahang/IIUM91 Spike protein obtained from SWISS-MODEL server shown in Fig. 4 will be used as a model to study the molecular docking of this mutated virus in another publication. The protein template 6xr8.1.A (distinct conformation states of SARS-CoV-2 Spike protein) [31] was selected for modelling protein with the sequence identity 99.92%, and Global Model Quality Estimation (GMQE) score 0.75. This 3D structure of Pahang/IIUM91-G614 variant Spike protein-to-6xr8. 1. A template covered the residues of 14-1162 (Fig. 5). The QMEAN score of the model is −1.26, and GMQE score is 0.71 (Table 4). Predicted protein–ligand binding residues of the predicted 3D structure of Pahang/IIUM91 were listed in Table 4. A close-up view of D614 residue of reference Spike protein, YP_009724390.1 and G614 resi-

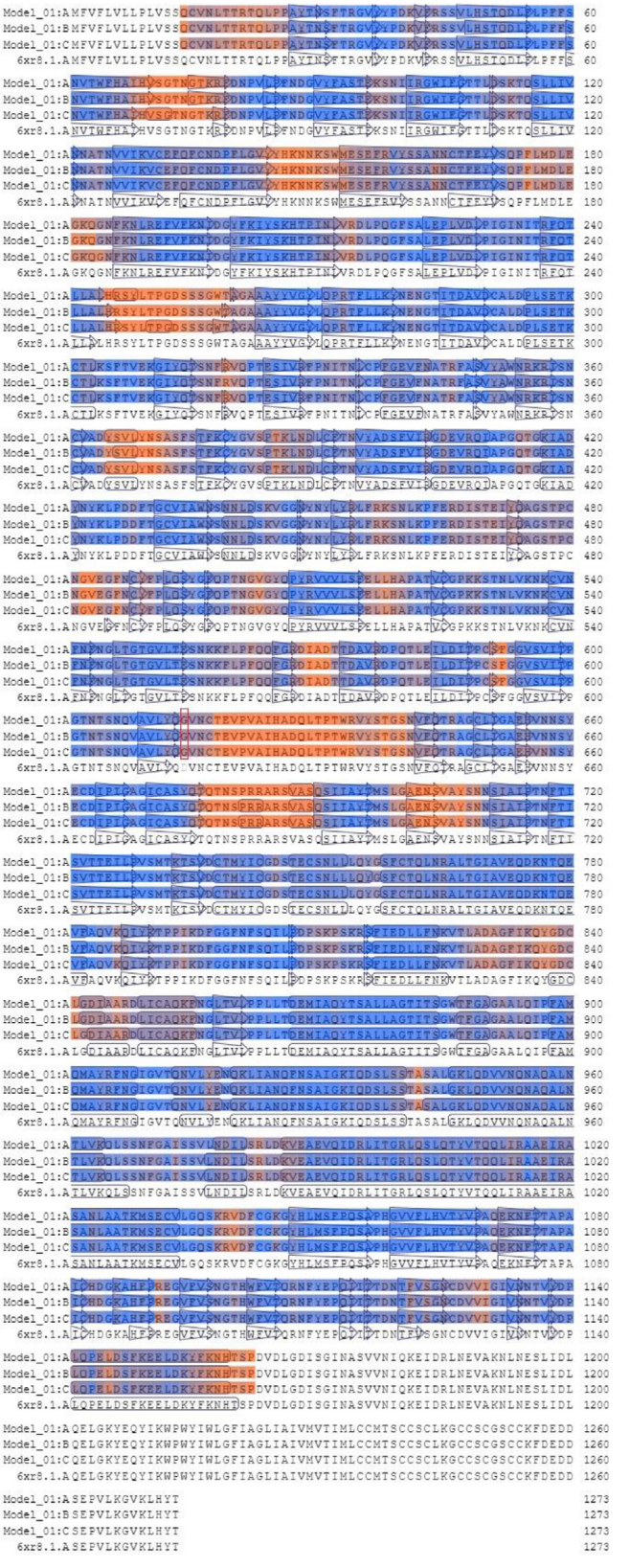

Fig. 5. Model-Template alignment coverage of Pahang/IIUM91 Spike protein-to-6xr8.1.A template generated using SWISS-MODEL. 3D structure model was built based on the residues 14-1162. Residues 1-13 and 1163-1273 were not included in the 3D structure due to lack of relevant template structures. The scheme colours indicate the quality estimation QMEAN Z- score. The blue specifies that the protein–protein interface was modelled with confidence, while orange specifies that the protein–protein interface was modelled with less confidence [26]. The red box indicates the location of the G614 residue. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 4
Summary of 3D structure model of Pahang/IIUM91Spike protein.

