Sains Malaysiana 53(8)(2024): 1767-1776
http://doi.org/10.17576/jsm-2024-5308-04

Genome Sequencing and Phylogenetic Analysis of SARS-CoV-2 Isolated from Patients with COVID-19 in Hospital Canselor Tuanku Muhriz (HCTM), Malaysia

(Penjujukan Genom dan Analisis Filogenetik SARS-CoV-2 Dipencilkan daripada Pesakit COVID-19 di Hospital Canselor Tuanku Muhriz (HCTM), Malaysia)

KON KEN WONG[1], NOOR ZETTI ZAINOL RASHID[1], UMI KALSOM ALI[1], NUR HAZLIN CHONG[2], SHAIRAH ABDUL RAZAK[2], NAJMA KORI[3], PETRICK PERIYASAMY[3] & ALFIZAH HANAFIAH[1]

[1]Department of Medical Microbiology & Immunology, Faculty of Medicine, Universiti Kebangsaan Malaysia, Jalan Yaacob Latiff, Bandar Tun Razak, 56000 Cheras, Kuala Lumpur, Malaysia
[2]Department of Biological Sciences and Biotechnology, Faculty of Science & Technology, Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Selangor, Malaysia
[3]Department of Medicine, Faculty of Medicine, Universiti Kebangsaan Malaysia, Jalan Yaacob Latiff, Bandar Tun Razak, 56000 Cheras, Kuala Lumpur, Malaysia

Received: 18 March 2024/Accepted: 26 June 2024

ABSTRACT

SARS-CoV-2 infection has reached pandemic status in numerous countries worldwide, including Malaysia. Monitoring the genetic diversity of SARS-CoV-2 is essential for identifying the emergence and prevalence of novel variants in different geographical areas. From May to October 2021, this research endeavor analyzed the SARS-CoV-2 genome of 40 COVID-19 patients diagnosed using real-time RT-PCR at the local hospital HCTM-UKM. The process involved extracting RNA from these patients, which was subsequently subjected to whole genome sequencing. Our discovery underscores the primary strains responsible for the fourth wave of the COVID-19 pandemic in Malaysia in May 2021 were the Beta variant, primarily associated with the B.1.351 lineage. However, by October 2021, the Delta variant had become predominant and was categorized within the AV.59 and AV.79 lineages. Genomic epidemiological analysis showed about 19 amino acid mutations in the spike protein that were prevalent during the Beta variant outbreak. Among these, the N501Y mutation is particularly noteworthy as it significantly enhances the virus's ability to bind to the ACE2 receptor. Additionally, 32 amino acid mutations were identified during the Delta variant outbreak, with the T478K mutation being linked to increased viral infectivity and affecting the virus's affinity for human cells. Throughout our research, we consistently noted the presence of Spike D614G mutations in all the strains we collected which is known to reduce S1 shedding and increase infectivity. These findings could contribute significantly to our understanding of specific variants due to their clinical implications and rapid spread within the community.

Keywords: COVID-19; genome sequencing; phylogenetic; spike protein mutations; variant

ABSTRAK

Jangkitan SARS-CoV-2 telah mencapai status pandemik di hampir semua negara di seluruh dunia, termasuk Malaysia. Pemantauan kepelbagaian genetik SARS-CoV-2 adalah penting untuk mengenal pasti penemuan dan prevalen varian baharu di kawasan geografi yang berbeza. Dari Mei hingga Oktober 2021, kami menganalisis genom SARS-CoV-2 bagi 40 pesakit COVID-19 dengan menggunakan kaedah '*realtime RT-PCR*' di hospital tempatan HCTM-UKM. Proses ini melibatkan pengekstrakan RNA daripada sampel pesakit ini dan dikaji dengan teknik jujukan lengkap genom. Penemuan utama kami adalah strain SARS-CoV-2 yang berleluasa semasa pandemik COVID-19 gelombang keempat di Malaysia pada Mei 2021 adalah varian Beta dengan garis keturunan B.1.351. Namun demikian, varian Delta telah menjadi dominan pada Oktober 2021, dan dikategorikan dalam garis keturunan AV.59 dan AV.79. Analisis epidemiologi genom menunjukkan kira-kira 19 mutasi asid amino dalam protein spike yang mendominasi semasa wabak varian Beta. Antara ini, mutasi N501Y adalah perlu diperhatikan kerana ia akan meningkatkan keupayaan virus untuk berikatan dengan reseptor ACE2 secara signifikan. Tambahan pula, sebanyak 32 mutasi asid amino dikenal pasti semasa wabak varian Delta dengan mutasi T478K dikaitkan dengan peningkatan jangkitan virus dan mempengaruhi keupayaan virus untuk berikatan dengan sel manusia. Sepanjang penyelidikan kami, mutasi Spike D614G dijumpai dalam semua strain yang kami kumpulkan. Mutasi ini dapat mengurangkan pembuangan S1 dan meningkatkan jangkitan virus. Dapatan kajian ini boleh memberi sumbangan yang signifikan kepada pemahaman kita mengenai varian tertentu disebabkan implikasi klinikal dan penyebaran yang pesat dalam komuniti.

Kata kunci: COVID-19; filogenetik; mutasi protein Spike; penjujukan genom; varian

Introduction

Emerging and re-emerging pathogens pose global challenges to public health (Gao 2018). One prominent example of such pathogens is coronaviruses, which are enveloped RNA viruses that are widely distributed among humans, other mammals, and birds. These viruses have been known to cause respiratory, enteric, hepatic, and neurological diseases (Weis & Leibowitz 2011). In particular, six coronavirus species are known to cause human diseases (Su et al. 2016). Four coronaviruses, 229E, OC43, NL63, and HKU1, are prevalent and typically cause common cold symptoms in immunocompetent individuals (Su et al. 2016). The other two strains, severe acute respiratory syndrome coronavirus (SARS-CoV) and Middle East respiratory syndrome coronavirus (MERS-CoV), are zoonotic in origin and have been linked to sometimes fatal illnesses (Cui, Li & Shi 2019). SARS-CoV-1 was the causal agent of the severe acute respiratory syndrome outbreaks in 2002 and 2003 in Guangdong Province, China (Drosten et al. 2003; Ksiazek et al. 2003; Zhong et al. 2003). MERS-CoV was the pathogen responsible for severe respiratory disease outbreaks in 2012 in the Middle East (Zaki et al. 2012).

