JMID/
Journal of Microbiology and Infectious Diseases

2022; 12 (1):139-147
doi: 10.5799/jmid.1218631

RESEARCH ARTICLE

Epidemiological Features and Phylogeny of SARS-CoV-2 Circulating in the Southeast Asia in Early Pandemic

Oktaviani Naulita Turnip[1], Chairunisa Fadhilah[2], Anwar Rovik[3], Ayu Rahayu[3,4]

[1]Department of Microbiology, Faculty of Medicine, Universitas Palangka Raya, Indonesia
[2]Biomedical Science, Faculty of Medicine, Universitas Indonesia, Jakarta, Indonesia
[3]Center of Tropical Medicine, Faculty of Medicine, Public Health, and Nursing, Universitas Gadjah Mada, Yogyakarta, Indonesia
Department of Microbiology, Faculty of Medicine, Public Health, and Nursing, Universitas Gadjah Mada, Yogyakarta, Indonesia

ABSTRACT

Objectives: This study aims to understand the epidemiological features and level of genetic similarity in the SARS-CoV-2 from different geographical areas in The Southeast Asia Region during an early pandemic.

Methods: The data on COVID-19 cases in Southeast Asia was collected from https://worldometer.info/ and extracted independently. Complete genome SARS-CoV-2 nucleotide sequence data was obtained from GISAID and NCBI online platforms. The sequences were aligned using MEGA X software and identified RdRp and Spike genes using UGENE software. The phylogenetic was constructed using MEGA X software to know the similarity of these genes among isolates in the Southeast Asia region.

Results: The result showed that the first case in Southeast Asia was reported in January 2020. The highest number of COVID-19 cases and death were reported from populous and suffering countries. The phylogenetic results showed an identical solid (100%) among isolates, except for the Philippines-5 isolate. The Wuhan-Hu-1 (China) SARS-CoV-2 isolate (Acc. NC_045512) was transmitted to other countries in Southeast Asia region with various mutations in the spike protein.

Conclusion: During the early pandemic, all countries in the Southeast Asia region reported COVID-19 cases. Indonesia became the country with the highest number of COVID-19 cases and deaths. The level of similarity of the RdRp gene in the SARS-CoV-2 in Southeast Asia is higher than the Spike genes. *J Microbiol Infect Dis 2022; 12(4):139-148*.

Keywords: Epidemiology, Phylogeny, SARS-Cov-2, Pandemic

INTRODUCTION

Wuhan city, with a population of more than 11 million, is one of China's most populous cities. Wuhan was the epicenter of the Coronavirus Disease 2019 (COVID-19), which is caused by SARS-CoV-2 infection [1]. The cases were traced back to a wholesale seafood store that offered a range of live animals. During the first four months of the pandemic, it expanded to over 210 nations worldwide, making Asia the first epidemic hub, followed by Europe and the Americas. Over 80 million COVID-19 cases were reported in a year, resulting in almost 2 million deaths. Since the virus is transmitted from human to human [2,3,4], it spreads quickly in crowded areas. Therefore, these numbers are still increasing rapidly for the incoming month and year. In addition, infectious diseases may damage countries via cross-border transmission. Since its breakout, COVID-19 has expanded from China to Southeast Asia in months, causing primary concern. As of December 11th, 2020, 1,291,859 confirmed cases of COVID-19 have

Correspondence: Dr. Ayu Rahayu, Department of Microbiology, Faculty of Medicine, Public Health, and Nursing, Universitas Gadjah Mada, Yogyakarta, Indonesia
E-mail : arahayu.id@gmail.com
Received: 13 May 2022 Accepted: 01 December 2022

been reported in the Southeast Asia region, with a total of 29,761 deaths [5].

As of October 2020, more than 120,000 SARS-CoV-2 genome data have been shared online [6]. It has allowed researchers to look at the zoonotic origins and patterns of SARS-CoV-2's global spread, the development of vaccine candidates, surveillance, and the discovery of novel genetic variants. SARS-CoV-2 and other RNA viruses exhibit high mutation rates linked to increased virulence and evolvability. The previous study estimated that general coronavirus mutation rates were 2x10-6 [7]. Globally, SARS-CoV-2 mutates at roughly 23.6 mutations/per year [8, 9]. In addition, many changes and deletions were found in the genomes of distinct SARS-CoV-2 strains collected from diverse locations, indicating a rapid development of the virus during the present pandemic [10].

The phylogenetic tree provides a clear picture of the evolution of the variants, even if it lacks evidence that some variants are less common. Therefore, understanding the phylogeny of SARS-CoV-2 by analyzing its sequence from different geographical areas could provide insight into how these mutations enable variable transmission of SARS-CoV-2 in different parts of the world. Considering the importance of public knowledge about the SARS-CoV-2 epidemiologic and phylogenetic analysis, we aimed to investigate and track the SARS-CoV-2 virus in The Southeast Asia region during the initial outbreak.

METHODS

Epidemiological of COVID-19 cases in the Southeast Asia region

The cumulative data of COVID-19 confirmed cases and death in the Southeast Asia region (Brunei Darussalam, Indonesia, Malaysia, Singapore, Myanmar, Laos, Cambodia, Thailand, Philippines, Viet Nam, and East Timor) as of December 12th, 2020 [5]. The supporting data were extracted independently using SPSS to analyze the cumulative cases, the correlation between population and cases, correlation cases, and the number of death [5].

