

Genomic analysis of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) strains isolated in Malaysia

Hooi-Leng Ser^{1†*}, Loh Teng-Hern Tan^{1†}, Jodi Woan-Fei Law¹, Vengadesh Letchumanan¹, Nurul-Syakima Ab Mutalib², Learn-Han Lee^{1*}

¹Novel Bacteria and Drug Discovery (NBDD) Research Group, Microbiome and Bioresource Research Strength, Jeffrey Cheah School of Medicine and Health Sciences, Monash University Malaysia, 47500 Bandar Sunway, Selangor Darul Ehsan, Malaysia

²UKM Medical Molecular Biology Institute (UMBI), UKM Medical Centre, Universiti Kebangsaan Malaysia, Kuala Lumpur, Malaysia

Abstract: Using seven complete genomes of human SARS-CoV-2 (retrieved from GISAID) isolated in Malaysia for phylogenetic tree construction, the current study showed that these strains formed four distinct clades when compared with other representative strains from Asia, Europe and US. In light of that, the genome sequences of these strains isolated in Malaysia suggested that there is currently more than one “type” of strain within the country. Complementing with epidemiological and experimental studies, these findings allow better understanding the prevalence of certain types in Malaysia and permits further in-depth studies on the virulence and pathogenic mechanisms of these strains which is particularly critical to speed up the development of effective treatment regime.

Keywords: SARS-CoV-2; Malaysia; genome; phylogenetic analysis; SNVs; mutations

Received: 4th May 2020

Accepted: 4th June 2020

Published Online: 18th June 2020

[†]These authors contributed equally in the writing.

***Correspondence:** Hooi-Leng Ser, Jeffrey Cheah School of Medicine and Health Sciences, Monash University Malaysia, 47500 Bandar Sunway, Selangor Darul Ehsan, Malaysia; ser.hooileng@monash.edu. Learn-Han Lee, Jeffrey Cheah School of Medicine and Health Sciences, Monash University Malaysia, 47500 Bandar Sunway, Selangor Darul Ehsan, Malaysia; lee.learn.han@monash.edu.

Citation: Ser H-L, Tan LT-H, Law JW-F, *et al.* Genomic analysis of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) strains isolated in Malaysia. *Prog Microbes Mol Biol* 2020; 3(1): a0000093. <https://doi.org/10.3687/pmbb.a0000093>.

Introduction

Taking a closer look at coronavirus, this notorious bug has caused several global outbreaks throughout human history, with notable instances such as 1918 (H1N1) influenza pandemic, and more recently Middle East Respiratory Syndrome (MERS) caused by MERS-CoV in 2012. Fast forward to the last month of 2019, the WHO China Country Office was informed of pneumonia cases with unknown etiology, which later on determined a new type of coronavirus known as SARS-CoV-2^[1]. As of 30th April 2020, the World Health Organization reported that there are more than 3 million confirmed COVID-19 cases globally and has impacted more than 210,000 COVID-related deaths^[2-6]. At the time of writing, Malaysia's government announced to ease the “partial lockdown” of more than six weeks, allowing almost all economic sectors to reopen, in parallel with the decreasing trend of COVID-19 confirmed cases during the end month of April 2020. In the South East Asia region, Malaysia was among the first few countries to implement the Movement Control Order (MCO) to curb the spread of the coronavirus on

18th March 2020 (Figure 1).

The first case of COVID-19 in Malaysia was detected on 24th January 2020^[7]. The first case involved a Chinese national from Wuhan, who had travelled from Singapore to Johor Bahru in a group of eight for holiday on 22nd January 2020. They were quarantined at a hotel on the following days after coming into contact with an infected patient in Singapore. Since then, there were around 22 positive COVID-19 cases as the first wave of outbreak cases in the country, of which all the patients were discharged upon recovery^[8]. However, a gradual increase in positive cases begun on 27th February before a sudden surge was observed on 15th March 2020 reaching as high as 428 cases after 11 days of zero reported case (i.e. from 16th to 26th February 2020). These cases were regarded as the second wave of outbreak and attributed to a religious gathering event which was attended by more than ten-thousands of people^[8,9]. As a consequence, the shift and firm response from the government into the MCO implementation has been critical to limiting the spread of COVID-19.

To gain further understanding on the molecular epidemiology of the COVID-19 outbreak in Malaysia, we used seven publicly available whole genome sequences sampled up 14th April 2020 to analyze the phylogenetic evolution of the SARS-CoV-2 strains isolated from the patients in Malaysia with worldwide SARS-CoV-2 genome sequences. Comparisons of the single nucleotide variants (SNV) are often used for evolutionary studies as

the virus genome is subjected to frequent mutations due to an error-prone RNA-dependent RNA polymerase responsible for the virus genome replication. These mutations might be associated with the changes in transmissibility and virulence of the virus. Thus, we also performed the SNV analysis to investigate the genotype changes during the transmission of SARS-CoV-2 in the country.