A novel human coronavirus, now named severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) (formerly called nCoV-19), emerged from Wuhan, China in late 2019 and has since contributed to the latest pandemic that has spread worldwide (WHO 2020). The disease caused by SARS-CoV-2 is known as coronavirus disease 2019 (COVID-19), which has rapidly become widespread in more than 150 countries around the globe. On January 30, 2020, the World Health Organization (WHO) declared the novel coronavirus pneumonia epidemic caused by SARS-CoV-2 as a public health emergency of international concern, calling for the urgent need for global cooperation to contain and mitigate the spread of this infectious disease.

Person-to-person transmission of COVID-19 has been described in both hospital and family settings, indicating the potential for widespread transmission (Chan et al. 2019). The primary mode of COVID-19 transmission is through respiratory droplets, with people also at risk of infection through contact with virus droplets. Postulations have been made regarding the potential transmission of coronaviruses through contaminated dry surfaces, including the possibility of self-inoculation of mucous membranes of the nose, eyes and digestive tract via the fecal-oral route (Dowell et al. 2004; Otter et al. 2016; Wu et al. 2020).

The presence of quasispecies has been previously reported for both SARS-CoV and MERS-CoV (Park et al 2016; Xu, Zhang & Wang 2004), indicating that these beta coronaviruses may be composed of complex and dynamically distributed closely related variants *in vivo*, similar to other RNA viruses. Moreover, the analysis of SARS-CoV-2 sequence variability has provided evidence for the existence of viral quasispecies in both clinical samples and primary isolates (Capobianchi et al. 2020). In a recent study, the team maps the transmission paths of confirmed cases of COVID-19 by phylogenetic networks to trace the origins of unreported COVID-19 infections, allowing for the isolation of these sources to curb the potential global recurrence of the disease (Forster et al 2020).

The genomic characterization of SARS-CoV-2 in this study will provide important information on the origins and clusters of the circulating virus among our patients. This may partly explain the infectiousness and transmissibility of SARS-CoV-2 in our population. Therefore, it is imperative to explore the whole genome sequencing (WGS) data of SARS-CoV-2 to understand its variation characteristics. Tracing the virus genome is a critical task, along with early initiation of treatment and infection control measures.

Materials and Methods

Study samples and viral RNA extraction

Nasopharyngeal/oropharyngeal swab samples from patients with severe acute respiratory syndrome (SARS) tested positive for SARS-CoV-2 by RT-PCR at Molecular Diagnostic Laboratory, HCTM-UKM from May 2021 – Nov 2021 were collected. Patients' history and clinical information were extracted from their medical record. The samples were handled under biosafety cabinet and in a dedicated and close room according to Ministry of Health, Malaysia guideline for COVID-19.

Viral RNA from the swab samples in viral transport medium (VTM) was extracted using GeneAll® RibospinTM vRD II RNA Extraction Kit (GeneAll, Korea) and followed by SARS-CoV-2 RNA detection with ADT Lystar™ Real-time RT-PCR Kit (ADT Biotech Sdn Bhd, Malaysia). RT-PCR was conducted follow the kit instructions and the result was analyzed using Q-Rex Software, in which a cycle threshold value (Ct value) of <45 for the amplification for E and RdRP viral gene regions, were defined as a positive result.

Genome sequencing

We performed whole genome sequencing on samples with SARS-CoV-2 RT-PCR positive. Briefly, cDNA was synthesized using the LunaScript RT SuperMix. PCR was performed using two pools of ARTIC V3 primers targeting regions of the SARS-CoV-2 genome. The ARTIC amplicons were then combined and purified. Following that, end repair and adaptor ligation were performed. The sequencing libraries were prepared using the NEBNext® ARTIC SARS-CoV-2 Library Prep Kit (Illumina®). The completed libraries were quantified with a Qubit 2.0 fluorometer. The size distribution of the libraries was assessed using TapeStation 2200. Sequencing was

performed on an Illumina MiSeq system (San Diego, CA, USA) using 2 × 250-bp paired-end reads.

PHYLOGENETIC ANALYSIS

The sequence data generated by the MiSeq sequencer underwent data preprocessing, followed by read alignment and lineage assignment. Adapter sequences were removed from the sequence data using Trimmomatic v0.39. Subsequently, these sequencing reads were aligned to the human reference genome GRCh38 to detect any human DNA contamination. The sequence data were then aligned to the SARS-CoV-2 reference genome (Wuhan-Hu-1, NCBI accession: MN908947.3) using Bowtie 2 (version 2.4.4) and SAMtools v1.13. Based on this alignment, the consensus FASTA file was then derived using iVar v1.0. Lineage assignment was conducted on the consensus sequences using Pangolin (Phylogenetic Assignment of Named Global Outbreak Lineages) v3.1.17, pangoLEARN v1.2.105 (6/12/2021) and Scorpio (serious constellations of recurring phylogeneticallyindependent origin) v0.3.16. Finally, a phylogenetic tree was generated based on Maximum Parsimony using UshER v0.4.8 from UCSC genome browser (https://genome.ucsc.edu/cgi-bin/hgPhyloPlace). Genome sequences were inserted into pre-constructed global phylogenetic tree comprising of 4.8 million publicly available SARS-CoV-2 genomes from GISAID database, National Center for Biotechnology (NCBI) GenBank, COVID-19 Genomics UK (COG-UK) Consortium database and China National Center for Bioinformation (CNCB) database. Visualization of the SARS-CoV-2 subtrees was done using NextStrain.

Results and Discussion

EPIDEMIOLOGICAL DATA ANALYSIS

In this study, we have only proceeded with genome sequencing of patient samples that have tested positive for SARS-CoV-2 using Real-time RT-PCR with a Ct value of less than 20. We have analyzed a total of 40 SARS-CoV-2 positive samples, consisting of 10 samples collected in May 2021 and 30 samples collected in October and November 2021. This period was categorized as the fourth wave, which began in the middle of May and peaked in the middle of June and the second peak emerged in the middle of October (COVID-19 pandemic in Malaysia 2022).

The patients' age range varied from 1 month to 80 years old, with an average age of 51 years. Among the patient group, 62.5% were females and 27.5% were males. Unfortunately, 32/40 patients were hospitalized for further treatment and one patient was succumbed due to viral pneumonia. The high coverage sequences obtained have been deposited in the GISAID EpiCoV newly emerging coronavirus SARS-CoV-2 platform and have been subjected to phylogenetic analysis to understand their molecular relationships and evolution. The sequences can be identified through the identifiers EPI_ISL_13338022 to EPI_ISL_13498261.

