JMB Journal of Microbiology and Biotechnology

J. Microbiol. Biotechnol. 2020. 30(7): 962–966
https://doi.org/10.4014/jmb.2006.06009

Minireview

An Overview of the Genetic Variations of the SARS-CoV-2 Genomes Isolated in Southeast Asian Countries

Polly Soo Xi Yap[1*], Tse Siang Tan[2], Yoke Fun Chan[1], Kok Keng Tee[1], Adeeba Kamarulzaman[3], and Cindy Shuan Ju Teh[1*]

[1]Department of Medical Microbiology, Faculty of Medicine, University of Malaya, 50603 Kuala Lumpur, Malaysia
[2]Department of Anaesthesiology and Intensive Care, UKM Medical Centre, 56000 Kuala Lumpur, Malaysia
[3]Department of Medicine, Faculty of Medicine, University of Malaya, 50603 Kuala Lumpur, Malaysia

Monitoring the mutation dynamics of human severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is critical in understanding its infectivity, virulence and pathogenicity for development of a vaccine. In an "age of mobility," the pandemic highlights the importance and vulnerability of regionalization and labor market interdependence in Southeast Asia. We intend to characterize the genetic variability of viral populations within the region to provide preliminary information for regional surveillance in the future. By analyzing 142 complete genomes from South East Asian (SEA) countries, we identified three central variants distinguished by nucleotide and amino acid changes.

Keywords: Coronavirus, SARS-CoV-2, genetic variation, mutation, COVID-19, Southeast Asia

The Study

Human mobility (in parentheses) is one of the main factors that contribute to the worldwide dissemination of microorganisms. The spread of coronavirus disease 2019 (COVID-19) was recently reported to transmit to neighboring countries with relocation diffusion [1]. With most of the studies focusing on China, Western Europe and the USA, little is known about its evolution and genome variability in Southeast Asian (SEA) countries. SEA is home to more than half a billion or 9% of the world's population. As the region grapples with a surge in infection cases since March 2020 [2], it is important to investigate purported mutations and the role of geographical proximity in shaping the genetic structure of the SARS-CoV-2 in SEA countries. On March 4, 2020, the World Health Organisation (WHO) outlined that only nine of the eleven countries have the capacity to test for COVID-19 [3], suggesting that the lack of testing facilities could hinder the preparedness and response planning of these countries towards COVID-19.

Among the SEA countries, Malaysia, Thailand and Singapore employ a large number of migrant workers, with Malaysia being the top importer with approximately 2.23 million people [4]. Concurrently, there has been a mass exodus of Malaysians seeking greater economic security in Singapore, with approximately 450,000 people crossing the Malaysia-Singapore border daily [5]. The Indonesia authorities reported that more than 64,000 Indonesian migrant workers had returned from Malaysia amid the country's ongoing lockdown [6]. Another type of human mobility is refugees. The political instability which holds sway in Myanmar has forced 10% of the population to emigrate in search of refuge [7]. More boats carrying Rohingya refugees were spotted off the coasts of Malaysia and southern Thailand, adding to the challenges faced by these countries fighting the pandemic outbreak [8]. This trend of massive internal mobility is expected to continue as the countries ease the lockdown in the foreseeable future, continuously shifting the genetic drift of the viral population. Studies have shown that human migration (gene flow) is a remarkable factor to consider in virus evolution [9]. Hence, characterization of the genetic variability of viral populations provides important insights in virus evolution and epidemiology for devising efficient and reliable infection control strategies. As of April 30, 2020, only 142 complete sequences (plural) of SARS-CoV-2 from six of the SEA countries, including Cambodia (n = 1), Malaysia (n = 16), the Philippines / Philippines (n = 12), Singapore (n = 74), Thailand (n = 31) and Vietnam (n = 8) (Supplementary Data 1) were deposited in the Global Initiative on Sharing Avian Influenza Data (GISAID) platform. All sequences have been included in the study. The majority of the strains were collected before their countries of origin imposed travel restrictions limiting entry (Supplementary Data 2). Therefore, in a data-limited situation, this study is intended to serve as an early snapshot of the genetic variations of the SARS-CoV-2 within the region and may help identify the endemic genotypes to provide fundamental information for regional surveillance in the future.