Oligo-State	Ligands	GMQE	QMEAN
Homo-trimer (matching prediction	1 × NAG: 2-acetamido-2-deoxy-beta-D-glucopyranose; 1 × NAG-NAG: 2-acetamido-2-deoxy-beta-D-glucopyranose-(1-4)-2-acetamido-2-deoxy-beta-D-glucopyranose; 1 × NAG-NAG-FUC: 2-acetamido-2-deoxy-beta-D-glucopyranose-(1-4)-[alpha-L-fucopyranose-(1-6)]2-acetamido-2-deoxy-beta-D-glucopyranose; 2 × NAG-NAG-MAN: alpha-D-mannopyranose-(1-4)-2-acetamido-2-deoxy-beta-D-glucopyranose-(1-4)-2-acetamido-2-deoxy-beta-D-glucopyranose;	0.71	−1.26

Note: The QMEAN Z-score above −4.0 is described as good 3D structure model [26]. The resulting GMQE score is expressed as a number between 0 and 1, reflecting the expected accuracy of a model built with that alignment and template, normalised by the target sequence's coverage. Higher numbers indicate higher reliability.

due of Pahang/IIUM91 Spike protein visualised in Fig. 6 showed interatomic interactions of wild type D614 and G614 mutant residues. An earlier study showed that the substitution of Asp614 with glycine changes hydrogen bonding around residue 614, as the Asp614-Thr859 hydrogen bond was eliminated while interaction with intradomain Ala647 was strengthened [32]. The change from D614 to G614 (Fig. 6) was previously reported not to cause any large structural rearrangement except for the loss of D614-K854 salt bridge in the fusion peptide proximal region (FPPR) [30,32]. Insight into the mechanism by which D614G increases infectivity, D614G mutation had increased both the stability (Table 5) and molecule flexibility (Table 6) of Pahang/IIUM91 Spike protein. Zhang et al. [33] suggest the virus with G614 variant has more stable G614 trimmer, as the receptor-binding domain (RBD) - down conformation is reinforced by both newly identified 630 loops and FPPR, consequently raising the barrier for the closed-to-open transition of the RBD. On top of that, disruption of the interprotomer latch between D614 in S1 and T859 in S2 due to D614G mutation results in increased distance between the promoters and a dramatic flip in the ratio of open to closed Spike protein particle, thus more open confirmation of its RBD [31]. Gained in molecule flexibility due to D614G mutation in Pahang/IIUM91 may increase protein thermostability, enabling the mutated virus to absorb more heat for the same increase in temperature than

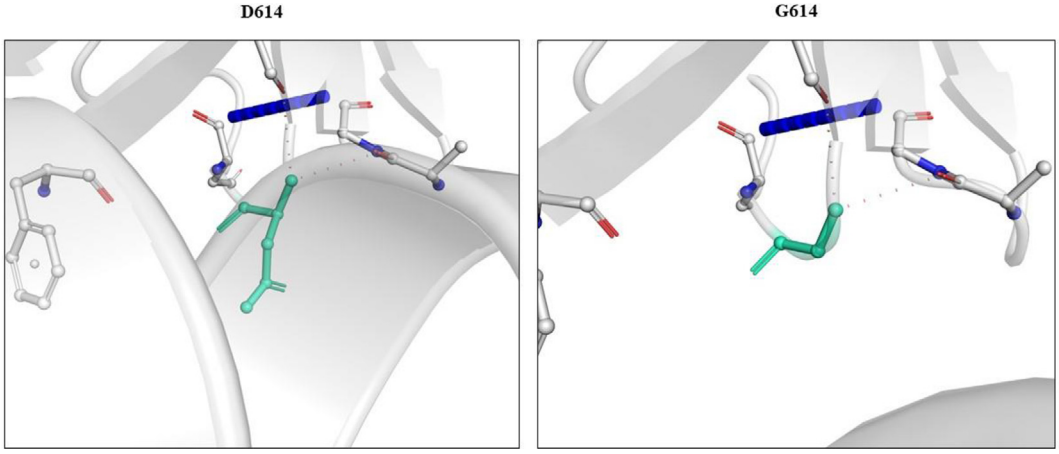

Fig. 6. A close-up view of D614 residue of reference Spike protein, YP_009724390.1 and G614 residue of Pahang/IIUM91 Spike protein. Interatomic interactions of wild type and mutant residues were coloured in light green and represented as sticks alongside the surrounding residues involved in any interaction. The models were visualized using DynaMut. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 5
The change in protein stability ΔΔG (kcal/mol) due to D614G mutation in Spike Protein.