SEQUENCING AND GENETIC CHARACTERISTICS OF SARS-COV-2 ISOLATES

The whole genome sequences of 40 isolates were obtained, covering the SARS-CoV-2 genome by 95.02%–99.60%, with a total base pair (bp) count ranging from 41,969,710 to 183,025,274. Based on data from the GISAID EpiCoV database, a total of six lineages were identified through whole genome sequencing, with the dominant lineages being AY.59 (n = 17), AY.79 (n = 10), B.1.351 (n = 9), AY.85 (n = 2), B.1.466.2 (n = 1), and B.1.617.2 (n = 2). Table 1 shows the sequencing information of the isolates, such as sample ID, Accession ID, Clade, Pangoline Lineage, Ambiguity score, and Scorpio call. In terms of the sequencing database, we found that the lineages of SARS-CoV-2 isolates in Malaysia during the fourth wave by May 2021 were the Beta variant (10/10) and Clade GH. When compared to the outbreak that occurred in October 2021, Clade GK, Delta variant was the dominant strain in Malaysia (29/30) (Figure 1). Table 1 demonstrates the distribution of clades and sub-lineages of each virus sequence of SARS-CoV-2.

We employed a phylogenetic tree as a tool to investigate and elucidate the inference, as well as to trace the evolutionary history and connections within various strains of SARS-CoV-2 worldwide. Our analysis involved a comparison of the complete genome sequences of 40 samples, drawing from a vast dataset comprising more than 12 million genomes sourced from GISAID, Genebank COG-UK, and CNCB as of October 12th, 2022. Subsequently, we constructed a minimum spanning tree by incorporating the sarcov2phylo 13-11-20 tree, to which we added more recent sequences using UShER, and included custom subtree tracks as needed.

According to our phylogenetic analysis, we observed that in March 2021, a significant majority of locally-acquired cases in our study (90%, or 9 out of 10 cases) were associated with the PANGO lineage B.1.351, while only one case was attributed to B.1.466.2. In contrast, by October 2021, the AY.59 lineage became predominant, accounting for approximately 56.7% of the total isolates, or 17 out of 30 cases. When employing Nextstrain classification, the collected isolates were categorized into four major global clades: 20H, 21A, 21I, and 21J. Notably, a significant proportion of the isolates fell within the phylogenetic clade 21I, representing 45% of the total (18 out of 40). It is worth mentioning that the lineage distribution of these 40 isolates from COVID-19 patients closely mirrored the genetic lineages of SARS-CoV-2 that were circulating worldwide during that specific time period (Figure 2).

TABLE 1. Genetic Characteristics variants of SARS-CoV-2 isolates with GISAID accession numbers

Lineage Assignment isolates collected on May 2021

No.	Sample ID	Accession ID	Clade	Lineage	Ambiguity score	Scorpio_call
1	UKM 1893	EPI_ISL_13338022	GH	B.1.351	1	Beta
2	UKM 1907	EPI_ISL_13338023	GH	B.1.466.2	0.997	Beta
3	UKM 1968	EPI_ISL_13338026	GH	B.1.351	0.995	Beta
4	UKM 2438	EPI_ISL_13375800	GH	B.1.351	0.986	Beta
5	UKM 2507	EPI_ISL_13375801	GH	B.1.351	0.988	Beta
6	UKM 2509	EPI_ISL_13375802	GH	B.1.351	0.991	Beta
7	UKM 2596	EPI_ISL_13375803	GH	B.1.351	0.991	Beta
8	UKM 2609	EPI_ISL_13375804	GH	B.1.351	0.991	Beta
9	UKM 2626	EPI_ISL_13375805	GH	B.1.351	0.996	Beta
10	UKM 2698	EPI_ISL_13375817	GH	B.1.351	0.986	Beta

Lineage Assignment isolates collected on October 2021

No.	Sample ID	Accession ID	Clade	Lineage	Ambiguity score	Scorpio_call
1	UKM 1247	EPI_ISL_13375818	GK	AY.85	0.977	Delta (B.1.617.2)
2	UKM 1293	EPI_ISL_13375819	GK	AY.59	0.999	Delta (B.1.617.2)
3	UKM 1444	EPI_ISL_13398733	GH	AY.59	0.995	Delta (B.1.617.2)
4	UKM 1452	EPI_ISL_13398734	GK	AY.79	0.982	Delta (B.1.617.2)
5	UKM 1472	EPI_ISL_14254439	GK	AY.79	0.968	Delta (B.1.617.2)
6	UKM 1503	EPI_ISL_13398735	GK	AY.79	0.980	Delta (B.1.617.2)
7	UKM 1570	EPI_ISL_13398763	GK	AY.59	0.996	Delta (B.1.617.2)
8	UKM 1949	EPI_ISL_13399302	GK	AY.59	0.983	Delta (B.1.617.2)
9	UKM 1985	EPI_ISL_13399476	GK	AY.79	1	Delta (B.1.617.2)
10	UKM 2066	EPI_ISL_13417462	GK	AY.59	0.983	Delta (B.1.617.2)
11	UKM 2150	EPI_ISL_13417463	GK	AY.85	0.999	Delta (B.1.415.1)
12	UKM 2241	EPI_ISL_13417552	GK	AY.59	1	Delta (B.1.617.2)
13	UKM 2413	EPI_ISL_13417553	GK	AY.79	0.983	Delta (B.1.617.2)
14	UKM 2618	EPI_ISL_13417554	GK	AY.59	1	Delta (B.1.617.2)
15	UKM 2774	EPI_ISL_13498246	GK	AY.59	1	Delta (B.1.617.2)
16	UKM 2940	EPI_ISL_13498247	GK	AY.59	1	Delta (B.1.617.2)
17	UKM 2969	EPI_ISL_13498248	GK	AY.59	0.983	Delta (B.1.617.2)
18	UKM 3078	EPI_ISL_13498249	GK	AY.59	0.970	Delta (B.1.617.2)
19	UKM 3113	EPI_ISL_13498250	GK	AY.59	0.994	Delta (B.1.617.2)
20	UKM 3191	EPI_ISL_13498251	GK	AY.59	1	Delta (B.1.617.2)
21	UKM 3268	EPI_ISL_13498252	GK	AY.59	1	Delta (B.1.617.2)
22	UKM 3369	EPI_ISL_13498253	GK	AY.59	0.983	Delta (B.1.617.2)
23	UKM 3434	EPI_ISL_13498254	GK	AY.59	1	Delta (B.1.617.2)
24	UKM 3638	EPI_ISL_13498255	GK	AY.79	0.995	Delta (B.1.617.2)
25	UKM 3660	EPI_ISL_13498256	GK	AY.59	0.983	Delta (B.1.617.2)
26	UKM 3910	EPI_ISL_13498257 EPI_ISL_13498246	GK	AY.79	0.977	Delta (B.1.617.2)
27	UKM 4072	EPI_ISL_13498258	GK	AY.79	0.936	Delta (B.1.617.2)
28	UKM 4323	EPI_ISL_13498259	GK	AY.79	1	Delta (B.1.617.2)
29	UKM 4344	EPI_ISL_13498260	GK	B.1.617.2	0.988	Delta (B.1.617.2)
30	UKM 4364	EPI_ISL_13498261	GK	AY.79	1	Delta (B.1.617.2)

