Original Research

A phylogenetic analysis of Indonesian SARS-CoV-2 isolates from March to December 2020: Compared with Delta and Mu variant

Arif Nur Muhammad Ansori, Yulanda Antonius
1 Faculty of Veterinary Medicine, Universitas Airlangga, Surabaya, Indonesia
2 Faculty of Biotechnology, University of Surabaya, Surabaya, Indonesia

Abstract: Indonesia’s economy and global health are endangered due to the SARS-CoV-2 outbreak. Our present study aims to examine the phylogenetic analysis of SARS-CoV-2 isolates in Indonesia and compare these isolates to other Southeast Asian countries. In the present study, we retrieved 105 isolates from GISAID EpiCoV and other isolates from GenBank, NCBI. Then, we extracted the full genome and focused on the spike (S) protein gene (3882 bp). We employed Molecular Evolutionary Genetics Analysis (MEGA X) software to construct a phylogenetic analysis using the maximum likelihood approach. Here, we demonstrated and revealed the relationship between Indonesian and other Southeast Asian SARS-CoV-2 isolates. In summary, our work presents the phylogenetic analysis of 105 isolates in Indonesia. Our study assists in monitoring the deployment of the disease spreading. Furthermore, we suggest that the genomics and epidemiological surveillance investigations on COVID-19 should be enhanced, especially in Indonesia.

Keywords: COVID-19, coronaviruses, SARS-CoV-2, phylogenetic analysis, spike protein

INTRODUCTION

The first case of SARS-CoV-2 emerged in Wuhan, China and it sporadically transmitted worldwide. The WHO declared that the infection was considered as a pandemic in March 2020. The outbreak of SARS-CoV-2 endangered the economy and global health. The problem calls for large-scale scientific research to reveal more information related to SARS-CoV-2, for instance, various aspects of the genome. In late December 2020, the virus had infected about 85 million people across the globe with more than 1.5 million global deaths. The data are supported by the CSSE, Johns Hopkins University, USA. The symptoms of COVID-19 are not much different from the symptoms of infection caused by other types of Coronaviruses (CoVs). Several regular/mild symptoms are cough and fever. Infection in the respiratory system that develops into pneumonia and exacerbates to ARD is the most severe case and results in death.

The coronaviruses (CoVs) are critical pathogenic agents that cause respiratory, neurological, gastrointestinal, and systemic diseases in humans and animals. The name “coronavirus” is derived from “corona” which reflects the appearance of the spiky outer protein cover of the virus. The coronavirus family consists of various genera, namely Gamma-, Delta-, Beta-, and Alphacoronavirus. The novel virus belongs to Betacoronavirus, a genus that has formerly caused epidemics, SARS-CoV-1 and MERS-CoV. It has a genome of 29,890 bp (GenBank NC_045512.2), similar to another class of CoVs. The CoVs
The S protein is a single-stranded RNA. E, M, N, and S, are the various types of structural proteins encoded by this viral genome. The S protein has recently emerged as a prime prospective antigen within vaccine formulation to fight SARS-CoV-2. Its main objective is to interact with host cells through the ACE2 receptor and immediately be recognized by the host immune system.

Indonesia is one of the Association of Southeast Asian Nations (ASEAN) that have reported the entire strain of SARS-CoV-2 genomes in their respective regions along with Brunei, Myanmar, Vietnam, Singapore, Malaysia, the Philippines, Thailand, Cambodia, and Laos. Presently, Laos is the ASEAN nation with the fewest whole-genome sequences of the virus published to databases, such as GenBank or GISAID. The data is essential to support the epidemiology investigation which is the notable instrument in the observation of emerging and re-emerging viruses. New concepts and insights must be applied and adopted carefully as information emerges every day at a rapid pace. Furthermore, GISAD EpiCoV recently defined the various clades for isolates originated from Indonesia, such as G, GH, GR, and so on. Therefore, due to the recent data, we unlocked the phylogenetic analysis of SARS-CoV-2 isolates in Indonesia and compared these isolates to other Southeast Asian countries.

**MATERIAL AND METHOD**

**a. SARS-CoV-2 Isolates**

All SARS-CoV-2 isolates from Indonesia were regained from the database (GenBank and GISAID EpiCoV) until December 2020. All 105 virus isolates were collected from GISAID EpiCoV. Moreover, we used the isolate of Wuhan-Hu-1 (extracted from GenBank, NCBI) as a reference, according to Ansori et al. (2020).

**b. Nucleotide Sequence Preparation**

We extracted S protein gene from all isolates. MSA of the sequences were completed using MUSCLE in MEGA X software (Pennsylvania State University, USA).

**c. Phylogenetic Tree Analysis**

In the present study, we constructed the molecular phylogenetic design and visualization by employing MEGA X software on the S protein gene of all isolates with a maximum likelihood approach. In addition, the molecular phylogenetic construction was tested according to our previous study.