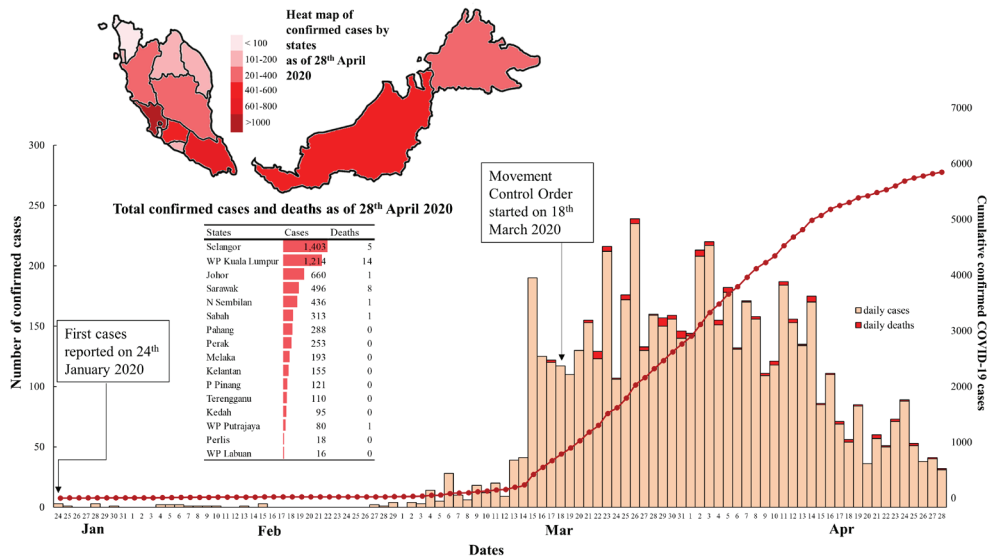


Figure 1. The daily confirmed COVID-19 cases and deaths reported in Malaysia up to 28th April 2020 based on statistics obtained from Ministry of Health Malaysia. The stacked bar graph shows the daily confirmed cases and deaths since the first case reported in Malaysia while the line graph shows the cumulative confirmed cases. A heat map shows the distribution of confirmed cases from different states in Malaysia.

Materials and methods

Phylogenetic and SNV analysis using complete genome of SARS-CoV-2

Complete genome sequences of SARS-CoV-2 were retrieved from GISAID (<https://www.gisaid.org>) and NCBI database^[10,11]. NC_045512 genome sequence fetched from NCBI was used for reference and genomic coordinates in this study are based on this reference genome. The alignment of sequences were formed via ClustalX software, verified manually and adjusted prior to the reconstruction of phylogenetic trees^[12]. Phylogenetic trees were constructed with the maximum-likelihood algorithm (Figure 2) using Molecular Evolutionary Genetics Analysis across Computing Platforms (MEGA) version 7.0^[13–15]. Support for the tree topology was estimated with 1,000 bootstrap replicates. SNV composition of the selected strains based on the phylogenetic tree was analyzed and compared with the reference sequence NC_045512 using Bioedit and 2019 Novel Coronavirus Resource (2019nCoV, <https://bigd.big.ac.cn/ncov>)^[16].

Data availability statement

All data are available in the main text and the supplementary materials. The genome sequences used in the current study is retrieved from GISAID (<https://www.gisaid.org/>) and NCBI databases (https://www.ncbi.nlm.nih.gov/nucleotide/NC_045512).