SARS-CoV-2 Genome Sequences

The SARS-CoV-2 sequences were retrieved from The Global Initiative on Sharing Avian Influenza Data (GISAID) and The National Center of Biotechnology Information (NCBI), an online platform [8, 11]. The data was filtered with the following criteria: human-hosted and whole genome. Five isolates were retrieved randomly from each country in the Southeast Asia region. In addition, the complete sequence of SARS-CoV-2 originating from Wuhan, China, was used as a reference genome.

RdRp (RNA-dependent RNA polymerase) and S (spike) specific-gene analysis among genomes

The multiple alignments obtained from whole-genome sequences were used to extract sub-alignments for each molecular target (S and RdRp genes) in Unipro UGENE v.33.0, considering the annotation of the isolate Wuhan-Hu-1 available in the National Center for Biotechnology Information [11]. Each sub-alignment was manually verified to remove sequences with Ns, unidentified positions (denoted with the IUPAC code), and gaps. Then, the collected whole-genome sequences were analyzed with Unipro UGENE software to cut the RdRp and S genes specific-site among genomes. These cut (partial) genomes were used for further analysis.

Phylogenetic analysis

RdRp and S gene sequences of SARS-Cov-2 circulating in the Southeast Asia region were aligned (MSA) using the MEGA X software package. The phylogenetic tree was constructed using a neighbor-joining (NJ) method with a 1,000 bootstraps replication.

RESULTS

The World Health Organization (WHO) declared the COVID-19 pandemic on March 11th, 2020. In the Southeast Asia region, Thailand reported the first case of COVID-19 in early January 2020, followed by Viet Nam, Malaysia, Cambodia, and the Philippines. COVID-19 transmission was reported in Indonesia, Brunei Darussalam, East Timor, Myanmar, and Laos in the next two months (Figure 1). As of December 11th, 2020, 1,291,859 confirmed cases of COVID-19 have been reported in the ASEAN region, with a total of 29,761 deaths [5]. Indonesia has recorded the highest cumulative COVID-19 cases in Southeast Asia, i.e., 611,631 confirmed cases with 18,653 deaths. The Philippines has the second-highest COVID-19 of 4448,331 confirmed cases. COVID-19 cases in Southeast Asia are increasing rapidly

in early 2020, especially in Indonesia, the Philippines, and Cambodia.

Meanwhile, Myanmar and Malaysia reported an increased number by September 2020. Singapore expressed its readiness against the COVID-19 pandemic since the cumulative cases were stable eight months after the first recorded case. East Timor and Laos reported the lowest cases in Southeast Asia, i.e., 31 and 41 COVID-19 confirmed cases, respectively, from January to December 2020 (Figure 2).

SARS-CoV-2 is highly transmittable from human to human. Therefore, it suggested that the COVID-19 pandemic may hardly hit the densest area or country. Figure 3 shows a positive correlation between cumulative COVID-19 cases and populations. COVID-19 infection resulted in mild to severe symptoms or even deaths. The Southeast Asia region reported 29,761 death during the year. The highest number of deaths recorded is from Indonesia, with 18,653 deaths, followed by the Philippines and Myanmar, with 8,730 and 2,220 deaths, respectively. Meanwhile, Cambodia, Laos, and East Timor reported no mortality cases in 2020. Figure 4 shows a strong positive correlation between cumulative COVID-19 cases and deaths.

A phylogenetic study based on the RdRp gene placed the circulating SARS-CoV-2 isolate in Southeast Asia in the same clade. The sequence analysis showed an identical solid (100%) among isolates, except for the Philippines-5 isolate (Figure 5). However, the study of S-gene-based phylogeny showed a different result. Most of the Philippines isolates and one isolate from Singapore were placed in different clades with other ASEAN's SARS-Cov-2 isolates, although they were 99% identical (Figure 6). It concluded that the Wuhan-Hu-1 (China) SARS-CoV-2 isolate (Acc. NC_045512) was transmitted into other countries in Southeast Asia region with various mutations in the spike protein.

DISCUSSION

The SARS-CoV-2 reportedly originated in Wuhan, China, and has spread worldwide since late 2019. In less than three months, the Asian region has become an epidemic center. The population of Southeast Asia accounts for about 8.54% of the world's population. Indonesia records the most cases and deaths due to the COVID-19 outbreak in Southeast Asia. SARS-CoV-2 shows a high potential for transmission due to its highly contagious nature [12].

In Southeast Asia, Indonesia and the Philippines are the two most populous countries, with 273,523,615 and 109,581,078 people, respectively, while Brunei Darussalam has fewer inhabitants of 437,479 people [5]. The results showed that Indonesia and the Philippines have the highest cumulative cases and deaths. Indonesia recorded the highest case fatality rate (CFR) of COVID-19 number of 3.049, more significant than the global CFR of 2.27. Cambodia, Laos, and East Timor have reported the minimum amount of CFR. The global CFR is disproportionate among countries. Different amounts of CFR in various countries may be due to the different situations of the outbreak, the quality of healthcare systems [13], demographic composition, and social and non-pharmaceutical interventions [14]. It indicated that many COVID-19 cases were reported from populous countries and suffering countries against the COVID-19 pandemic.