Overall, all strains isolated before national implementation of border control were largely invariant (n = 43), while Clusters I, II, and III were sampled relatively recently and showed approximately 80% of synonymous mutations, suggesting possible ongoing adaptation of SARS-CoV-2 in the region. This is in agreement with a

Received: June 8, 2020
Accepted: June 21, 2020

First published online:
June 24, 2020

**Corresponding authors*
P.S.X.Y
Phone: +603-79676601
Fax: +603-79676678
E-mail: polly_yapsooxi@yahoo.com
C.S.J.T.
Phone: +603-79676674
Fax: +603-79676678
E-mail: cindysjteh@um.edu.my

Supplementary data for this paper are available on-line only at http://jmb.or.kr.

pISSN 1017-7825
eISSN 1738-8872

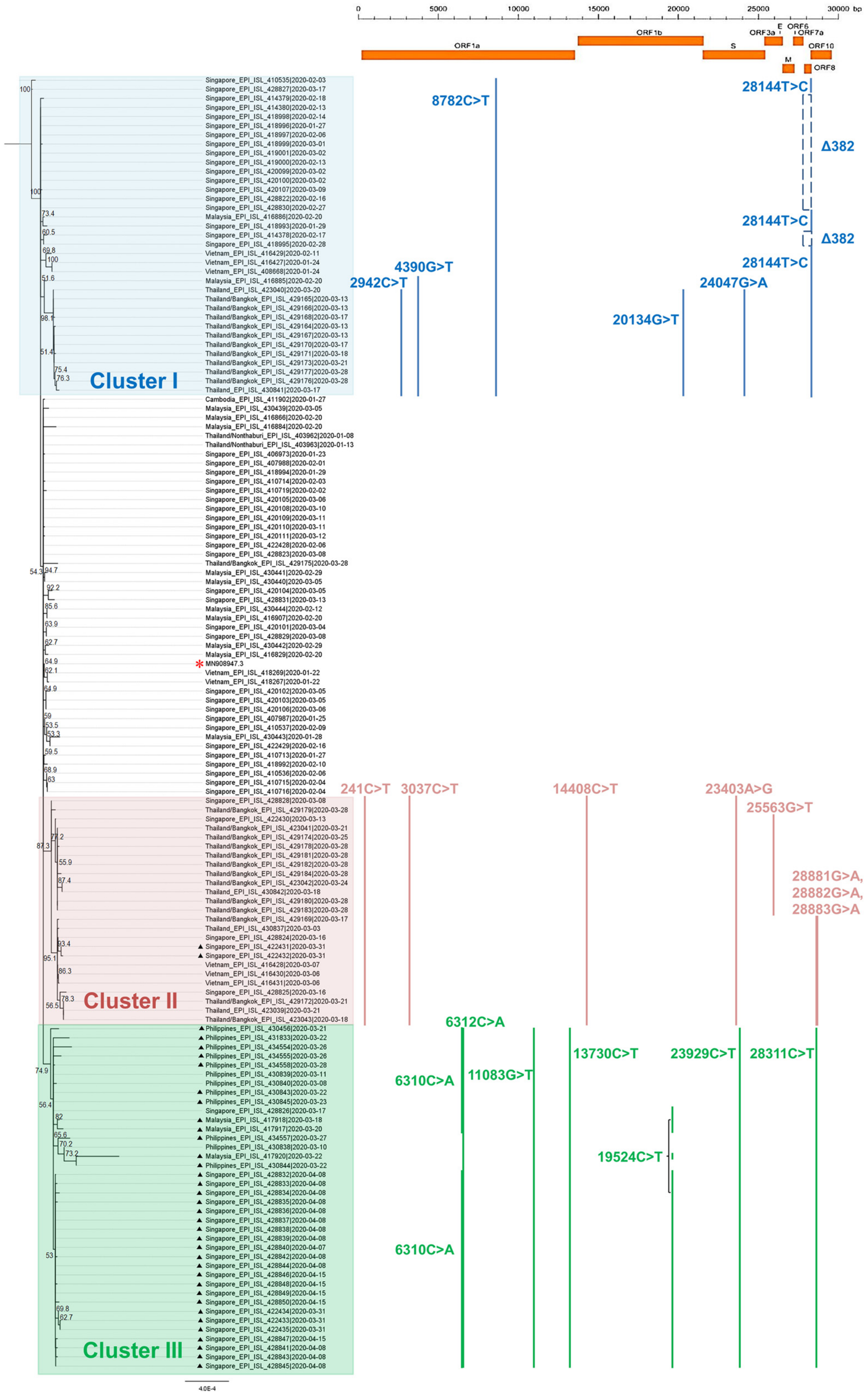

Fig. 1. Phylogenetic tree inferred by using neighbor-joining method and Tamura-Nei model with 1,000 bootstrap replicates, representing complete SARS-CoV-2 genomes from SEA countries against the reference genome Wuhan-Hu-1 (*). Clusters identified were colored accordingly: Cluster I (blue), Cluster II (red) and Cluster III (green). Isolates with symbol (▲) denotes samples collected after country border control implementation whereas isolates without symbol denotes samples collected before country border control implementation.