Results

Phylogenetic analysis using maximum likelihood algorithm

In order to observe the genomic relationship between SARS-CoV-2 strains isolated in Malaysia, a dataset of 7 publicly available complete genomes of SARS-CoV-2 from different countries was retrieved from GISAID (<https://www.gisaid.org>, accessed on 14th April 2020; Supp Table 1)^[10,11]. A phylogenetic tree which was constructed using the maximum-likelihood algorithm revealed that the seven strains isolated in Malaysia formed four major clades (Figure 2). Majority of the strains ($n = 3$) showed that they were closely related, forming Clade III with the closest related strain isolated from Singapore (GISAID accession ID: EPI_ISL_407987). On the other hand, another strain isolated from a 45-year old male patient in Clade I (GISAID accession ID: EPI_ISL_416886) displayed a close evolutionary relationship with strain from US (GISAID accession ID: EPI_ISL_404895) and China (Fujian, GISAID accession ID: EPI_ISL_411060). The fourth clade (i.e. Clade IV) which consist of strain isolated from a 67-year old female patient (GISAID accession ID: EPI_ISL_416907) was closely related to another strains from Asia region including China (GISAID accession ID: EPI_ISL_408486, EPI_ISL_403930), Japan (GISAID accession ID: EPI_ISL_410532, EPI_ISL_413459, EPI_ISL_412969) and Singapore (GISAID accession ID: EPI_ISL_406973). In addition, the other two strains EPI_ISL_417917 and EPI_ISL_417418 which were isolated later in late March formed a separate clade (Clade II).

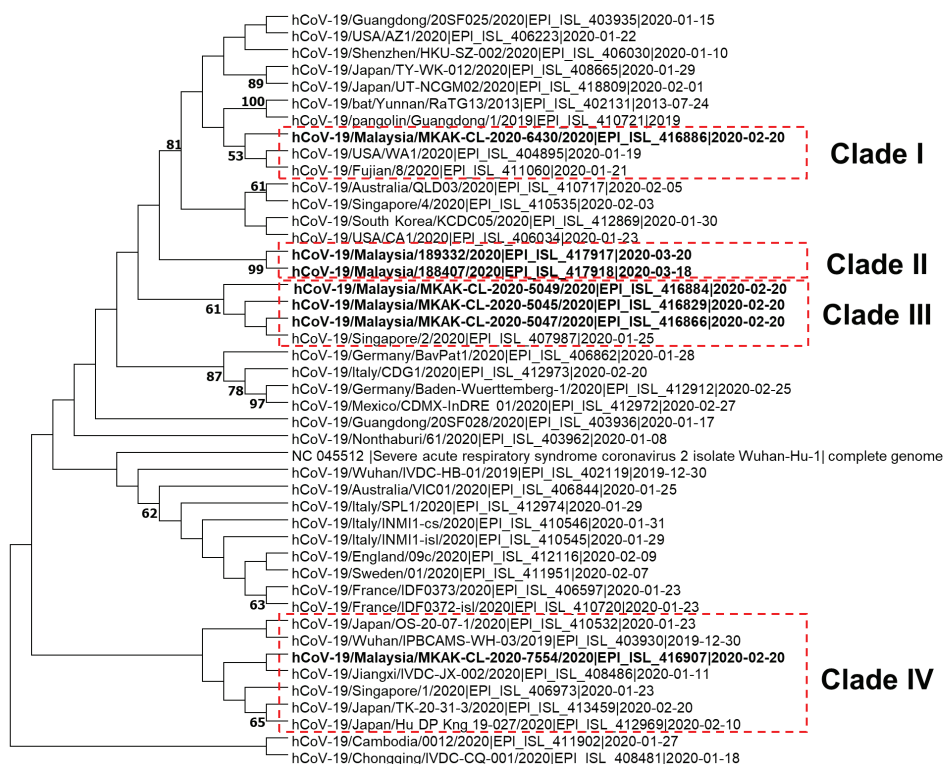


Figure 2. Phylogenetic analysis of forty-five SARS-CoV-2 complete genome sequences (~29880 nucleotides) showing relationship between seven Malaysian strains with representatives' complete sequences from different countries. It is a maximum-likelihood tree with bootstrap values (>50%) based on 1000 re-sampled datasets are shown at branch nodes.

Single nucleotide variants (SNVs) analysis

The genome-wide SNVs for selected strains are as reported in Table 1 using a strain isolated from China in December 2019 (NCBI accession number: NC_045512) as reference strain. A total of 30 SNVs was identified. EPI_ISL_417917 showed the highest number of SNVs, with 9 positions showing variance compared to the reference strain. Furthermore, one of the closely related strain isolated in Malaysia, EPI_ISL_417418 (based on phylogenetic analysis) revealed 8 SNVs; this strain shared 7 common SNVs with EPI_ISL_417917 with addition to a SNVs at 25473 nt (compared to NC_045512) where gene M is located. Additionally, all members of Clade III showed 27147G>C and one missense variation in gene M which encodes for matrix protein. Among the seven strains isolated in Malaysia, EPI_ISL_416866 displayed the lowest number of SNVs only one at 27147 nt (gene M). Similar pattern was observed with amino acid variations as shown in Table 2.