**RESULTS AND DISCUSSION**

The Alphacoronavirus and Betacoronavirus infect animals and humans, whereas Deltacoronavirus and Gammacoronavirus only infect animals. There were six CoVs that caused problems for humans by the end of 2019. The SARS-CoV-2, the seventh CoVs, is the most recent identified CoV strain which was detected in Wuhan, China in December 2019. It was postulated to be transmitted to humans through animals in the live animal markets in Wuhan, China. Up until today, based on the CSSE, Johns Hopkins University, USA, more than 100 million people worldwide have been infected by this novel virus.

The S protein mediates the access of SARS-CoV-2 through the membrane fusion of human cells and has become the main purpose for few researches of vaccines and antiviral drugs. Both S1 and S2 domains of the S protein are part of the novel virus that are essential for infection. In brief, the S1 domain is the most essential domain for binding into cellular receptors of the host. The efficacy of various remedies, including fusion blockers, disrupting protease inhibitors, neutralizing antibodies, S protein inhibitors, small RNAs, and ACE2 blockers show that in vitro researches are unacceptable. Various tools have been performed to generate vaccines by employing the S protein
Therefore, in this study we focus on the S protein of the novel virus isolates in Indonesia.

As new data on the novel virus is issued rapidly, updated theories and arrangements should be persistently adopted. Recently, the database established various clades of SARS-CoV-2, such as S, G, V, and so on. In general, viruses have a much higher mutation rate than prokaryotes and eukaryotes. Viruses with a genome in the form of RNA have a high mutation rate, about one million times higher compared to their host, thus increasing their virulence. The mutation rate for CoVs is estimated at $4 \times 10^{-4}$ nucleotide substitutions/site/year. Therefore, it can be said that the novel virus has high mutation rate. Nonetheless, its mutation rate increases the potency of zoonotic viral pathogenicity for human-to-human transmission and it might be more virulent.

Phylogenetic analysis is an analysis which is commonly used for targeting both the fundamental and applied issues of virology, including evolution, taxonomy, diagnostics, phylogeography, origin, and epidemiology. It might supply an overview of the virus evolution, which can be investigated to know the cluster of viruses. In this study, we demonstrated the relation of 105 Indonesian SARS-CoV-2 isolates to another Southeast Asian SARS-CoV-2 isolates from humans, bats, mink, and pangolin (Figure 1). Interestingly, we found that based on the viral S protein gene isolated from Indonesia and various other countries, there was not much difference in them. Previously, an investigation of molecular phylogenetic analysis of Indonesian isolates was established by Arif Nur Muhammad Ansori et al. (2020). However, it still has limitation regarding to the number of isolates.

Our study demonstrated that the other CoVs isolated from humans, such as CoV-229E, -NL63, and -HKU1 are reflected as a pathogen lead to upper respiratory infection and conscientious for more than 15% of the common cold. The HCoV-229E replicates within the upper respiratory tract epithelial cells. Unlike SARS-CoV that spreads from the upper airway and causes a severe lower respiratory infection. Moreover, we also used another CoVs sample originated from bats, namely CoV-HKU4-1, -ZC45, -YN2018D, -ZXC21, and so on. In addition, the results of a five-year study in twenty countries in three continents found that bats harbor a high number of probably zoonotic CoVs. Since the emergence of SARS-CoV, it is known that various animal-borne CoVs have mutated and made the leap to humans, causing severe infections. All of SARS-CoV-2, HCoV-229E, HCoV-NL63, MERS-CoV, and SARS-CoV are believed to originate from bats, whereas HCoV-OC43 and HCoV-HKU1 are believed to derive from rodents. Additionally, all seven CoVs that cause human diseases have crossed the species barrier, as the progenitor viruses are found in different host animals.

In recent findings, it is known that CoVs exists in several wild animals in Asia, especially mammals. Thus, research related to this matter is very important in order to investigate the possible host roles of this new virus. Pangolin CoV derived from Malayan pangolin or Manis javanica, is 91.02% identical to the novel virus in the whole-genome level. Hitherto, another study mentioned that the novel virus shares 96% of the whole genome with a BatCoV RaTG13 isolated from China. Snakes are also considered as feasible virus reservoir to human infection. Minks and bats are also prospective hosts of SARS-CoV-2. Additionally, there is a possibility that domesticated animals could serve as an intermediate host and enable the transmission of the virus from their natural reservoirs to humans. Moreover, there are various intermediate hosts involved in virus transmission, such as camelids for human CoV-229E, dromedary camels for MERS-CoV, and civets for SARS-CoV. Therefore, we suggest that supporting surveillance researches should be conducted on minks,
pangolins, bats, and other mammals in wild habitats, remarkably in East Asia, in order to understand the risk of forthcoming zoonotic diseases.