Clade			S		L									V		G				
																			GH	GR
Country	Strain Name	Collection Date	C8782T	T28144C	C241T	C3037T	A23403G	C8782T	G11083T	G25563T	G26144T	T28144C	G28882A	G11083T	G26144T	C241T	C3037T	A23403G	G25563T	G28882A
Cambodia	EPI_ISL_411902	2020-01-27																		
Malaysia	EPI_ISL_430443	2020-01-28																		
	EPI_ISL_430444	2020-02-12																		
	EPI_ISL_416829	2020-02-20																		
	EPI_ISL_416866	2020-02-20																		
	EPI_ISL_416884	2020-02-20																		
	EPI_ISL_416885	2020-02-20	■	■				■				■								
	EPI_ISL_416886	2020-02-20	■	■				■				■								
	EPI_ISL_416907	2020-02-20																		
	EPI_ISL_430441	2020-02-29																		
	EPI_ISL_430442	2020-02-29																		
	EPI_ISL_430440	2020-03-05																		
	EPI_ISL_430439	2020-03-05																		
	EPI_ISL_417919	2020-03-14																		
	EPI_ISL_417918 ▲	2020-03-18							■					■						
	EPI_ISL_417917 ▲	2020-03-20							■					■						
	EPI_ISL_417920 ▲	2020-03-22							■					■						
Philippines	EPI_ISL_430840	2020-03-08							■					■						
	EPI_ISL_430838	2020-03-10							■					■						
	EPI_ISL_430839	2020-03-11							■					■						
	EPI_ISL_430456 ▲	2020-03-21																		
	EPI_ISL_431833 ▲	2020-03-22							■					■						
	EPI_ISL_430843 ▲	2020-03-22							■					■						
	EPI_ISL_430844 ▲	2020-03-22							■					■						
	EPI_ISL_430845 ▲	2020-03-23							■					■						
	EPI_ISL_434554 ▲	2020-03-26							■					■						
	EPI_ISL_434555 ▲	2020-03-26							■					■						
	EPI_ISL_434557 ▲	2020-03-27							■					■						
	EPI_ISL_434558 ▲	2020-03-28							■					■						
Singapore	EPI_ISL_406973	2020-01-23																		
	EPI_ISL_407987	2020-01-25																		
	EPI_ISL_418996	2020-01-27	■					■												
	EPI_ISL_410713	2020-01-27									■				■					
	EPI_ISL_418993	2020-01-29	■	■				■				■								
	EPI_ISL_418994	2020-01-29							■					■						
	EPI_ISL_407988	2020-02-01							■					■						
	EPI_ISL_410719	2020-02-02																		
	EPI_ISL_410535	2020-02-03	■	■				■				■								
	EPI_ISL_410714	2020-02-03									■				■					
	EPI_ISL_410716	2020-02-04									■				■					
	EPI_ISL_410715	2020-02-04									■				■					
	EPI_ISL_418997	2020-02-06	■					■												
	EPI_ISL_422428	2020-02-06																		
	EPI_ISL_410536	2020-02-06									■				■					
	EPI_ISL_410537	2020-02-09																		
	EPI_ISL_418992	2020-02-10									■				■					
	EPI_ISL_414380	2020-02-13	■					■												
	EPI_ISL_419000	2020-02-13	■					■												
	EPI_ISL_418998	2020-02-14	■					■												
	EPI_ISL_422429	2020-02-16																		
	EPI_ISL_428822	2020-02-16	■					■												
	EPI_ISL_414378	2020-02-17	■					■												
	EPI_ISL_414379	2020-02-18	■					■												
	EPI_ISL_428830	2020-02-27	■					■												
	EPI_ISL_418995	2020-02-28	■					■												
	EPI_ISL_418999	2020-03-01	■					■												
	EPI_ISL_419001	2020-03-02	■					■												
	EPI_ISL_420099	2020-03-02	■					■												
	EPI_ISL_420100	2020-03-02	■					■												
	EPI_ISL_420101	2020-03-04																		
	EPI_ISL_420102	2020-03-05																		
	EPI_ISL_420103	2020-03-05																		
	EPI_ISL_420104	2020-03-05							■		■			■	■					
	EPI_ISL_420105	2020-03-06																		
	EPI_ISL_420106	2020-03-06																		
	EPI_ISL_428823	2020-03-08																		
	EPI_ISL_428828	2020-03-08			■	■	■									■	■	■		
	EPI_ISL_428829	2020-03-08																		
	EPI_ISL_420107	2020-03-09	■					■												
	EPI_ISL_420108	2020-03-10																		
	EPI_ISL_420109	2020-03-11																		