Indonesia confirmed that COVID-19 cases are rising at a very rapid rate. The characteristic of Indonesian COVID-19 patients was similar to China and Italy [15]. Indonesia is the most populous country in Southeast Asia and is included in low and middle-income countries (LMIC). It has a high rate of disease transmissions, including malaria, tuberculosis, HIV, and other tropical infections [16], as well as cardiovascular diseases, cancer, pulmonary diseases, and diabetes [17]. As in many LMICs, substantial proportions of the population face barriers to accessing quality healthcare services due to under-resourced and fragile health systems [18]. These factors may increase the cases and aggravate COVID-19 patients in Indonesia.

The analysis result demonstrated that the Southeast Asia region have a high sequence similarity based on the RdRp gene. The RNA-dependent RNA polymerase complex (RdRp) possesses proofreading capacity, implying that the mutation rate of these viruses is lower than that observed for other RNA viruses [19]. Figure 5 shows that the RdRp gene has high similarities among circulating SARS-CoV-2 isolates in Southeast Asia, compared to the reference isolate of Wuhan-Hu-1. They were

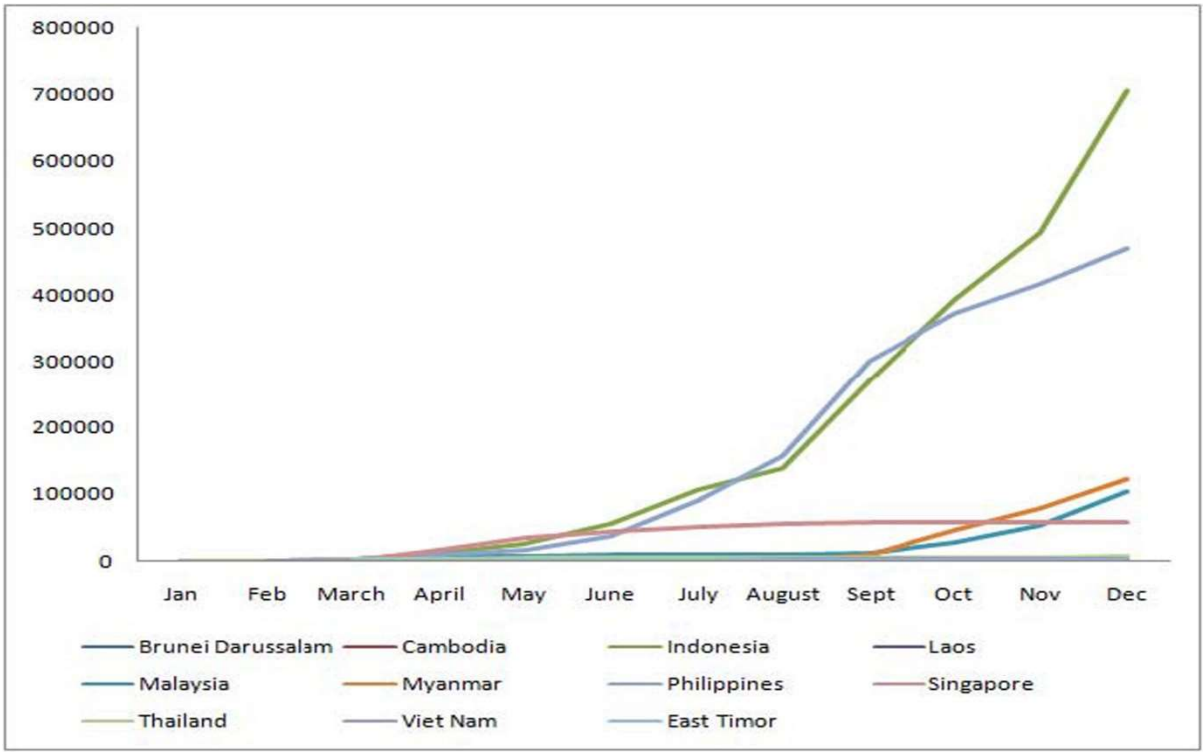

Figure 1. The COVID-19 cases in Southeast Asia (updated by December 12th, 2020, from https://worldometer.info/).