Discussion

The availability and data-sharing of SARS-CoV-2 whole genome sequences allow researchers to study the evolutionary relationships and patterns of molecular divergence between coronaviruses around the globe. In fact, nearly two-thirds of the viral genome falls within the first ORF (ORF1a/b) which translate to (non-structural) two polyproteins (pp1a and pp1ab), while the remaining genome encodes four essential structural proteins, including spike (S) glycoprotein, small

envelope (E) protein, matrix (M) protein, and nucleocapsid (N) protein^[17,18]. Along with these, there are a total of six accessory proteins coded by ORF3a, ORF6a, ORF7a, ORF7b, ORF8 and ORF10 genes located across the ~29kb genome of SARS-CoV-2. In the current analysis, a total of seven strains isolated in Malaysia was used to perform in the comparison along with representative strains from Asia, US and Europe (note: all sequences of this study were retrieved from NCBI or GISAID database). Most of these strains were closely related with strains isolated from Asia region based on Figure 2. However, some strains displayed more SNVs than the rest of the strains isolated in Malaysia (Figure 3). For instance, EPI_ISL_417917 and EPI_ISL_417918 formed Clade II; analyses on SNVs and amino acid variations reflected that these strains have more than 90% common nucleotide/amino acid in several genes. For these strains, most of the variations occurred within the ORF1ab genes including 6310C>A, 6312C>A, 11083G>T, 13730C>T, 19524C>T. These SNVs then corresponded for four missense variants and one synonymous variant for polyproteins 1a/1ab. In addition, there are two more variations occurring only in EPI_ISL_417917 (but not EPI_ISL_417918) at position 2737T>A (synonymous variant) and 13975G>A (missense variant). In addition to that, there is a SNV observed in gene N at 28311C>T in both strains which encodes for nucleocapsid protein (structural protein) that wraps the genomic RNA (gRNA) into a helical structure. One of the important points to note is that these two strains were obtained from patients in March 2020. As suggested by previous literature, RNA viruses like coronavirus can have high mutation rates and more often than not, these mutations are correlated with virulence modulation and

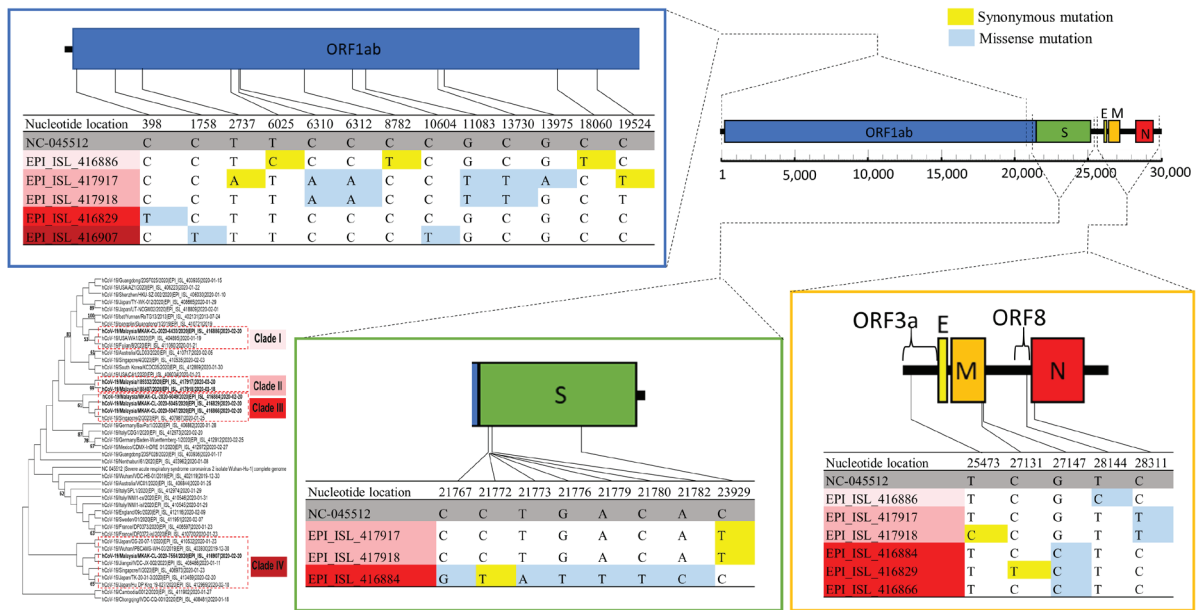


Figure 3. Single nucleotide variants detected from the seven Malaysia strains of SARS-CoV-2 in comparison to the reference strain (NC_045512).

evolvability, thereby improving their adaptation ability^[19-21].