	EPI_ISL_420110	2020-03-11																		
	EPI_ISL_420111	2020-03-12																		
	EPI_ISL_422430	2020-03-13			■	■	■			■						■	■	■	■	
	EPI_ISL_428831	2020-03-13							■		■			■	■					
	EPI_ISL_428824	2020-03-16			■	■	■						■			■	■	■		■
	EPI_ISL_428825	2020-03-16			■	■	■						■			■	■	■		■
	EPI_ISL_428826	2020-03-17							■					■						
	EPI_ISL_428827	2020-03-17	■	■				■				■								
	EPI_ISL_422431 ▲	2020-03-31			■	■	■						■			■	■	■		■
	EPI_ISL_422432 ▲	2020-03-31			■	■	■						■			■	■	■		■
	EPI_ISL_422433 ▲	2020-03-31							■					■						
	EPI_ISL_422434 ▲	2020-03-31							■					■						
	EPI_ISL_422435 ▲	2020-03-31							■					■						
	EPI_ISL_428840 ▲	2020-04-07							■					■						
	EPI_ISL_428832 ▲	2020-04-08							■					■						
	EPI_ISL_428833 ▲	2020-04-08							■					■						
	EPI_ISL_428834 ▲	2020-04-08							■					■						
	EPI_ISL_428835 ▲	2020-04-08							■					■						
	EPI_ISL_428836 ▲	2020-04-08							■					■						
	EPI_ISL_428837 ▲	2020-04-08							■					■						
	EPI_ISL_428838 ▲	2020-04-08							■					■						
	EPI_ISL_428839 ▲	2020-04-08							■					■						
	EPI_ISL_428841 ▲	2020-04-08							■					■						
	EPI_ISL_428842 ▲	2020-04-08							■					■						
	EPI_ISL_428843 ▲	2020-04-08							■					■						
	EPI_ISL_428844 ▲	2020-04-08							■					■						
	EPI_ISL_428845 ▲	2020-04-08							■					■						
	EPI_ISL_428846 ▲	2020-04-15							■					■						
	EPI_ISL_428847 ▲	2020-04-15							■					■						
	EPI_ISL_428848 ▲	2020-04-15							■					■						
	EPI_ISL_428849 ▲	2020-04-15							■					■						
	EPI_ISL_428850 ▲	2020-04-15							■					■						
Thailand	EPI_ISL_403962	2020-01-08																		
	EPI_ISL_403963	2020-01-13																		
	EPI_ISL_430837	2020-03-03			■	■	■						■			■	■	■		■
	EPI_ISL_429164	2020-03-13	■	■				■				■								
	EPI_ISL_429165	2020-03-13	■	■				■				■								
	EPI_ISL_429166	2020-03-13	■	■				■				■								
	EPI_ISL_429167	2020-03-13	■	■				■				■								
	EPI_ISL_429168	2020-03-17	■	■				■				■								
	EPI_ISL_429169	2020-03-17			■	■	■						■			■	■	■		■
	EPI_ISL_429170	2020-03-17	■	■				■				■								
	EPI_ISL_430841	2020-03-17	■	■				■				■								
	EPI_ISL_423043	2020-03-18			■	■	■						■			■	■	■		■
	EPI_ISL_429171	2020-03-18	■	■				■				■								
	EPI_ISL_430842	2020-03-18			■	■	■			■						■	■	■	■	
	EPI_ISL_423040	2020-03-20	■	■				■				■								
	EPI_ISL_423039	2020-03-21			■	■	■						■			■	■	■		■
	EPI_ISL_423041	2020-03-21			■	■	■			■						■	■	■	■	
	EPI_ISL_429172	2020-03-21			■	■	■						■			■	■	■		■
	EPI_ISL_429173	2020-03-21	■	■				■				■								
	EPI_ISL_423042	2020-03-24			■	■	■			■						■	■	■	■	
	EPI_ISL_429174	2020-03-25			■	■	■			■						■	■	■	■	
	EPI_ISL_429175	2020-03-28							■					■						
	EPI_ISL_429176	2020-03-28	■	■				■				■								
	EPI_ISL_429177	2020-03-28	■	■				■				■								
	EPI_ISL_429178	2020-03-28			■	■	■			■						■	■	■	■	
	EPI_ISL_429179	2020-03-28			■	■	■									■	■	■		
	EPI_ISL_429180	2020-03-28			■	■	■			■						■	■	■	■	
	EPI_ISL_429181	2020-03-28			■	■	■			■						■	■	■	■	
	EPI_ISL_429182	2020-03-28			■	■	■			■						■	■	■	■	
	EPI_ISL_429183	2020-03-28			■	■	■			■						■	■	■	■	
	EPI_ISL_429184	2020-03-28			■	■	■			■						■	■	■	■	
Vietnam	EPI_ISL_418269	2020-01-22																		
	EPI_ISL_418267	2020-01-22																		
	EPI_ISL_416427	2020-01-24	■	■				■				■								
	EPI_ISL_408668	2020-01-24	■	■				■				■								
	EPI_ISL_416429	2020-02-11	■	■				■				■								
	EPI_ISL_416430	2020-03-06			■	■	■						■			■	■	■		■
	EPI_ISL_416431	2020-03-06			■	■	■						■			■	■	■		■
	EPI_ISL_416428	2020-03-07			■	■	■						■			■	■	■		■