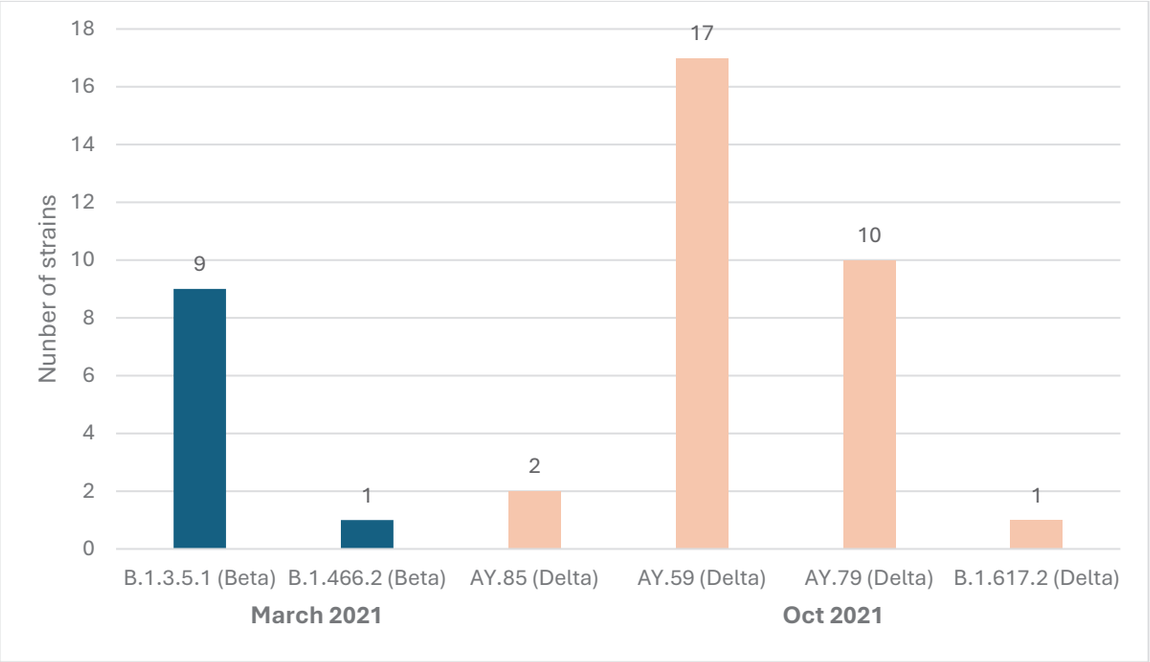


FIGURE 1. Strain distribution of SARS-CoV-2 genomes in Malaysia during the 4th wave of COVID-19

MUTATIONS IDENTIFIED IN THE SEQUENCED SARS-COV-2 GENOMES

We conducted a comprehensive analysis of full-length genome mutations across all isolates using the GISAID database CoVsurver platform, which can be accessed at (https://gisaid.org/database-features/covsurver-mutations-app/). Our primary focus was on comparing the amino acid (a.a) changes within the spike protein of our collected isolates when compared to the SARS-CoV-2 reference strain hCoV-19/Wuhan/WIV04/2019. It can gain insights into how these genetic changes might have impacted the virus's interaction with the ACE2 receptor and, consequently, its infectivity and transmissibility. To enhance our understanding of these mutations, we harnessed the 3D structural visualization tool available on the CoVsurver platform (Figure 3). The results of our analysis unveiled the presence of approximately 19 amino acid mutations within the spike protein during the Beta variant outbreak. Similarly, we identified a total of 32 amino acid mutations within the spike protein for Delta variant outbreak. All defining mutations and prevalence are shown in Figure (4).

The COVID-19 epidemic, which started in December 2019 with the highly infectious and pathogenic SARS-CoV-2 virus, has deprived millions of lives globally, and it still poses a significant threat to public health all over the world. Therefore, there has been rapid research conducted on the new SARS-CoV-2 virus. This has led to the gradual improvement of relevant diagnostic tests and treatments, as well as the successful launch and promotion of vaccines. However, despite these advancements, our understanding of the basic biological issues of the new SARS-CoV-2 virus remains limited. For instance, we still have a limited understanding of the virus's life cycle and its molecular mechanism of interaction with the host. The impact of virus mutation on pathogenicity and transmission ability also remains the focus of current attention.

Throughout the COVID-19 pandemic, many significant Variants of Interest (VOIs) and Variants of Concern (VOCs) of SARS-CoV-2 have emerged. According to data collected by the WHO, there are currently four VOCs - Alpha, Beta, Gamma, and Delta - that have spread worldwide during the second and third waves of the pandemic (CDC 2022). Our study has highlighted several fast-spreading variants, including the Beta variant that was detected in March 2021 and the Delta variant that was detected in October 2021.

The Beta variant was first identified in the Cape province of South Africa in October 2020 and has 7 critical mutations, which have resulted in a 50% increase in transmission (CMMID Repository 2022). The Beta variant is reported to have increased infectivity and a higher risk of transmission. It also shows resistance to neutralization by monoclonal antibody therapy and convalescent sera. However, the Delta variant may have improved replication ability, which can lead to the production of more virus particles and increased