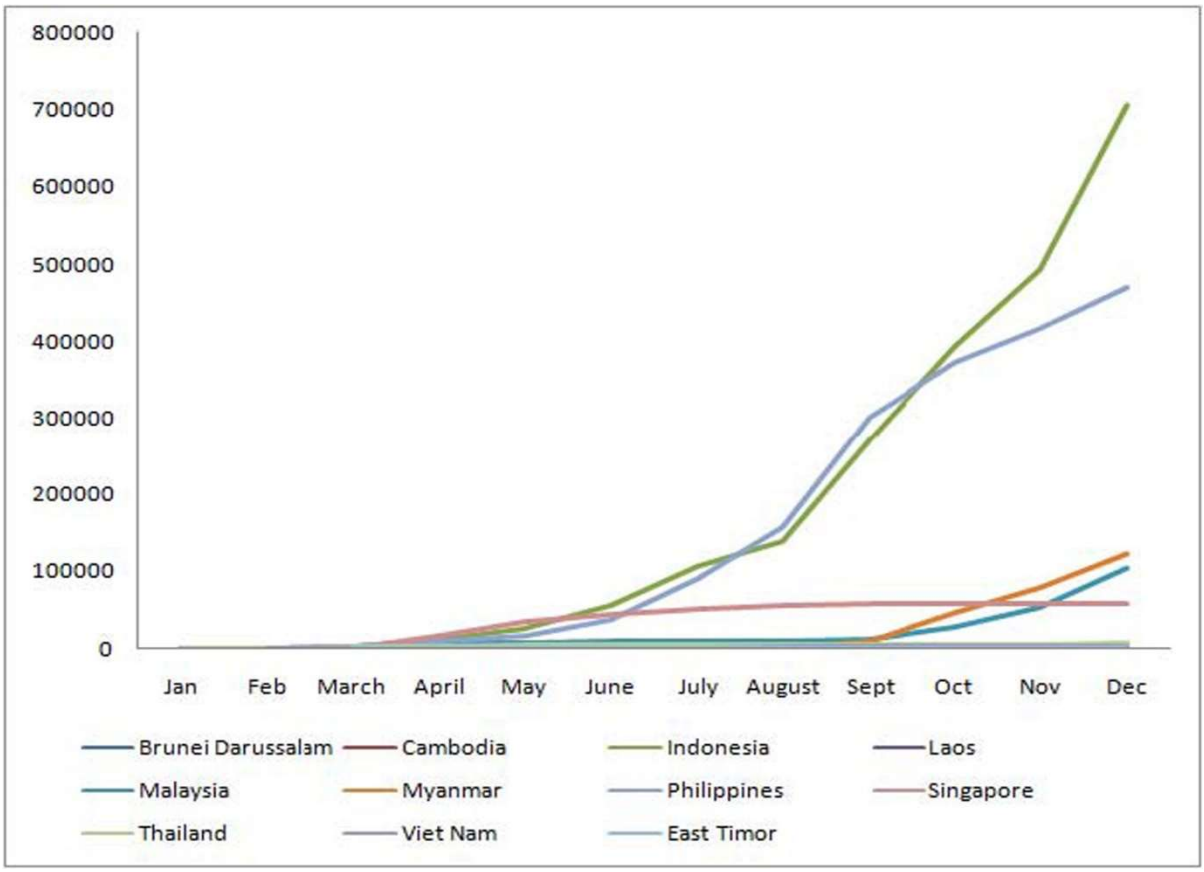

Figure 2. Cumulative number of COVID-19 cases in Southeast Asia (recorded from January to December 2020).