Fig. 2. Heatmap showing the presence of marker variants of Clades S, L, V, G, GH and GR named by GISAID [12] in SEA SARS-CoV-2 genomes. Grey denotes the presence and empty space denotes the absence of the genes listed. Isolates with symbol (▲) denotes samples collected after country border control implementation whereas isolates without symbol denotes samples collected before country border control implementation.

previous report in which local evolution of strains related to the mass gathering in Malaysia was observed [10]. We focused on mutations that have emerged multiple times and identified 22 recurrent mutations in the SEA SARS-CoV-2 genomes (Fig. 1). We also note that nearly 75% of the hits also overlap with candidate mutations, which may affect the phenotype of SARS-CoV-2 identified by van Dorp *et al.* [11]. There are 5 sites in the ORF1ab and 1 site each in the Spike and N protein that are characterized by a particularly large number of recurrent mutations (>20%). The GISAID SARS-CoV-2 portal [12] defined Clades S, G, and V, according to nucleotide variants that produce amino acid changes. These changes are located in ORF8 L84S (28144T>C); S protein D614G (23403A>G) and nsp3 G251V (26144G>T), respectively. We also mapped the strains with presence or absence of marker variants of larger clades named by GISAID [12] according to their geographical distributions (Fig. 2). Prior reports showed that the most common mutations were 8782C>T in ORF1ab and 28144T>C in ORF8 [13, 14]. ORF1ab is an orthologous gene with other human-associated betacoronaviruses, in particular SARS-CoV [15] and MERS-CoV [16], and was consistently identified as SARS-CoV-2 mutation hot spots [11]. Intriguingly, 92% of the strains isolated after country border control carried mutation at 11083G>T; otherwise this variant was hardly identified among strains before border control. Yeh & Contreras (Bull World Health Organ, https://www.who.int/bulletin/online_first/20-255752.pdf) reported that 11083G>T substitution was related to viral transmission among patients from a cruise with a 3-week quarantine period, and the study hypothesized that this mutation had taken place via RNA recombination with positive pressure. Further study is required to determine whether 11083G>T mutation plays a role in increasing the fitness in the carrier.

The current studied genomes showed phylogenetic relation with common recurrent mutations. Cluster I exhibited common recurrent mutation at 8782C>T in ORF1ab (n = 35). Forster *et al.* [17] observed that the ancestral S variant with these two mutations at 8782C>T and 28144T>C was predominantly identified in East Asia, but this variant outside of Asia was observed with striking, long mutational branch lengths. Instead of harboring the 28144T>C mutation, 15 Singaporean genomes from two different submitting laboratories were observed with a distinctive 382-nt deletion (Δ382) covering almost the entire ORF8. The variant was isolated between January 27 and March 9 of 2020, and was not identified in the subsequent isolates. As reported by Su *et al.* (bioRxiv, https://doi.org/10.1101/2020.03.11.987222), the Δ382 deletion removes the ORF8 transcription-regulatory sequence (TRS), resulted in enhanced downstream transcription of the N gene. Intriguingly, Malaysia closed its border on March 18, 2020, but the Δ382 has not been observed in its neighboring countries. As observed in other SARS-CoVs, mutations or deletions in ORF8 have been associated with viral replicative fitness in facilitating host adaption for interspecies transmission [18, 19]. Hence, core mutations (Δ382) in the S clade observed in the current study suggest an urgent need for comprehensive studies combining genomic, epidemiological and clinical data to understand the evolutionary pressure of this virus.

Cluster II was distinguished by not only the spike mutation D614G (23403A>G; G clade), but also at 241C>T, 3037C>T, and 14408C>T (n = 25). A further subcluster additionally containing mutations at 3 neighboring positions (28881G>A, 28882G>A, 28883G>A) in the N protein is also apparent within this clade. The G variant was rarely sampled in Asia but corresponded / corresponded to the most frequent variant in Europe [12]. However, there is no evidence that the increasing predominance of this mutation was caused by convergent selection or bottleneck events. In the current study, G variants were identified in strains predominantly from Thailand, followed by Singapore and Vietnam. Cluster III belonged to clades outside the reported S, G and V clades, and was distinguished by shared mutations at 6312C>A, 11083G>T, 13730C>T and 19524C>T in ORF1ab, 23929C>T in spike and 28311C>T in the N protein (n = 38). This variant was beginning to be observed in strains from Malaysia, the Philippines and Singapore from mid-March onwards. On April 19, 2020, a new cluster emerged from students returning to Malaysia from Indonesia [20]. Therefore, it will be valuable to include strains from Indonesia as well as other SEA countries in the future for endemic genetic variation surveillance. Similar to other SARS-CoVs, the spike protein and its host receptor have been the key targets for drug and vaccine development because the protein is relatively more conserved and is critical for viral infection [21]. In the current study, each of the clusters was identified with one mutation at the spike protein. Further analysis with PROVEAN [22] prediction of the amino acid variants showed that mutations at 23403A>G as well as 23929C>T in the spike protein were deleterious/damaging to protein function (PROVEAN score ≤ -2.5) (Supplementary Figure). Notably, both spike mutations were observed only from March 8th onwards, suggesting the emergence of positive selection. All Philippine strains carried 23929C>T mutations although this mutation was observed only after border control was implemented in Malaysia and Singapore. Nevertheless, the impact of these mutations towards the virus affinity to host receptor remains to be elucidated. There has been a lack of association between genotypes and clinical presentations of COVID-19 [23]. Hence, this characterization of SARS-CoV-2 variants could lead to better treatment strategies in the future.