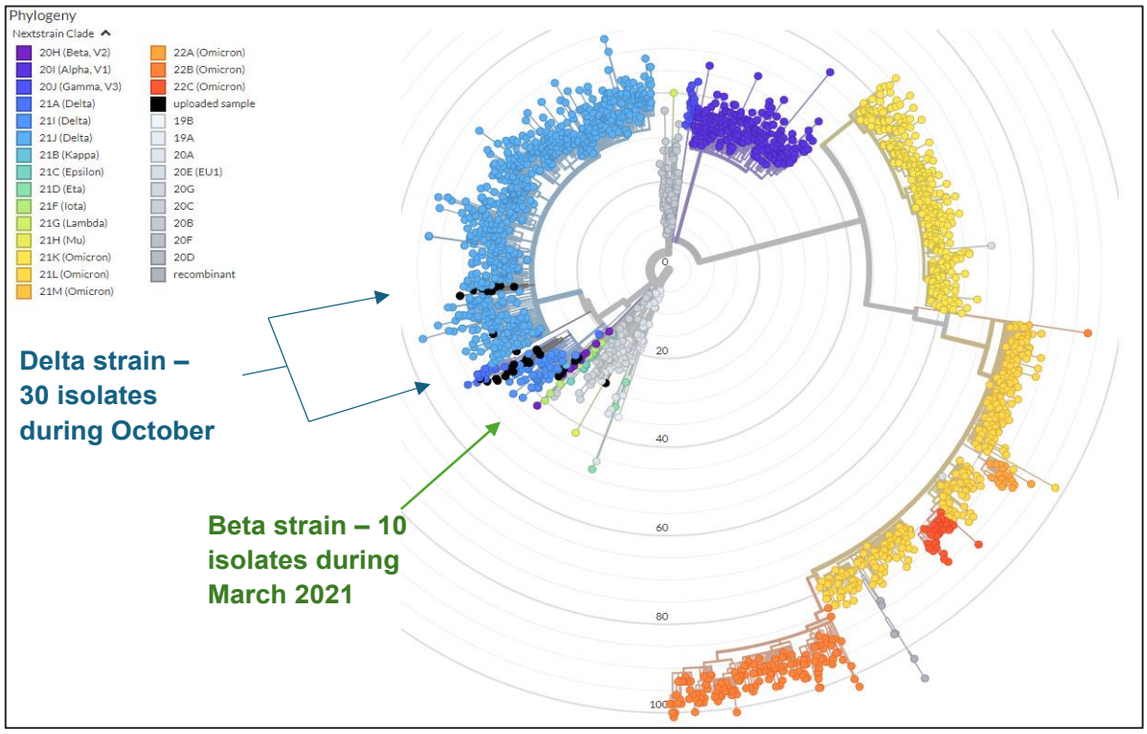


FIGURE 2. Phylogenetic tree showing the placement of samples in global SARS-CoV-2 tree. Subtree with hCoV- uploaded samples. The genomic epidemiology focused sampling of 40 SARS-CoV-2 genomes

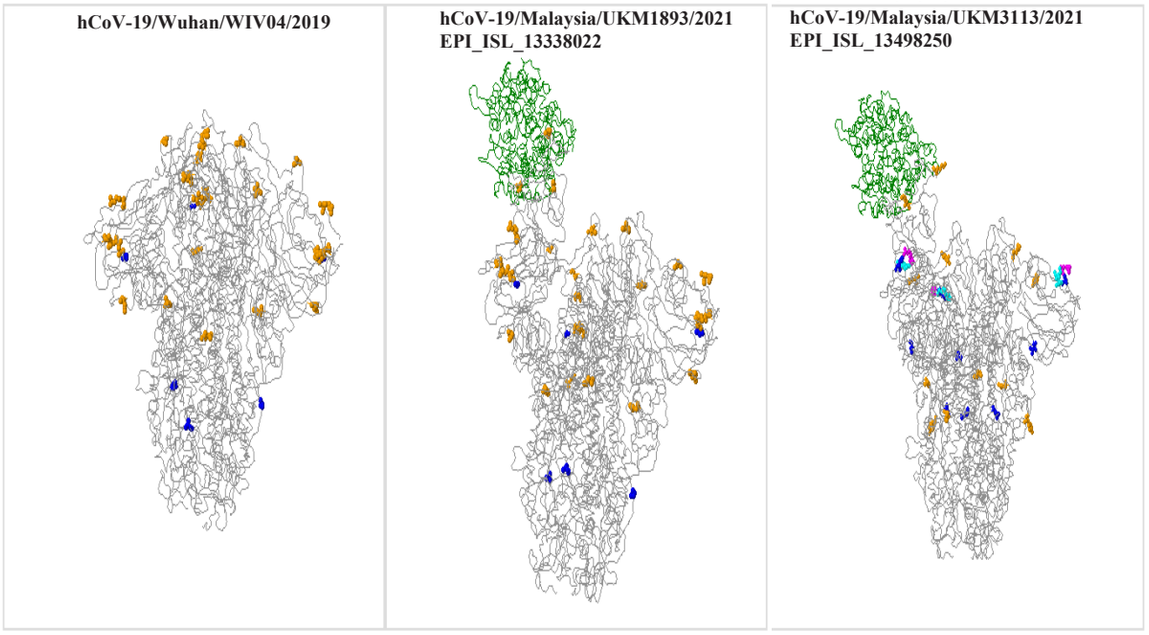
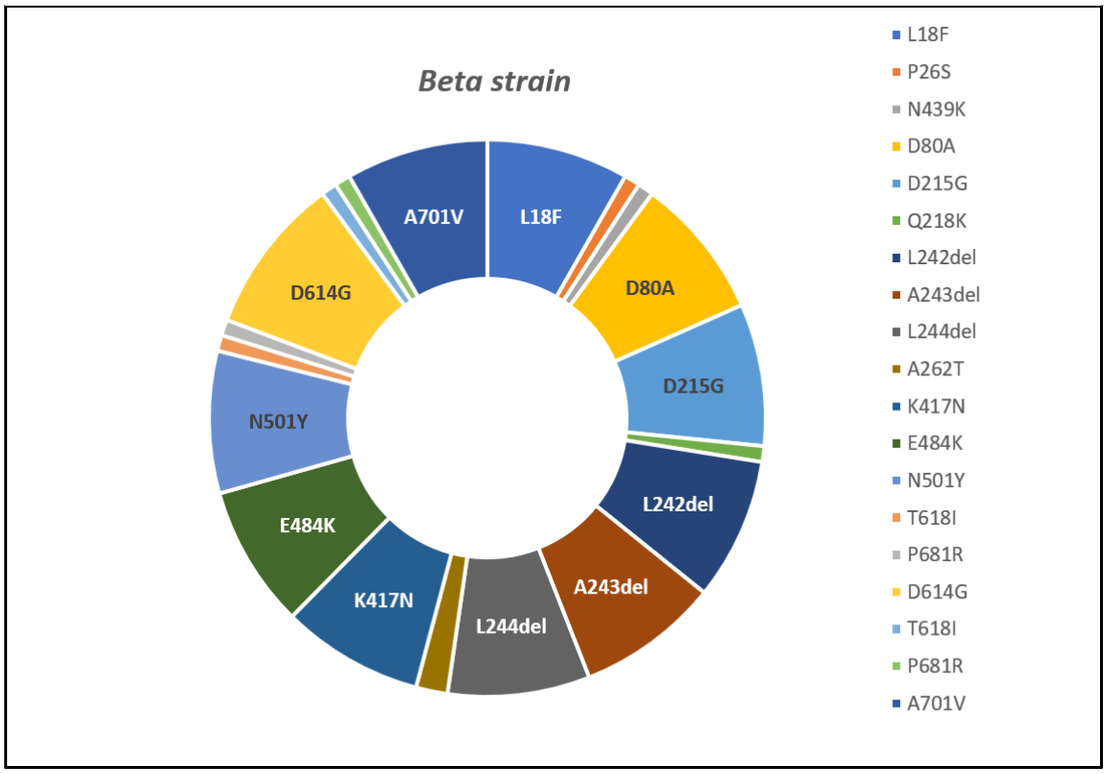