Virus	Change in protein stability, ΔΔG (kcal/mol) by DynaMut	Stabilising/Destabilising	ΔΔG (kcal/mol) by other structure-based methods	Stabilising/Destabilising
Pahang/IIUM91	0.28	Stabilizing	mCSM: −0.21 SDM: 2.330 DUET: 0.475	Destabilizing Stabilising Stabilising

Note: ΔΔG ≥ 0 as stabilizing and ΔΔG less than 0 as destabilizing [27]. The analysis was performed using DynaMut server.

Table 6
The effect of D614G on the entropy energy between wild-type and mutant structures (in kcal/mol/K) of Spike protein.

Virus	Δ Vibrational entropy energy between wild-type and mutant, ΔΔSVib ENCoM, (kcal.mol-1.K-1)	Molecule flexibility (Increase/ Decrease)
Pahang/IIUM91	0.065	Increase

*The analysis was performed using DynaMut server.

wild-type virus [13,34]. These conformational changes in G614 trimmer, rendered virus with D614G mutating to be more immunogenic than wild-type D614 virus [31]. This mutational feature may substantially contribute to understanding the variability of COVID-19 susceptibility, severity, and outcomes in the population [35].

4. Conclusion

In conclusion, increases trend of positive COVID-19 in Malaysia may be contributed by major SARS-CoV-2 lineage B.5 which harbour D614G mutation in Spike protein. Here we report that COVID-19 with D614G mutation has been circulating in our society earlier than the case reported by MOH. Establishment of complete SARS-CoV-2 strain genome database in Malaysia is critical to proceed. This genome database would be beneficial at baseline to improve diagnosis and vaccine development treatment in the near future. We also reported that D614G mutated Pahang/IIUM91 virus was circulating in Pahang since April 2020. This virus was not related to the mutant D614G virus introduced by Sivagangga cluster. The 3D structure model of Pahang/IIUM91 a G614 variant will be used as a model for future analysis, particularly on the vaccine effectiveness study and potential phytochemicals. D614G mutation in Spike protein enabled Pahang/IIUM91 to increase its stability and fitness, thus contributing to a massive increase in COVID-19 positive cases detected in Pahang.

CRediT authorship contribution statement

Aini Syahida Mat Yassim: Data curation, Formal analysis, Visualization, Writing - original draft, Writing - review & editing. **Mohd Fazli Farida Asras:** Writing - original draft, Writing - review & editing. **Ahmad Mahfuz Gazali:** Writing - original draft, Writing - review & editing. **Martin S. Marcial-Coba:** Writing - original draft, Writing - review & editing. **Ummu Afeera Zainulabid:** Writing - original draft, Writing - review & editing. **Hajar Fauzan Ahmad:** Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing - original draft, Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We humbly acknowledge the authors of GISAID database and the COVID-19 Taskforces from UMP and IIUM. We also thank the Universiti Malaysia Pahang, Malaysia and Ministry of Higher Education Malaysia for supporting this work *via* RDU190364 and FRGS/1/2019/WAB13/UMP/03/1, respectively.