The evolution and transmission of the SARS-CoV-2 is potentially affected by distinctive travel histories, founder events, host characteristics as well as geographical and climate factors. It is prudent to consider the possibility that mutational variants might influence the virus spread and subsequently the clinical presentation and outcome. Therefore, the described core mutations and phylogenetic classification in this work may provide information regarding outbreak control as well as evaluating the clinical and epidemiological outcomes of SARS-CoV-2 infection.

Methodology

A phylogenetic tree was constructed using 142 complete genome sequences from the South East Asia countries plus the reference genome Wuhan-Hu-1 (MN908947.3). The SARS-CoV-2 genomes were obtained from the GISAID (https://www.gisaid.org/), as of April 30, 2020. Genome sequence alignment was performed using

Multiple Sequence Comparison by Log-Expectation (MUSCLE) and the phylogenetic tree was inferred by using the neighbor-joining method and Tamura-Nei model with 1,000 bootstrap replicates in Geneious software (version 2020.1; Biomatters Ltd.). The sequences were also aligned to the reference genome. Sites were masked in the first 130bp and last 50 bp, as were other ambiguous positions following the protocol advocated by van Dorp *et al.* [11] and De Maio *et al.* (http://virological.org/t/issues-with-sars-cov-2-sequencing-data/473, accessed May 12, 2020). Estimation of the best fitting substitutions in nucleotide and amino acid were also detected within Geneious. We only considered variants at a site if present with frequency ≥5% [24, 25].

Data Availability

The genome sequences of the SARS-CoV-2 used in this study are available from the GISAID database (https://www.gisaid.org/), upon registration.

Conflict of Interest

The authors have no financial conflicts of interest to declare.

Acknowledgments

We gratefully acknowledge the authors from originating and submitting laboratories of sequence data on which the analysis is based.