FIGURE 3. 3D structural visualization of the spike glycoprotein with aa changes identified in the query sequences shown as colored balls (GISAID CoVsurver v1.22.06)



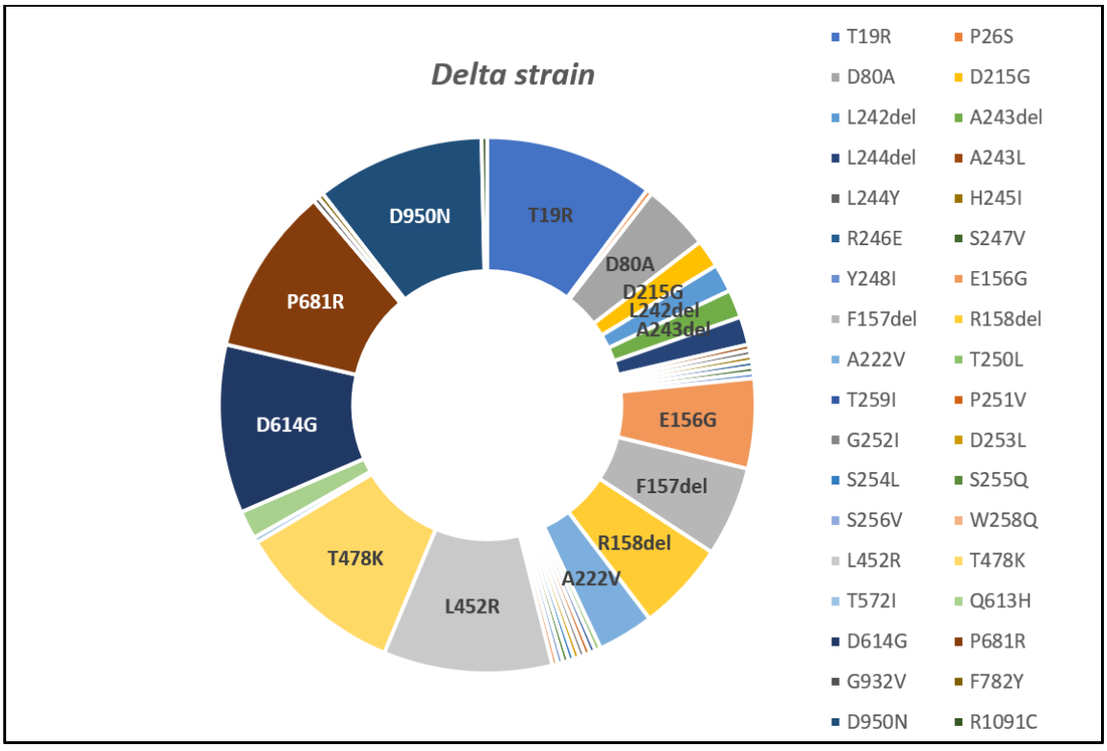


FIGURE 4. Prevalence of acid amino mutation in human SARS-CoV-2 spike proteins for Beta variant and Delta variant

transmission (Aleem, Akbar Samad & Slenker 2022). Among the 19 amino acid changes found in the spike protein of the Beta variant, it is crucial to focus on four specific mutations: The N501Y mutation, in particular, is noteworthy as it substantially increases the virus's binding affinity to the ACE2 receptor. In fact, this mutation boosts the binding affinity by up to 4.62 times when compared to the original SARS-CoV-2 strain (Ramanathan et al. 2021). This heightened affinity for ACE2 may contribute to the Beta variant's increased infectivity. Additionally, the K417N and E484K mutations play a role in the Beta variant's ability to escape neutralizing antibodies. These mutations are thought to alter the shape of the spike protein (Wise 2021), potentially making it more challenging for the immune system to recognize and neutralize the virus effectively. Furthermore, the presence of the L18F mutation, which is specific to the Beta variant, can lead to a steric clash between neighboring residues, which can further affect its interaction with host receptors and immune responses.

The Delta variant was first identified in Maharashtra, India in October 2020. It was declared a variant of concern (VOC) by the World Health Organization (WHO) on May 11, 2021, as it had become the dominant lineage detected in over 100 countries. Among the 32 amino acid changes in the spike protein in Delta variant, the T478K mutation was found to be responsible for increased viral infectivity and influences the virus's affinity to human cells (Di Giacomo et al. 2021). A recent study found that the Delta variant has immunity against specific antibodies, which is due to a P681R mutation that changes an amino acid at the furin cleavage site and makes the Delta variant more efficient at furin cleavage (Science 2022). Another crucial mutation in the Delta variant is L452R, which can enhance transmission efficiency into human cells. A study group had shown that the Delta variant has 18–24% higher transmissibility and 20 times reduction in neutralizing titers from vaccinated individuals (Deng et al. 2021). This further increases the survival and transmission of the Delta variant. The spike mutation T19R in the Delta variant removes a potential N-glycosylation site at position 17, which may also affect the strain's antigenic and other properties. Specifically, the motif at positions 17-19 changed from NLT (glyco) to NLR (no glyco).

The D614G mutation became the center of attention during the early stages of the COVID-19 pandemic due to its high correlation with the widespread infection and virulence, as well as changes in antigenicity (Korber et al. 2020b). In their research, Zhang et al. (2020) demonstrated that the D614G mutation was linked to increased viral loads across the globe. Our analysis highlights the fixation of the D614G mutation in the S protein, as well as the P323L mutation in the RNA-dependent RNA polymerase (RdRp). In October 2020, the Beta (B.1.351) variant was first identified in the Cape province of South Africa, with 17 critical mutations that resulted in a 50% increase in transmission. This variant also evades 21% (95% CrI: 11-36%) of previously acquired immunity (Aleem, Akbar Samad & Slenker 2022). Based on the hypothetical loss of a hydrogen bond between D614 in S1 and T859 in S2, it is suggested that the D614G mutation could promote the shedding of the S1 domain and is also associated with enhanced infectivity (Zhang et al. 2020).