**Figure 1.** Phylogenetic tree showing the close relation of SARS-CoV-2 isolates in Indonesia to the Southeast Asian SARS-CoV-2 isolates and other CoVs (bat, human, mink, and pangolin).

Outbreak of the novel virus has led to economic and medical emergency worldwide. Thus, unlocking the features of the novel virus genome and
establishing the procedures to observe the novel virus during the pandemic is a crucial move for monitoring the COVID-19 pandemic.\textsuperscript{48,49} Genomic data should be used to monitor and track the spread of the novel virus. This is also related to the recognition of genotypes related to temporal infectious clusters and specific geographics.\textsuperscript{10}

Furthermore, recent finding showed that another variant of SARS-CoV-2 had spread in Indonesia. The new variant which termed as B.1.617.2 variant or Delta variant has become the dominant variant to cause 78.8\% cases. It was detected on April 2021 in Indonesia and transmitted widely in countrywide.\textsuperscript{50} In brief, the lineage of Delta variant was consisted of various mutation in N-terminal domain (NTD). Therefore, those mutation positions were stated as B.1.617.1, B.1.617.2, and B.1.617.3. Result demonstrated that Delta mutation had ability to be resistant for neutralization by antibodies, including anti-NTD. It was caused by inability of antibody to bind into spike protein of SARS-CoV-2.\textsuperscript{51} Based on the phylogenetic analysis in general, Delta variant or B.1617 variant was showed as derived from D614G lineage.\textsuperscript{52}

Moreover, the SARS-CoV-2 virus seems keep evolving and the new variant is detected as MU variant. It is termed as B.1.621 based on the official Phylogenetic Assignment of Named Global Outbreak lineage designation. In detail, this variant is characterized by several substitution in spike protein, such as T95I, Y144T, Y145S, 146N, R346K, E484K, and N501Y. It known for having an ability for virus to escape from immune effect.\textsuperscript{53} In Indonesia, this variant is estimated carefully in order to minimize the effect of infection. However, the specific number of infections remains unknown.

Hence, the quick finding of variation in genomic level within the investigation of the novel virus in Indonesia is urgently needed for a streamlined retort to the COVID-19 pandemic.\textsuperscript{11-13} Furthermore, recognizing unique variants of the novel virus and linking them to employ a molecular epidemiology approach might enable scientists to establish the ancestry of a unique variant and observe the virus spreading. This data might be a crucial instrument in controlling the COVID-19 pandemic.

CONCLUSION

In summary, our work presents the phylogenetic analysis of 105 isolates in Indonesia. Our study assists in monitoring the deployment of the disease spreading. Furthermore, we suggest the genomics and epidemiological surveillance investigation to be enhanced on COVID-19, especially in Indonesia.

AUTHORS’ CONTRIBUTIONS

Arif Nur Muhammad Ansori prepared the samples, designed the protocols, executed the protocols, and wrote the manuscript. Yulanda Antonius reviewed and supervised the manuscript. All authors have read and approved the final manuscript.

ACKNOWLEDGEMENT

This study is supported by the Faculty of Biotechnology, University of Surabaya, Surabaya, Indonesia. Our heartfelt sympathy to the victims of COVID-19 and tribute to the frontliners worldwide, notably in Indonesia. We thank Rr. Putri Amaristya Purwono for editing the manuscript. Additionally, Arif Nur Muhammad Ansori would specifically like to highlight the ongoing and unwavering support of Yulanda Antonius.

DATA AVAILABILITY STATEMENT

The datasets used during the present work are available in the GISAID (https://www.gisaid.org/) and GenBank, NCBI
We gratefully acknowledge the authors from originating and submitting laboratories of GISAID data in this study.

DISCLOSURE STATEMENT
The views and opinions expressed in this article are those of the authors and do not necessarily reflect the official policy or position of any affiliated agency of the authors.

REFERENCES


Arif Nur Muhammad Ansori is a Ph.D. candidate in Veterinary Science at Universitas Airlangga. He completed his B.Sc. in Biology and M.Sc. in Vaccinology and Immunotherapeutics, Universitas Airlangga. Currently, he is an awardee of PMDSU Scholarship (Batch III) at Universitas Airlangga. His research projects related to virology, bioinformatics, and molecular biology. His actual research focus is the application of molecular biology to the development of COVID-19 vaccine.

Yulanda Antonius received her B.Sc. in Biology, University of Brawijaya, Indonesia and M.Sc. in Medical Science, University of Tsukuba, Japan. She embroiled in various research about cell biology, molecular biology, and bioinformatics. Currently, she works as a lecturer at the University of Surabaya, Indonesia and concerns with the molecular biology research.