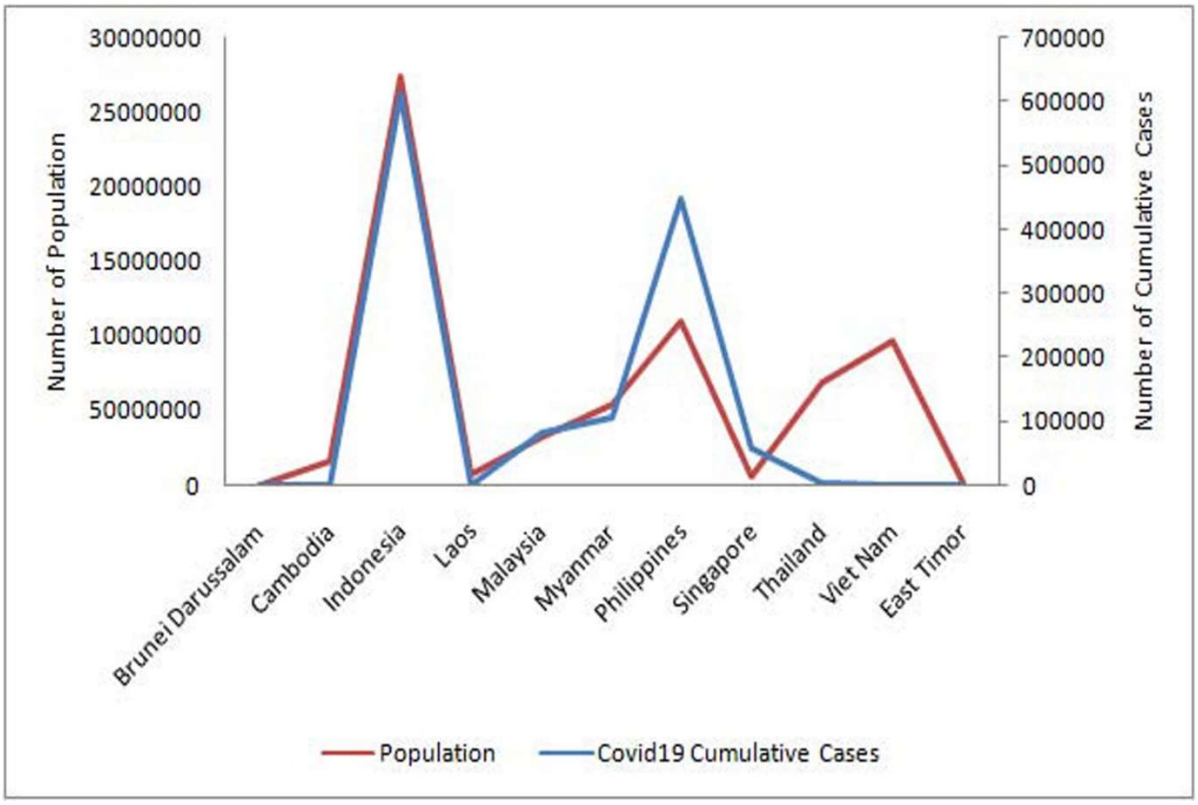

Figure 3. The correlation of cumulative COVID-19 cases to population number in the Southeast Asia (recorded from January to December 2020).

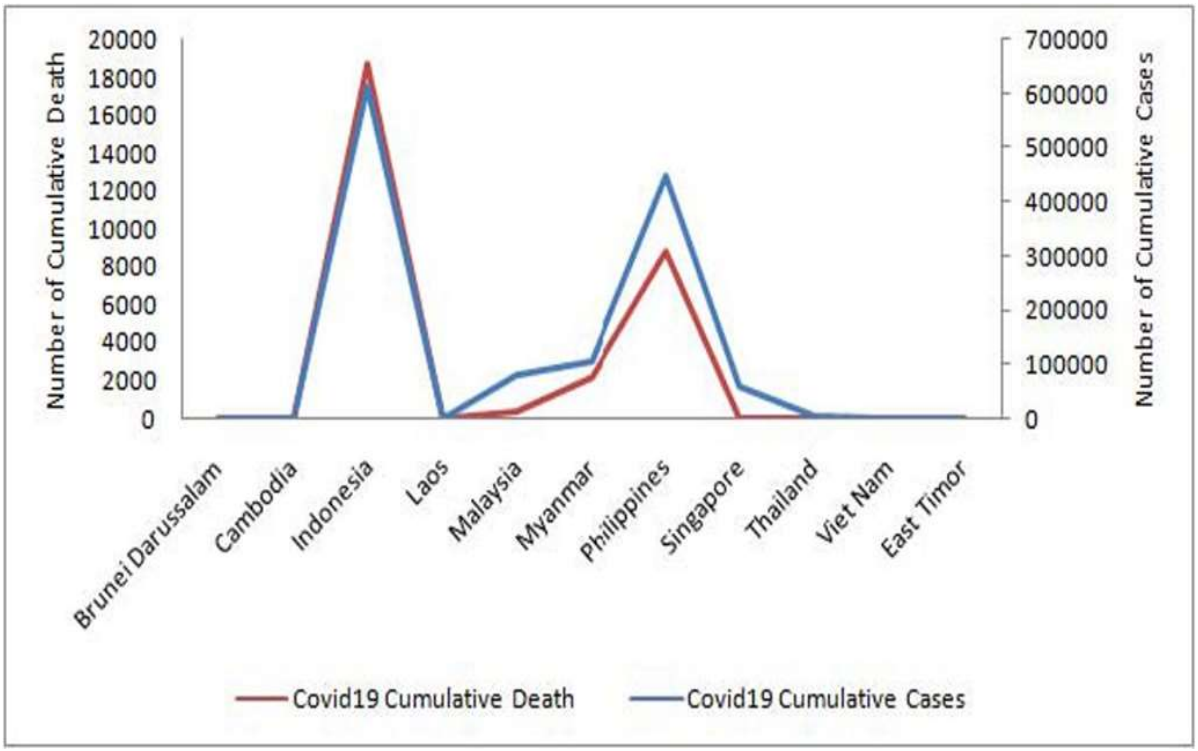

Figure 4. The correlation between cumulative COVID-19 cases and the number of deaths in Southeast Asia (recorded from January to December 2020).