Five of the seven strains, EPI_ISL_416886, EPI_ISL_416866, EPI_ISL_416884, EPI_ISL_416829, EPI_ISL_416907 (Clade I, III and IV) were isolated in late February in Malaysia. Interestingly, EPI_ISL_416886 reflected several SNVs within two well-studied genomic regions including ORF1ab (6025T>C, 8782C>T and 18060C>T) and ORF8 (28144T>C). As a matter of fact, EPI_ISL_416886 is the only unique strain isolated from Malaysia that exhibited co-variations at these two locations (i.e. 8782C>T and 28144T>C). As it has been noted that most of 8782C>T and 28144T>C variant sub-strains are found outside of Wuhan, one of the two closely related strains of EPI_ISL_416886 was identified to be strain from Fujian (EPI_ISL_411060) which also carried SNVs at these two locations. A simple pairwise comparison showed that EPI_ISL_416886 exhibited ~99.7 % similarity to closely related strains EPI_ISL_411060 and EPI_ISL_404895 which isolated in US. There have been ongoing discussions on the topic of characterizing SARS-CoV-2 strains based on variation at 8782 (C or T) and 28144 (T or C), either (a) to trace the routes of infections by observing their network nodes, or (b) to study the virulence of strains (i.e. S or L subtypes)^[22,23]. Having said that, the actual function of ORF8 in SARS-CoV-2 is yet to be discovered given that it lacks a known useful motif or region and seems to be highly divergent from ORF8b in SARS-CoV which induce intracellular stress pathways^[17,24,25]. In addition to this, EPI_ISL_416886 carried a SNV at position 18060 nucleotide belonging to ORF1ab (which encodes for nsp14). Even though Pachetti *et al.*^[20] reported that this SNV was mostly related to SARS-CoV-2 strain isolated from North America, further investigation into the travel history of the patient whom EPI_ISL_416886

was isolated from may be worthwhile to obtain a clearer understanding on how the strain acquired these SNVs, whether it's a spontaneous mutation and/or how fast before these SNVs emerged. Still and all, the changes at this position (i.e. 18060) from C>T resulted in a synonymous variant in amino acid sequence, thus the actual effect of this substitution is still pending to be discovered.

S glycoprotein encoded by gene S is one of the crucial factors in determining host range and pathogenicity of SARS-CoV-2. A study in early February by Zhou *et al.*^[26] has confirmed that SARS-CoV-2 shares similarity in the target receptor for cellular entry, recognized as angiotensin converting enzyme II (ACE2) receptors. In other words, the S glycoprotein of SARS-CoV-2 can attach to ACE2 receptors available on several types of human cells before hijacking the host machinery^[26-30]. From Table 1, a total of 9 SNVs was found to be associated with S glycoprotein (or gene S). Despite of that, three strains (EPI_ISL_416884, EPI_ISL_417917 and EPI_ISL_417918) isolated in Malaysia carried one or more SNVs within this gene. EPI_ISL_416884 (Clade III) displayed highest number of SNVs within gene S with 7 SNVs, and subsequently resulted in 1 synonymous and 5 missense (amino acid) variants in the S glycoprotein, while EPI_ISL_417917 and EPI_ISL_417918 only reflected one SNV (i.e. 23929C>T) without changes in the amino acid composition (i.e. synonymous variant). At the time of writing, these 7 SNVs found in EPI_ISL_416884 (21767C>G, 21772C>T, 21773T>A, 21776G>T, 21779A>T, 21780C>T, 21782A>C) seems to be novel¹⁷ and the true effect of these changes are still unknown. Taking note that these SNVs are mostly located near the N-terminal domain of the S1 subunit of S protein which is not directly involved with binding to ACE2, hence the variations may not be influencing/altering the binding ability of the virus^[31]. On the flip side, it is quite conservative or safe to mention that the actual functional impact of