References

[1] World Health Organization, WHO Coronavirus Disease (COVID-19) Dashboard, https://covid19.who.int/ (accessed January 16, 2020).
[2] World Health Organization, WHO (Representative Office for Malaysia, Brunei Darussalam and Singapore): COVID-19 in Malaysia Situation Report 01, https://www.who.int/malaysia/internal-publications-detail/covid-19-in-malaysia-situation-report-01 (accessed January 14, 2021).
[3] Ministry of Health (Malaysia), COVID-19 Malaysia, http://covid-19.moh.gov.my/ (accessed January 16, 2021).
[4] S. Elbe, G. Buckland-Merrett, Data, disease and diplomacy: GISAID's innovative contribution to global health, Global Challenges 1 (1) (2017) 33–46.
[5] Y. Shu, J. McCauley, GISAID: global initiative on sharing all influenza data–from vision to reality, Eurosurveillance 22 (13) (2017) 30494.
[6] A. Rambaut et al., A dynamic nomenclature proposal for SARS-CoV-2 to assist genomic epidemiology, Nat. Microbiol. 5 (2020) 1403–1407.
[7] Y.M. Chong et al., SARS-CoV-2 lineage B. 6 is the major contributor to transmission in Malaysia, PLoS Negl. Trop. Dis. 14 (11) (2020) e0008744.
[8] Y.M. Chong et al., Complete genome sequences of SARS-CoV-2 strains detected in Malaysia, Microbiol. Resour. Announcements 9 (20) (2020) e00383–e420.
[9] S. Duchene, L. Featherstone, M. Haritopoulou-Sinanidou, A. Rambaut, P. Lemey, G. Baele, Temporal signal and the phylodynamic threshold of SARS-CoV-2, Virus Evol. 6 (2) (2020) veaa061.
[10] BERNAMA, COVID-19 D614G mutation found in five clusters - Health DG, in: BERNAMA.com, September 25 ed, 2020.
[11] B. Korber et al., Tracking changes in SARS-CoV-2 Spike: evidence that D614G increases infectivity of the COVID-19 virus, Cell 182(4) (2020) 812–827. e19.
[12] L. Zhang et al., SARS-CoV-2 spike-protein D614G mutation increases virion spike density and infectivity, Nat. Commun. 11 (1) (2020) 1–9.
[13] J.A. Plante et al., Spike mutation D614G alters SARS-CoV-2 fitness, Nature (2020) 1–6.
[14] D.C. Groves, S.L. Rowland-Jones, A. Angyal, The D614G mutations in the SARS-CoV-2 spike protein: implications for viral infectivity, disease severity and vaccine design, Biochem. Biophys. Res. Commun. S0006-291X(20) (2020) 32038-6.
[15] Q. Li et al., The impact of mutations in SARS-CoV-2 spike on viral infectivity and antigenicity, Cell 182(5) (2020) 1284–1294. e9.
[16] N. Müller et al., Viral genomes reveal patterns of the SARS-CoV-2 outbreak in Washington State, Medrxiv[Preprint] (2020).
[17] Q. Josh, nCoV-2019 sequencing protocol, https://www.protocols.io/view/ncov-2019-sequencing-protocol-bbmuik6w (accessed August 1, 2020).
[18] E.S. Wright, Using DECIPHER v2. 0 to analyse big biological sequence data in R, R J. 8 (1) (2016) 352–359.
[19] D. Charif, J.R. Lobry, SeqinR 1.0-2: a contributed package to the R project for statistical computing devoted to biological sequences retrieval and analysis, in: U. Bastolla, M. Porto, H.E. Roman, M. Vendruscolo (Eds.), Biological and Medical Physics, Biomedical Engineering, Springer, Berlin, Heidelberg, 2007, ch. Structural Approaches to Sequence Evolution, pp. 207–232.
[20] S. Kumar, G. Stecher, M. Li, C. Knyaz, K. Tamura, MEGA X: molecular evolutionary genetics analysis across computing platforms, Mol. Biol. Evol. 35 (6) (2018) 1547–1549.
[21] N. Saitou, M. Nei, The neighbor-joining method: a new method for reconstructing phylogenetic trees, Mol. Biol. Evol. 4 (4) (1987) 406–425.
[22] J. Felsenstein, Confidence limits on phylogenies: an approach using the bootstrap, Evolution 39 (4) (1985) 783–791.
[23] D.T. Jones, W.R. Taylor, J.M. Thornton, The rapid generation of mutation data matrices from protein sequences, Bioinformatics 8 (3) (1992) 275–282.
[24] T. Schwede, J. Kopp, N. Guex, M.C. Peitsch, SWISS-MODEL: an automated protein homology-modeling server, Nucleic Acids Res. 31 (13) (2003) 3381–3385.
[25] P. Benkert, M. Biasini, T. Schwede, Toward the estimation of the absolute quality of individual protein structure models, Bioinformatics 27 (3) (2011) 343–350.
[26] P. Benkert, S.C. Tosatto, D. Schomburg, QMEAN: a comprehensive scoring function for model quality assessment, Proteins Struct. Func. Bioinformat. 71 (1) (2008) 261–277.
[27] C.H. Rodrigues, D.E. Pires, D.B. Ascher, DynaMut: predicting the impact of mutations on protein conformation, flexibility and stability, Nucleic Acids Res. 46 (W1) (2018) W350–W355.
[28] C. Clarissa, Mutated strain more infectious, in: The Star, October 3 ed, 2020.
[29] X.-Y. Zhan et al., Molecular evolution of SARS-CoV-2 structural genes: Evidence of positive selection in spike glycoprotein, BioRxiv (2020).
[30] F.M.T. Reporters, Super-spreader strain found in Sabah, Kedah Covid-19 clusters, in: Free Malaysia Today, September 25 ed, 2020.
[31] Y. Cai et al., Distinct conformational states of SARS-CoV-2 spike protein, BioRxiv (2020).
[32] L. Yurkovetskiy et al., Structural and functional analysis of the D614G SARS-CoV-2 spike protein variant, Cell 183(3) (2020) 739–751. e8.
[33] J. Zhang et al., Structural impact on SARS-CoV-2 spike protein by D614G substitution, BioRxiv (2020).
[34] A. Karshikoff, L. Nilsson, R. Ladenstein, Rigidity versus flexibility: the dilemma of understanding protein thermal stability, FEBS J. 282 (20) (2015) 3899–3917.
[35] C.-H.-G. Initiative, The COVID-19 Host Genetics Initiative, a global initiative to elucidate the role of host genetic factors in susceptibility and severity of the SARS-CoV-2 virus pandemic, Eur. J. Hum. Genet. (2020) 1.