References

1. Sirkeci I, Yüceşahin MM. 2020. Coronavirus and migration: Analysis of human mobility and the spread of COVID-19. *Migrat. Lett.* **17:** 379-398.
2. Southeast Asia Covid-19 Tracker. 2020. Available from: https://www.csis.org/programs/southeast-asia-program/southeast-asia-covid-19-tracker-0. Accesed May 12, 2020.
3. World Health Organisation. 2020. WHO emphasizes on agile response capacities, as South-East Asia Region confirms more COVID-19 cases. Available from: https://www.who.int/southeastasia/news/detail/04-03-2020-who-emphasizes-on-agile-response-capacities-as-south-east-asia-region-confirms-more-covid-19-cases. Accessed May 18, 2020.
4. TRIANGLE in ASEAN International Labour Organisation. Available from: https://www.ilo.org/asia/projects/WCMS_428584/lang--en/index.htm. Accessed May 20, 2020.
5. Today Online. 2018. Amid cross border tensions, Malaysians working in Singapore hope rice bowls unaffected. Available from: https://www.todayonline.com/world/amid-cross-border-tensions-malaysians-working-singapore-hope-rice-bowls-unaffected. Accessed May 20, 2020.
6. The Jakarta Post. 2020. High influx of Indonesian migrant workers from Malaysia, despite call to stay amid outbreak. Available from: https://www.thejakartapost.com/news/2020/04/23/high-influx-of-indonesian-migrant-workers-from-malaysia-despite-call-to-stay-amid-outbreak.html. Accessed 23 April, 2020.
7. The Center for Migration Studies. The Stateless Rohingya in Thailand. Available from: https://cmsny.org/the-stateless-rohingya-in-thailand/. Accessed May 20, 2020
8. The Guardian. 2020. Malaysia cites Covid-19 for rounding up hundreds of migrants. Available from: https://www.theguardian.com/global-development/2020/may/02/malaysia-cites-covid-19-for-rounding-up-hundreds-of-migrants. Accessed May 2, 2020.
9. Moya A, Holmes EC, Gonzalez-Candelas F. 2004. The population genetics and evolutionary epidemiology of RNA viruses. *Nat. Rev. Microbiol.* **2:** 279-288.
10. Chong YM, Sam IC, Ponnampalavanar S, Syed Omar SF, Kamarulzaman A, Munusamy V, *et al.* 2020. Complete Genome Sequences of SARS-CoV-2 Strains Detected in Malaysia. *Microbiol. Resour. Announc.* **9:** e00383-20.
11. van Dorp L, Acman M, Richard D, Shaw LP, Ford CE, Ormond L, *et al.* 2020. Emergence of genomic diversity and recurrent mutations in SARS-CoV-2. *Infect. Genet. Ecol.* **83:** 104351.
12. GISAID. 2020. GISAID_hCoV-19_Analysis_Update_2020-06-12_1900UTC. Available from: https://www.gisaid.org/. Accessed June 16, 2020.