The SARS-CoV-2 virus tends to evolve in a particular direction, leading to either increased infectivity or a stronger immune escape ability. The function of the S protein mutation site needs to be coordinated, and a single site can only achieve a single function; it is difficult to have both simultaneously. We conducted an analysis of the mutation sites observed in the human SARS-CoV-2 spike protein, including mutation percentages, protein sequence, and non-mutated sites, to classify the variants according to WHO labels. There is a need for tracking mutations in the Spike (S) protein, as it is responsible for facilitating the virus's entry into human cells. This approach may aid in the identification of potential inhibitors, epitope sites for drug and vaccine development, and antibody design (Korber et al. 2020a; Guruprasad 2022). There is an uneven effort globally to test for SARS-CoV-2 infections in the population and sequence the viral genomes of those infected. Consequently, emerging mutations may go unreported until they are widespread among the population. Viral genome information provides valuable insights into viral function and biology and allows for the implementation of effective viral surveillance strategies to prevent further transmission and infection.

Conclusion

In conclusion, our findings offer intriguing insights into the various strains of SARS-CoV-2. The emergence of new variants, particularly those with critical mutations in the spike protein, has resulted in increased infectivity and lethality. Further research will be necessary to monitor the mutation rate relative to the severity of COVID-19. To detect epidemiological events, it is essential to strengthen the capacity for surveillance and conduct whole-genome sequencing of indigenous strains prevalent in different regions of the country. The differences in mutation paths lead to changes in antigenicity, indicating that a single-strain vaccine may not be effective in protecting the population against other mutant strains on different evolutionary paths. We will continue to monitor the virus's mutation in real-world settings to develop an advanced broad-spectrum vaccine. In the next five to ten years, a single dose of vaccine may offer protection for up to a decade.

ACKNOWLEDGEMENTS

The authors thank all staff at Microbiology Laboratory, Department of Diagnostic Laboratory Services, Hospital Canselor Tuanku Muhriz, Kuala Lumpur, Malaysia for their technical assistant during samples collection and processing. This research was funded by Faculty of Medicine, Universiti Kebangsaan Malaysia, grant number FF-2020-208.

REFERENCES

Aleem, A., Akbar Samad, A.B. & Slenker, A.K. 2022. Emerging variants of SARS-CoV-2 and novel therapeutics against coronavirus (COVID-19). *StatPearls Publishing* 54(1): 524-540.

Capobianchi, M.R., Rueca, M., Messina, F., Giombini, E., Carletti, F., Colavita, F., Castilletti, C., Lalle, E., Bordi, L., Vairo, F., Nicastri, E., Ippolito, G., Gruber, CEM. & Bartolini, B. 2020. Molecular characterization of SARS-CoV-2 from the first case of COVID-19 in Italy. *Clinical Microbiology and Infection* 26(7): 954-956.

Centers for Disease Control and Prevention (CDC). 2022. *About Variants of the Virus that Causes COVID-19.* https://www.cdc.gov/coronavirus/2019-ncov/variants/variant-classifications.html#concern (Accessed on 6 October 2022).

Chan, J.F., Yuan, S., Kok, K.H., To, K.K., Chu, H., Yang, J., Xing, F., Liu, J., Yip, C.C., Poon, R.W., Tsoi, H.W., Lo, S.K., Chan, K.H., Poon, V.K., Chan, W.M., Ip, J.D., Cai, J.P., Cheng, V.C., Chen, H., Hui, C.K. & Yuen, K.Y. 2020. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: A study of a family cluster. *Lancet* 395(10223): 514-523.

CMMID Repository. 2022. *Estimates of Severity and Transmissibility of Novel Sars-Cov-2 Variant 501y.V2 in South Africa.* https://cmmid.github.io/topics/covid19/sa-novel-variant.html [30 Dec 2022].

COVID-19 pandemic in Malaysia. 2022. https://en.wikipedia.org/wiki/COVID-19_pandemic_in_Malaysia (Accessed on 12 October 2022).

Cui, J., Li, F. & Shi, Z.L. 2019. Origin and evolution of pathogenic coronaviruses. *Nature Reviews Microbiology* 17: 181-192.

Deng, X., Garcia-Knight, M.A., Khalid, M.M., Servellita, V., Wang, C., Morris, M.K., Sotomayor-González, A., Glasner, D.R., Reyes, K.R., Gliwa, A.S., Reddy, N.P., Martin, C.S.S., Federman, S., Cheng, J., Balcerek, J., Taylor, J., Streithorst, J.A., Miller, S., Kumar, G.R., Sreekumar, B., Chen, P.Y., Schulze-Gahmen, U., Taha, T.Y., Hayashi, J., Simoneau, C.R., McMahon, S., Lidsky, P.V., Xiao, Y., Hemarajata, P., Green, N.M., Espinosa, A., Kath, C., Haw, M., Bell, J., Hacker, J.K., Hanson, C., Wadford, D.A., Anaya, C., Ferguson, D., Lareau, L.F., Frankino, P.A., Shivram, H., Wyman, S.K., Ott, M., Andino, R. & Chiu, C.Y. 2021. Transmission, infectivity, and antibody neutralization of an emerging SARS-CoV-2 variant in California carrying a L452R spike protein mutation. *MedRxiv* doi: http://doi.org/10.1101/2021.03.07.21252647. Preprint.

Di Giacomo, S., Mercatelli, D., Rakhimov, A. & Giorgi, F.M. 2021. Preliminary report on severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) spike mutation T478K. *Journal of Medical Virology* 93(9): 5638-5643.

Dowell, S.F., Simmerman, J.M., Erdman, D.D, Wu, J.S., Chaovavanich, A., Javadi, M., Yang, J.Y., Anderson, L.J., Tong, S. & Ho, M.S. 2004. Severe acute respiratory syndrome coronavirus on hospital surfaces. *Clinical Infectious Diseases* 39: 652-657.

Drosten, C., Günther, S., Preiser, W., van der Werf, S., Brodt, H.R., Becker, S., Rabenau, H., Panning, M., Kolesnikova, L., Fouchier, R.A., Berger, A., Burguière, A.M., Cinatl, J., Eickmann, M., Escriou, N., Grywna, K., Kramme, S., Manuguerra, J.C., Müller, S., Rickerts, V., Stürmer, M., Vieth, S., Klenk, H.D., Osterhaus, A.D., Schmitz, H. & Doerr, H.W. 2003. Identification of a novel coronavirus in patients with severe acute respiratory syndrome. *The New England Journal of Medicine* 348(20): 1967-1976.