NT	Gene	Ref	Clade I		Clade II		Clade III		Clade IV						Proportion of strains with SNVs			
			EPI_ ISL_416886	EPI_ ISL_404895	EPI_ ISL_411060	EPI_ ISL_417917	EPI_ ISL_417918	EPI_ ISL_416884	EPI_ ISL_416829	EPI_ ISL_416866	EPI_ ISL_407987	EPI_ ISL_410532	EPI_ ISL_403930	EPI_ ISL_416907		EPI_ ISL_408486	EPI_ ISL_406973	EPI_ ISL_413459
398		C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	1/16
1758		C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	1/16
2737		T	T	T	A	T	T	T	T	T	T	T	T	T	T	T	T	1/16
6025		T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	1/16
6310		C	C	C	A	A	A	C	C	C	C	C	C	C	C	C	C	2/16
6312		C	C	C	A	A	A	C	C	C	C	C	C	C	C	C	C	2/16
6996		T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	1/16
8274	OR-F1ab	G	G	G	G	G	G	G	G	G	G	G	G	A	G	G	G	1/16
8782		C	T	T	T	T	C	C	C	C	C	C	C	C	C	C	C	3/16
10604		C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	1/16
11083		G	G	G	T	T	G	G	G	G	G	G	G	G	G	T	T	4/16
13730		C	C	C	T	T	C	C	C	C	C	C	C	C	C	C	C	2/16
13975		G	G	G	A	G	G	G	G	G	G	G	G	G	G	G	G	1/16
18060		C	T	T	T	T	C	C	C	C	C	C	C	C	C	C	C	3/16
19524		C	C	C	T	T	C	C	C	C	C	C	C	C	C	C	C	2/16
21767		C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	1/16
21772		C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	1/16
21773		T	T	T	T	T	A	T	T	T	T	T	T	T	T	T	T	1/16
21776	S	G	G	G	G	G	T	G	G	G	G	G	G	G	G	G	G	1/16
21779		A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	1/16
21780		C	C	C	C	C	T	C	C	C	C	C	C	C	C	C	C	1/16
21782		A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	1/16
23929		C	C	C	C	C	T	C	C	C	C	C	C	C	C	C	C	2/16
25060		A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	1/16
25473	ORF3a	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	1/16
27131	M	C	C	C	C	C	T	C	C	C	C	C	C	C	C	C	C	1/16
27147		G	G	G	G	G	C	C	C	C	C	C	C	C	C	C	C	4/16
28144	ORF8	T	C	C	C	C	T	T	T	T	T	T	T	T	T	T	T	3/16
28311	N	C	C	C	C	C	T	C	C	C	C	C	C	C	C	C	C	2/16
29635	ORF10	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	T	2/16
Proportion of SNVs		-	4/30	3/30	3/30	8/30	9/30	8/30	8/30	1/30	1/30	1/30	1/30	1/30	1/30	2/30	2/30	

Table 1: Single nucleotide variations (SNVs) deduced by comparison of complete whole genome sequences of SARS-CoV-2 isolated in Malaysia and selected closely-related sequences (n = 16 sequences) using NC_045512 as reference genome (Blue: variant).

Aa	Gene	Ref	Clade I			Clade II			Clade III			Clade IV						Proportion of strains with variants	
			EPL ISL_416886	EPL ISL_404895	EPL ISL_411060	EPL ISL_417917	EPL ISL_417918	EPL ISL_416884	EPL ISL_416829	EPL ISL_416866	EPL ISL_407987	EPL ISL_410532	EPL ISL_403930	EPL ISL_416907	EPL ISL_408486	EPL ISL_406973	EPL ISL_413459		EPL ISL_412969
45	OR-F1ab	NC_044512	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	1/16	
498		A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	1/16
824		T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	1/16
1920		Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	1/16
2015		S	S	S	R	R	S	S	S	S	S	S	S	S	S	S	S	S	2/16
2016		T	T	T	K	K	T	T	T	T	T	T	T	T	T	T	T	T	2/16
2244		I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	1/16
2670		C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	1/16
2839		S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	3/16
3447		P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	1/16
3606		L	L	L	F	F	L	L	L	L	L	L	L	L	L	L	L	L	4/16
4489		A	A	A	V	V	A	A	A	A	A	A	A	A	A	A	A	A	2/16
4571		G	G	G	S	S	G	G	G	G	G	G	G	G	G	G	G	G	1/16
5932		L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	3/16
6420		L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	2/16
69		S	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	1/16
70	V		V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	V	1/16
71	S		S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	1/16
72	G		G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	1/16
73	T		T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	1/16
74	N		N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	1/16
789	Y		Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	2/16
1166	L		L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	1/16
27	ORF3a		D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	1/16
203	M		N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	1/16
209		D	D	D	D	D	H	H	H	H	D	D	D	D	D	D	D	4/16	
84	ORF8	L	S	S	S	L	L	L	L	L	L	L	L	L	L	L	L	3/16	
13	N	P	P	P	P	L	L	L	L	L	L	L	L	L	L	L	L	2/16	
26	ORF10	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	2/16	
Proportion of variants		0/29	4/29	3/29	3/29	9/29	8/29	7/29	3/29	1/29	3/29	0/29	1/29	2/29	1/29	2/29	2/29		

Table 2. Amino acid variations deduced by comparison of complete whole genome sequences of SARS-CoV-2 isolated in Malaysia and selected clade-related sequences (total n = 16 sequences) using NC_045512 as reference strain (Blue: missense; yellow: synonymous).

these variants remain unclear and further investigations are necessary to assess its importance, particularly for the development of vaccines^[25,32].