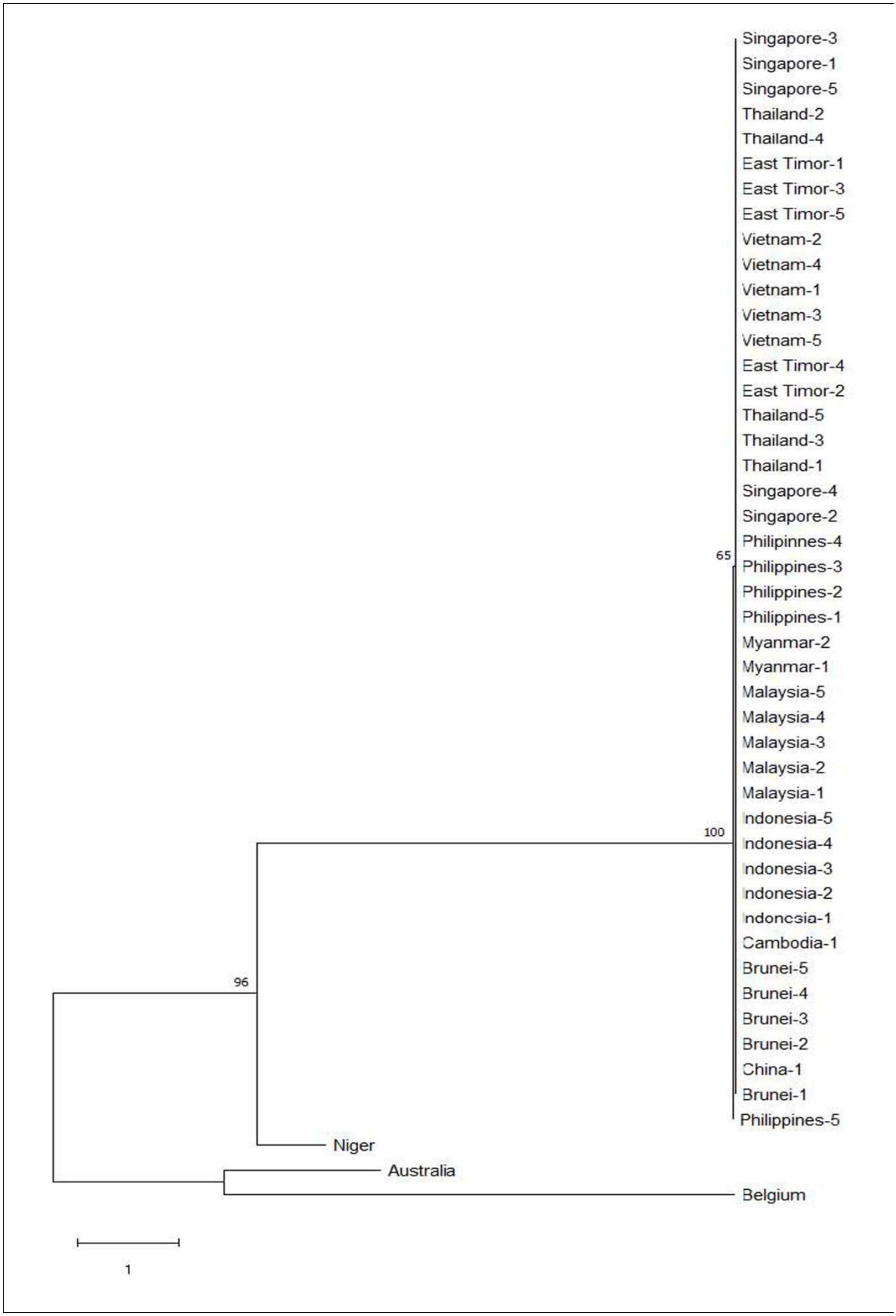

Figure 5. Phylogenetic analysis of SARS-CoV-2 circulating in Southeast Asia based on their functional protein: RNA-dependent RNA polymerase (RdRp) gene (the phylogeny of five randomly chosen isolates were constructed by Neighbor-Joining analysis with 1,000 bootstraps in MEGA X software).