13. Khailany RA, Safdar M, Ozaslan M. 2020. Genomic characterization of a novel SARS-CoV-2. *Gene Rep.* **19:** 100682.
14. Wang C, Liu Z, Chen Z, Huang X, Xu M, He T, *et al.* 2020. The establishment of reference sequence for SARS-CoV-2 and variation analysis. *J. Med. Virol.* **92:** 667-674.
15. Lau SK, Woo PC, Li KS, Huang Y, Tsoi HW, Wong BH, *et al.* 2005. Severe acute respiratory syndrome coronavirus-like virus in Chinese horseshoe bats. *Proc. Natl. Acad. Sci. USA* **102:** 14040-14045.
16. Kim JI, Kim YJ, Lemey P, Lee I, Park S, Bae JY, *et al.* 2016. The recent ancestry of Middle East respiratory syndrome coronavirus in Korea has been shaped by recombination. *Sci. Rep.* **6:** 18825.
17. Forster P, Forster L, Renfrew C, Forster M. 2020. Phylogenetic network analysis of SARS-CoV-2 genomes. Proc. *Natl. Acad. Sci. USA* **117:** 9241-9243.
18. Muth D, Corman VM, Roth H, Binger T, Dijkman R, Gottula LT, *et al.* 2018. Attenuation of replication by a 29 nucleotide deletion in SARS-coronavirus acquired during the early stages of human-to-human transmission. *Sci. Rep.* **8:** 15177.
19. Lau SK, Feng Y, Chen H, Luk HK, Yang WH, Li KS, *et al.* 2015. Severe Acute Respiratory Syndrome (SARS) coronavirus ORF8 protein is acquired from SARS-Related coronavirus from greater horseshoe bats through recombination. *J. Virol.* **9:** 10532-10547.
20. The Star. 2020. New Covid-19 cluster emerges from students returning from Indonesia, says Health DG. Available from: https://www.thestar.com.my/news/nation/2020/04/19/new-covid-19-cluster-emerges-from-students-returning-from-indonesia-says-health-dg. Accessed April 19, 2020.
21. Du L, He Y, Zhou Y, Liu S, Zheng BJ, Jiang S. 2009. The spike protein of SARS-CoV--a target for vaccine and therapeutic development. *Nat. Rev. Microbiol.* **7:** 226-236.
22. Choi Y, Chan AP. 2015. PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics* **31:** 2745-2747.
23. Jin X, Lian JS, Hu JH, Gao J, Zheng L, Zhang YM, *et al.* 2020. Epidemiological, clinical and virological characteristics of 74 cases of coronavirus-infected disease 2019 (COVID-19) with gastrointestinal symptoms. *Gut* **69:** 1002-1009.
24. Gire SK, Goba A, Andersen KG, Sealfon RS, Park DJ, Kanneh L, *et al.* 2014. Genomic surveillance elucidates Ebola virus origin and transmission during the 2014 outbreak. *Science* **345:** 1369-1372.
25. Holmes EC, Dudas G, Rambaut A, Andersen KG. 2016. The evolution of Ebola virus: Insights from the 2013-2016 epidemic. *Nature* **538:** 193-200.