Above all, these findings of this study were consistent with reports from other research teams, with most of the strains showing high genomic similarities between strains derived from the Asia region. In the current study, regardless of sampling time (February or March), all the strains isolated in Malaysia reflected highly similar “pattern” of SNVs when compared to the other strains derived from Asia including China (EPI_ISL_411060, EPI_ISL_403930 and EPI_ISL_408486), Singapore (EPI_ISL_407987 and EPI_ISL_406973), and Japan (EPI_ISL_410532, EPI_ISL_413459 and EPI_ISL_412969). When compared within the seven strains isolated in Malaysia, members of Clade II EPI_ISL_417917 and EPI_ISL_417918 were shown to be more closely related to each other (<0.15% difference by pairwise comparison) compared to the other five strains (Supp Table 2). A recently published study by Forster *et al.*^[22] reconstructing evolutionary paths of SARS-CoV-2 using 160 complete human SARS-CoV-2 and found three central variants distinguished by amino acid changes which was named A, B, and C, with A being the ancestral type according to the bat outgroup coronavirus. Based on this finding, two members of Clade I (EPI_ISL_404895 and EPI_ISL_411060) in this study were determined to be A type, while most of the members of Clade III and IV were determined as derived-B type which as suggested by its name derived from A type. Forster *et al.*^[22] mentioned that SARS-CoV-2 would first mutate into derived B type (from A type), before finally turning into B type which is the most common type in East Asia. Altogether, these findings may indicate that there are currently more than one subtype present in Malaysia. Nonetheless, there is still much more to be explored, particularly on the transmittance of SARS-CoV-2: how these variations/mutations emerged over time in Malaysia, and then how these changes caused to the behavior of the virus (e.g. on their pathogenicity and virulence). Given the recent emergence of SARS-CoV-2 around the world, researchers have witnessed the growing number of genome sequences deposited in publicly available database like GISAID and NCBI. The policy on data sharing can potentially hasten the identification of COVID-19 infection sources while expediting the drug designing process and vaccine(s) development based on the availability of these complete viral genomes derived from different parts of the world^[33]. All and all, by garnering a more thorough perspective on COVID-19 infection sources, the current study could serve as the launching pad to inspect and understand the dynamics of the local transmission of SARS-CoV-2 in Malaysia. Complementing with epidemiological studies, these findings could essentially gather valuable information on the prevalence of certain strains in Malaysia, whereas in-depth experimental research would then

shed some light on the virulence and pathogenic mechanisms of these strains.

Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Author Contributions

The experiment and data analysis were performed by H-LS, LT-HT, L-HL, the manuscript was written and proof by H-LS, LT-HT, JW-FL, VL, N-SAM and L-HL. The project was founded by H-LS and L-HL.

Acknowledgments

We sincerely appreciate the researchers worldwide who sequenced and shared the complete genome data of SARS-CoV-2 from GISAID (<https://www.gisaid.org>) and NCBI databases (https://www.ncbi.nlm.nih.gov/nuccore/NC_045512). This research is dependent on these precious data.