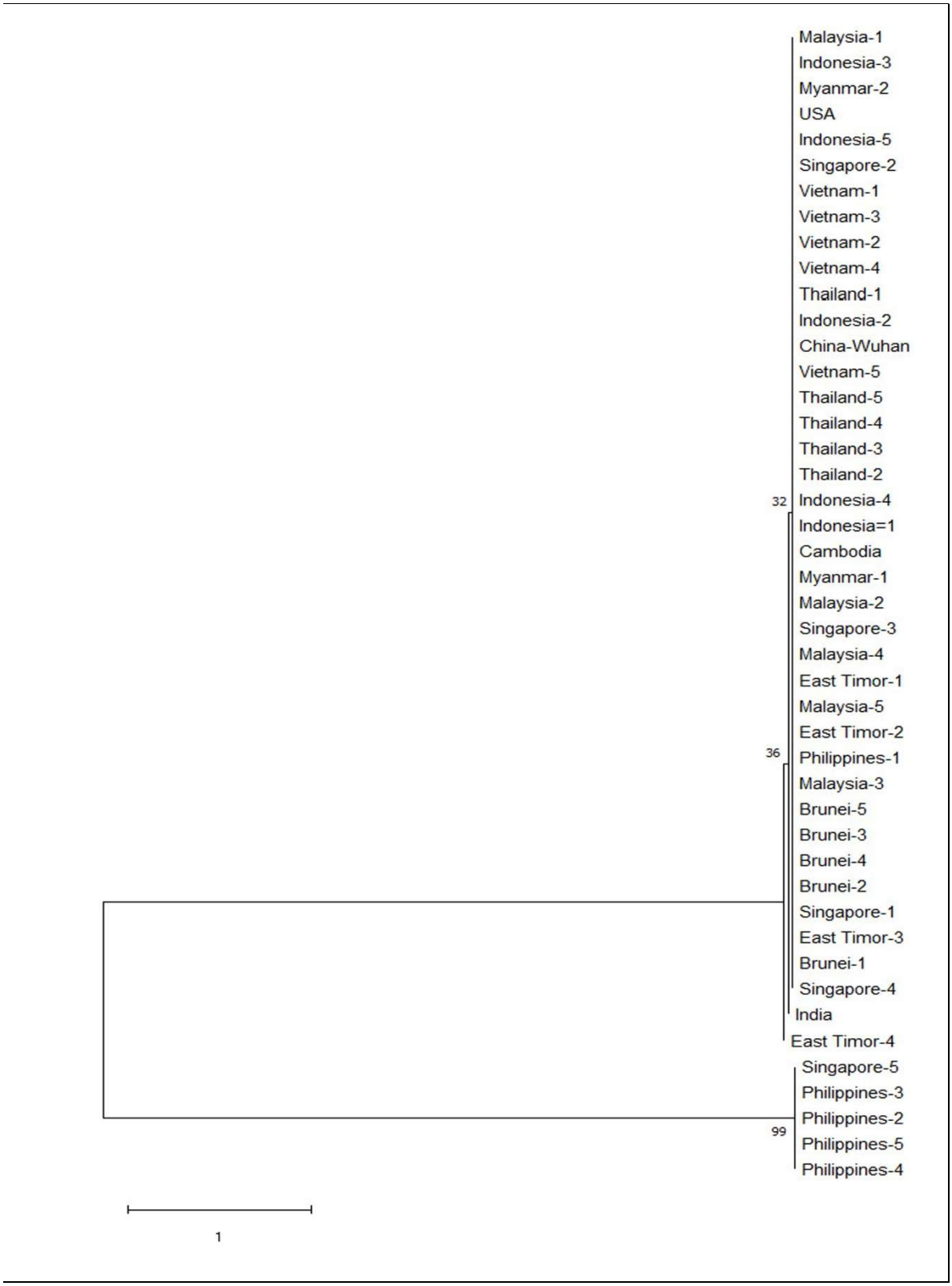

Figure 6. Phylogenetic analysis of SARS-CoV-2 circulating in Southeast Asia based on their structural protein: spike gene (S) (the phylogeny of five randomly chosen isolates were constructed by Neighbor-Joining analysis with 1,000 bootstraps in MEGA X software).

separately placed in different clades with collected SARS-CoV-2 isolates from Niger, Australia, and Belgium. Although, the sequence similarity was 96% identical. It showed that the Wuhan-Hu-1 (China) SARS-CoV-2 isolate (Acc. NC_045512) was

transmitted into other countries in Southeast Asia region with less mutation on the RdRp gene. The previous study found that Iranian SARS-CoV-2 isolates have the most similarity to China's isolate based on the RdRp gene [20]. This present study showed that the Philippines-5 isolate has less difference among others, with two nucleic acid deletions. Evaluation of the correlation of some mutations of RdRp with COVID-19 mortality rates will be clinically useful [13]. Consequently, it is essential to investigate and characterize the SARS-CoV-2 RdRp mutations to detect possible drug-resistant of SARS-CoV-2 traits.

The importance of the spike glycoprotein has been highlighted; additionally, it has been discovered that SARS-CoV-2 has a high number of mutations in its genome. The spike gene of SARS-CoV and SARS-CoV-2 viruses, which is responsible for viral entrance, mediates the binding of angiotensin-converting enzyme 2 (ACE2) to the host cell membrane. The surface glycoprotein of the SARS-CoV spike is made up of two parts: S1 and S2. SARS-CoV-2's S protein attaches to the host receptor ACE2 by its S1 subunit, which contains RBD, and then fuses the viral and host membranes via the S2 subunit, which contains the fusion peptide primed by host protease [13, 21]. Because SARS-CoV-2 identifies ACE2 as its host receptor for binding to viral S protein, it is crucial to identify the RBD in SARS-CoV-2 S protein as the most likely target for novel inhibitors, neutralizing antibodies and vaccines aimed against the virus's attachment mechanism.

Figure 6 shows that the spike-encoding gene is less conserved among circulating SARS-CoV-2 isolates in Southeast Asia compared to the reference isolate of Wuhan-Hu-1. The spike protein is the most variable genomic component of SARS-CoV [1]. This present study showed that most Philippines isolates and one isolate from Singapore were placed in different clades with ASEAN's SARS-CoV-2 isolates, although they were 99% identical. The previous study reported that the sequence of the S gene from Iranian isolates is highly similar to the sequence from Wuhan [20]. In addition, the S protein sequence of Wuhan's SARS-CoV-2 shows about 80% similarity with bat coronaviruses [22]. A mutation in the S gene is a severe clinical and public health concern. It can alter a virus's tropism, including its ability to adapt to new hosts or increase its pathogenicity. Detecting and analyzing mutations in the spike protein from various nations could provide insight into the continual shift in its structure and how these mutations enable varying transmission of SARS-CoV-2 in various regions of the world [6]. The mutation in the spike protein may be associated with higher case fatality rates [23] and its transmissibility to humans [24].

Scientists discovered that the nations hardest struck by the COVID-19 outbreak have a high fraction of the worldwide genetic diversity of SARS-CoV-2, implying widespread global transmission of SARS-CoV-2 early in the epidemic and the absence of a single initial patient in most countries and territories [13]. A study of analysis sequences is crucial for future investigations into the pathogenesis, prevention, and treatment of SARS-CoV-2 infection. Identifying genotypes connected to specific geographic locations suggests tracking the origin of variants and monitoring the transmission of SARS-CoV-2. It could be an essential tool in controlling the outbreak.