Forster, P., Forster, L., Renfrew, C. & Forster, M. 2020. Explaining phylogenetic network analysis of SARS-CoV-2 genomes. *Proceedings of the National Academy of Sciences* 117(23): 12524-12525.

Gao, G.F. 2018. From "A"IV to "Z"IKV: Attacks from emerging and re-emerging pathogens. *Cell* 172: 1157-1159.

Guruprasad, K. 2022. Mutations in human SARS-CoV-2 spike proteins, potential drug binding and epitope sites for COVID-19 therapeutics development. *Current Research in Structural Biology* 4: 41-50.

Korber, B., Fischer, W.M., Gnanakaran, S., Yoon, H., Theiler, J., Abfalterer, W., Foley, B., Giorgi, E., Bhattacharya, T., Parker, M., Partridge, D., Evans, C., Freeman, T., de Silva, T., LaBranche, C. & Montefiori, D. 2020a. Spike mutation pipeline reveals the emergence of a more transmissible form of SARS-CoV-2. *BioRxiv* doi: https://doi.org/10.1101/2020.04.29.069054 Preprint.

Korber, B., Fischer, W.M., Gnanakaran, S., Yoon, H., Theiler, J., Abfalterer, W., Hengartner, N., Giorgi, E.E., Bhattacharya, T., Foley, B., Hastie, K.M., Parker, M.D., Partridge, D.G., Evans, C.M., Freeman, T.M., de Silva, T.I., Sheffield COVID-19 Genomics Group; McDanal, C., Perez, L.G., Tang, H., Moon-Walker, A., Whelan, S.P., LaBranche, C.C., Saphire, E.O. & Montefiori, D.C. 2020b. Tracking changes in SARS-CoV-2 spike: Evidence that D614G increases infectivity of the COVID-19 virus. *Cell* 182(4): 812-827.

Ksiazek, T.G., Erdman, D., Goldsmith, C.S., Zaki, S.R., Peret, T., Emery, S., Tong, S., Urbani, C., Comer, J.A., Lim, W., Rollin, P.E., Dowell, S.F., Ling, A.E., Humphrey, C.D., Shieh, W.J., Guarner, J., Paddock, C.D., Rota, P., Fields, B., DeRisi, J., Yang, J.Y., Cox, N., Hughes, J.M., LeDuc, J.W., Bellini, W.J., Anderson, L.J. & SARS Working Group. 2003. A novel coronavirus associated with severe acute respiratory syndrome. *The New England Journal of Medicine* 348(20): 1953-1966.

Otter, J.A., Donskey, C., Yezli, S., Douthwaite, S., Goldenberg, S.D. & Weber, D.J. 2016. Transmission of SARS and MERS coronaviruses and influenza virus in healthcare settings: The possible role of dry surface contamination. *Journal of Hospital Infection* 92: 235-250.

Park, D., Huh, H.J., Kim, Y.J., Son, D.S., Jeon, H.J., Im, E.H., Kim, J.W., Lee, N.Y., Kang, E.S., Kang, C.I., Chung, D.R., Ahn, J.H., Peck, K.R., Choi, S.S., Kim, Y.J., Ki, C.S. & Park, W.Y. 2016. Analysis of intrapatient heterogeneity uncovers the microevolution of Middle East respiratory syndrome coronavirus. *Cold Spring Harbor Molecular Case Studies* 2(6): a001214.

Ramanathan, M., Ferguson, I.D., Miao, W. & Khavari, P.A. 2021. SARS-CoV-2 B.1.1.7 and B.1.351 spike variants bind human ACE2 with increased affinity. *The Lancet Infectious Diseases* 21(8): 1070.

Science. 2022. Delta variant triggers dangerous new phase in the pandemic. https://www.sciencemag.org/news/2021/06/deltavariant-triggers-dangerous-new-phase-pandemic (Accessed on 31 December 2022).

Su, S., Wong, G., Shi, W., Liu, J., Lai, A.C.K., Zhou, J., Liu, W., Bi, Y. & Gao, G.F. 2016. Epidemiology, genetic recombination, and pathogenesis of coronaviruses. *Trends Microbiol.* 24(6): 490-502.

Weis, S.R. & Leibowitz, J.L. 2011. Coronavirus pathogenesis. *Advances in Virus Research* 81: 85-164.

Wise, J. 2021. COVID-19: The E484K mutation and the risks it poses. *British Medical Journal* 372: n359.

World Health Organization (WHO) 2020. *Coronavirus Disease (COVID-2019) Situation Reports- 39.* Geneva: WHO. https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports (Accessed on 1 October 2022).

Wu, Y., Guo, C., Tang, L., Hong, Z., Zhou, J., Dong, X., Yin, H., Xiao, Q., Tang, Y., Qu, X., Kuang, L., Fang, X., Mishra, N., Lu, J., Shan, H., Jiang, G. & Huang, X. 2020. Prolonged presence of SARS-CoV-2 viral RNA in faecal samples. *Lancet Gastroenterol. Hepatol.* 5(5): 434-435.

Xu, D., Zhang, Z. & Wang, F.S. 2004. SARS-associated coronavirus quasispecies in individual patients. *The New England Journal of Medicine* 350(13): 1366-1367.

Zaki, A.M., van Boheemen, S., Bestebroer, T.M., Osterhaus, A.D. & Fouchier, R.A. 2012. Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *The New England Journal of Medicine* 367: 1814-1820.

Zhang, L., Jackson, C.B., Mou, H., Ojha, A., Rangarajan, E.S., Izard, T., Farzan, M. & Choe, H. 2020. The D614G mutation in the SARS-CoV-2 spike protein reduces S1 shedding and increases infectivity. *BioRxiv* http://doi.org/10.1101/2020.06.12.148726 Preprint.

Zhong, N.S., Zheng, B.J., Li, Y.M., Poon, X.Z.H., Chan, K.H., Li, P.H., Tan, S.Y., Chang, Q., Xie, J.P., Liu, X.Q., Xu, J., Li, D.X., Yuen, K.Y., Peiris & Guan, Y. 2003. Epidemiology and cause of severe acute respiratory syndrome (SARS) in Guangdong, People's Republic of China. *Lancet* 362(9393): 1353-1358.

*Corresponding author; email: alfizah@ppukm.ukm.edu.my