References

1. World Health Organization. WHO Director-General's opening remarks at the media briefing on COVID-19 - 13 April 2020. [Internet]. 2020 [cited 2020 May 8]; Available from: <https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19--13-april-2020>.
2. World Health Organization. Coronavirus disease 2019 (covid-19). Situation report 101. [Internet]. 2020 [cited 2020 May 8]; Available from: <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports>.
3. European Centre for Disease Prevention and Control. Situation update for the EU/EEA and the UK, as of 15 April 2020. [Internet]. 2020 [cited 2020 May 8]; Available from: <https://www.ecdc.europa.eu/en/cases-2019-ncov-eueea>.
4. Letchumanan V, Ab Mutalib NS, Goh BH, *et al.* Novel coronavirus 2019-nCoV: Could this virus become a possible global pandemic. *Prog Microbes Mol Biol* 2020; 3(1): a0000068.
5. Tan LT, Letchumanan V, Ser HL, *et al.* PMMB COVID-19 Bulletin: United Kingdom (22nd April 2020). *Prog Microbes Mol Biol* 2020; 3(1).
6. Ser HL, Letchumanan V, Law JW, *et al.* PMMB COVID-19 Bulletin: Spain (18th April 2020). *Prog Microbes Mol Biol* 2020; 3(1).
7. World Health Organization. Coronavirus disease (COVID-19) in Malaysia. [Internet]. 2020 [cited 2020 May 8]; Available from: <https://www.who.int/malaysia/emergencies/coronavirus-disease-covid-19-in-malaysia>.
8. Khor V, Arunasalam A, Azli S, *et al.* Experience from Malaysia During the COVID-19 Movement Control Order. *Urol* 2020.
9. Director General of Health Malaysia. From the Desk of the Director-General of Health Malaysia. [Internet]. 2020 [cited 2020 May 8]; Available from: <https://kpksehatan.com/>.
10. Elbe S and Buckland-Merrett G. Data, disease and diplomacy: GISAID's innovative contribution to global health. *Glob Chall* 2017; 1(1): 33–46.
11. Shu Y and McCauley J. GISAID: Global initiative on sharing all influenza data—from vision to reality. *Eurosurv* 2017; 22(13).
12. Thompson JD, Gibson TJ, Plewniak F, *et al.* The CLUSTAL_X windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nuc Acids Res* 1997; 25(24): 4876–4882.
13. Felsenstein J. Evolutionary trees from DNA sequences: A maximum likelihood approach. *J Mol Evol* 1981; 17(6): 368–376.
14. Tamura K, Peterson D, Peterson N, *et al.* MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* 2011; 28(10): 2731–2739.
15. Kumar S, Stecher G, Tamura K. MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol* 2016; 33(7): 1870–1874.

16. Zhao WM, Song SH, Chen ML, *et al.* The 2019 novel coronavirus resource. *Yi chuan= Hereditas.* 2020; 42(2): 212–221.
17. Khailany RA, Safdar M, Ozaslan M. Genomic characterization of a novel SARS-CoV-2. *Gene Rep* 2020; 100682.
18. Zhou G and Zhao Q. Perspectives on therapeutic neutralizing antibodies against the Novel Coronavirus SARS-CoV-2. *Int J Biol Sci* 2020; 16(10): 1718.
19. Duffy S. Why are RNA virus mutation rates so damn high? *PLoS Biol* 2018; 16(8): e3000003.
20. Pachetti M, Marini B, Benedetti F, *et al.* Emerging SARS-CoV-2 mutation hot spots include a novel RNA-dependent-RNA polymerase variant. *J Transl Med* 2020; 18: 1–9.
21. Wang C, Liu Z, Chen Z, *et al.* The establishment of reference sequence for SARS-CoV-2 and variation analysis. *J Med Virol* 2020; 92(6): 667–674.
22. Forster P, Forster L, Renfrew C, *et al.* Phylogenetic network analysis of SARS-CoV-2 genomes. *PNAS* 2020; 117(17): 9241–9243.
23. Tang X, Wu C, Li X, *et al.* On the origin and continuing evolution of SARS-CoV-2. *Natl Sci Rev* 2020; 1–12.
24. Shi CS, Nabar NR, Huang NN, *et al.* SARS-Coronavirus Open Reading Frame-8b triggers intracellular stress pathways and activates NLRP3 inflammasomes. *Cell Death Dis* 2019; 5(1): 1–2.
25. Yuen KS, Ye ZW, Fung SY, *et al.* SARS-CoV-2 and COVID-19: The most important research questions. *Cell Biosci* 2020; 10(1):1–5.
26. Zhou P, Yang XL, Wang XG, *et al.* A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* 2020; 579(7798): 270–273.
27. Tortorici MA and Veesler D. Structural insights into coronavirus entry. *Adv Virus Res* 2019; 105: 93–116.
28. South AM, Diz DI, Chappell MC. COVID-19, ACE2, and the cardiovascular consequences. *Am J Physiol* 2020; 318(5): H1084–1090.
29. Xu H, Zhong L, Deng J, *et al.* High expression of ACE2 receptor of 2019-nCoV on the epithelial cells of oral mucosa. *Int J Oral Sci* 2020; 12(1): 1–5.
30. Zhang H, Penninger JM, Li Y, *et al.* Angiotensin-converting enzyme 2 (ACE2) as a SARS-CoV-2 receptor: Molecular mechanisms and potential therapeutic target. *Intensive Care Med* 2020; 1–5.
31. Walls AC, Park YJ, Tortorici MA, *et al.* Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. *Cell* 2020.
32. Wang N, Shang J, Jiang S, *et al.* Subunit vaccines against emerging pathogenic human coronaviruses. *Front Microbiol* 2020;11:298.
33. Moorthy V, Restrepo AM, Preziosi MP, *et al.* Data sharing for novel coronavirus (COVID-19). *Bull World Health Organ* 2020;98(3): 150.