Our study has some limitations: 1) Testing capacity was low during the early pandemic. Therefore, there is a possibility of different daily testing capacities among ASEAN countries. 2) Deposited genomic data of SARS-CoV-2 from ASEAN remind low. Here, we only use five isolates from each country.

Conclusion

In the early pandemic, Indonesia recorded the highest number of confirmed COVID-19 cases and deaths in Southeast Asia. Sequence and phylogenetic analysis showed that the RdRp gene is conserved among circulating SARS-CoV-2 in Southeast Asia, while the Spike-encoding gene is less conserved.

ACKNOWLEDGMENTS

The author thanks the originating laboratories responsible for obtaining the specimens and the submitting laboratories where genetic sequence data were generated and shared via the GISAID and NCBI Initiative, on which this research is based. Also, the Laboratory of Microbiology, Faculty of Biology, Universitas Jenderal Soedirman, Jawa Tengah, for fruitful discussion.

Declaration of conflicting interest: The author(s) declare no potential conflicts of

interest concerning this article's research, authorship, and/or publication.

Financial disclosure: No financial support was received for this study

REFERENCES

1. Lu R, Zhao X, Li J, et al. 2020. Genomic characterization and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. Lancet 2020; 395: 565–574.
2. Khailany RA, Safdar M, Ozaslan M. Genomic characterization of a novel SARS-CoV-2. Gene Rep 2020; 19: 100682.
3. Zhou L, Ayeh SK, Chidambaram V, et al. Modes of transmission of SARS-CoV-2 and evidence for preventive behavioral interventions. BMC Infect Dis 2021; 21: 496.
4. Ozaslan M, Safdar M, Halil KI, Khailany RA. Practical measures to prevent COVID-19: a mini-review. J Biol Sci 2020; 20: 100–102.
5. COVID Live. Accessed by https://www.worldometers.info/coronavirus/.
6. Tordoff DM, Greninger AL, Roychoudhury P, et al. Phylogenetic estimates of SARS-CoV-2 introductions into Washington State. Lancet Reg Health Am 2021;1:100018.
7. 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):179.
8. Global Initiative on Sharing ALL Influenza Data (GISAID). Accessed by https://www.gisaid.org/.
9. Yao H, Lu X, Chen Q, et al. Patient-derived SARS-CoV-2 mutations impact viral replication dynamics and infectivity in vitro and with clinical implications in vivo. Cell Discov 2020; 6: 76.
10. Hakim MS, Annisa L, Supriyati E, et al. Current understanding of the origin, molecular biology, and continuing evolution of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). J Med Sci 2020; 52(3): 54-66.
11. The National Center of Biotechnology Information (NCBI). Accessed by https://ncbi.nlm.nih.gov/.
12. Zhu N, Zhang D, Wang W, et al. A Novel Coronavirus from Patients with Pneumonia in China, 2019. N Engl J Med 2020; 382:8.
13. Abdullahi IN, Emeribe AU, Ajayi OA, Oderinde BS, Amadu DO, Osuji AI. Implications of SARS-CoV-2 genetic diversity and mutations on pathogenicity of the COVID-19 and biomedical interventions. J Taibah Univ Med Sci 2020; 15(4): 258-264.
14. Dowd JB, Andriano L, Brazel DM, Mills MC. Demoghrapich science aids in understanding the spread and fatality rates of Covid19. PNAS 2020; 117: 9696-9698.
15. Aisyah DN, Mayadewi CA, Diva H, Kozlakidis Z, Siswanto, Adisasmito W. A spatial-temporal description of the SARS-CoV-2 infections in Indonesia during the first six months of the outbreak. PLoS ONE 2020; 15(12): e0243703.
16. Mboi N, Surbakti IM, Trihandini I, et al. On the road to universal health care in Indonesia, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet 2018; 392: 581–591.
17. Sanderson JE, Mayosi B, Yusuf S, et al. Global burden of cardiovascular disease. Heart 2007; 93(10):1175.
18. Clark A, Jit M, Gash CW, et al. Global, regional, and national estimates of the population at increased risk of severe COVID-19 due to underlying health conditions in 2020: a modeling study. Lancet Glob Health 2020; 8: e1003–17.
19. Duffy S. Why are RNA virus mutation rates so damn high? PLoS Biol 2018; 16(8): e3000003.
20. Tabibzadeh A, Zamani F, Laali A, et al. SARS-CoV-2 Molecular and Phylogenetic analysis in COVID-19 patients: A preliminary report from Iran. Infection, Genetics, and Evolution; 2020 ;84: 104387.
21. Walls AC, Park Y-J, Tortorici MA, Wall A, McGuire AT, Veesler D. Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. Cell 2020; 180: 281–292.
22. Chan JF, Kong K, Zhu Z, et al. Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan. Emerg. Microbes Infect 2020; 9.
23. Becerra-Flores M, Cardozo T. SARS-CoV-2 viral spike G614 mutation exhibits a higher case fatality rate. Int J Clin Pract 2020; 74(8): e13525.
24. Zhang H, Penninger JM, Li Y, Zhong N, Slutsky AS. Angiotensin-converting enzyme 2 (ACE2) as a SARS-CoV-2 receptor: molecular mechanisms and potential therapeutic target. Intensive Care Med 2020; 46(4): 586-590.