Distribution of COVID-19 and Phylogenetic Tree Construction of SARS-CoV-2 in Indonesia

Dora Dayu Rahma Turista¹, Aesthetica Islamy¹, Viol Dhea Kharisma²,³,⁴ and Arif Nur Muhammad Ansori³,⁴,⁵*

¹Department of Medical Laboratory Technology, STIKes Hutama Abdi Husada Tulungagung, Indonesia. ²Master Program in Biology, Faculty of Mathematics and Natural Sciences, Universitas Brawijaya, Malang, Indonesia. ³Division of Molecular Biology and Genetics, Generasi Biologi Indonesia (GENBINESIA) Foundation, Gresik, Indonesia. ⁴Department of Biotechnology, Faculty of Mathematics and Natural Sciences, Universitas Negeri Malang, Malang, Indonesia. ⁵Doctoral Program in Veterinary Science, Faculty of Veterinary Medicine, Universitas Airlangga, Surabaya, Indonesia.

Abstract

Coronavirus disease 2019 (COVID-19) is a disease caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). COVID-19 has spread quickly across the world and has been declared a pandemic. Indonesia has many COVID-19 cases, with a high mortality rate. This study aimed to describe the distribution of COVID-19 in Indonesia and constructed the SARS-CoV-2 phylogenetic tree from Indonesian isolates and those from other countries, including other CoVs to determine their relationship. The distribution data of COVID-19 in Indonesia were obtained from the COVID-19 Management Handling Unit and descriptively analyzed. SARS-CoV-2 isolates were retrieved from the GenBank® (National Center of Biotechnology Information, USA) and GISAID EpiCoV™ databases and were used to construct phylogenetic trees using MEGA X software. Of the 37 provinces in Indonesia, five provinces with the highest case fatality rates were DKI Jakarta, Jawa Barat, Jawa Timur, and Banten, and the five provinces with the highest cure rate were Kepulauan Riau, Bali, Aceh, Gorontalo, and DI Yogyakarta. SARS-CoV-2 Indonesian isolates were closely related to SARS-CoV-2 isolates from other countries. The rapid and widespread distribution of SARS-CoV-2 in Indonesia was caused by the lack of compliance with territorial restrictions and dishonesty with medical personnel. These data revealed that mutations can occur during the transmission process, which can be caused by a history of travel and increased patient immunity.

Keywords: COVID-19, Indonesia, Phylogenetic Tree, SARS-CoV-2

*Correspondence: arif.nma-17@fkh.unair.ac.id

(Received: May 09, 2020; accepted: May 23, 2020)
INTRODUCTION

An unknown cause of acute respiratory disease occurred in Wuhan, China on December 2019 (Wong et al., 2020). The World Health Organization (WHO) established it as an outbreak and a global public health emergency on January 2020. On February 11, 2020, WHO named the disease COVID-19, caused by SARS-CoV-2, which was previously named as 2019-nCoV (Ghinai et al., 2020; Lai et al., 2020; World Health Organization, 2020a). WHO characterized COVID-19 as a pandemic on March 2020 (World Health Organization, 2020b), and as of May 9, 2020, COVID-19 has been informed in 215 countries with 3,822,382 confirmed cases and 263,658 mortalities (World Health Organization, 2020c), showing that the transmission of SARS-CoV-2 is very rapid.

SARS-CoV-2 is transmitted through droplets or direct contact (Lai et al., 2020). SARS-CoV-2 transmission route between people is direct transmission, such as sneezing, contact transmission, droplet inhalation, coughing, including contact with the mucous membranes of the nose, eyes, and mouth (Peng et al., 2020). In the process of transmission, SARS-CoV-2 infects humans due to the recognition of receptors; specifically, angiotensin-converting enzyme (ACE2) (Gheblawi et al., 2020; Shang et al., 2020; Zhang et al., 2020). SARS-CoV-2 enters cells through the bond between the spike glycoprotein and the host receptor, namely, ACE2 (Hoffmann et al., 2020; Zhang et al., 2020). ACE2 is present in various human organs and tissues (Hamming et al., 2004; Donoghue et al., 2000; Li et al., 2020) so every acute respiratory syndrome coronavirus 2 (SARS-CoV-2; in the lungs, ACE2 is present in type II pneumocytes (Chen et al., 2010; Uhal et al., 2013).

The location of ACE2 in type 2 alveolar cells is the main cause of acute pneumonia in patients. The clinical symptoms of these patients include dry cough, fever, difficulty in breathing (dyspnea), an increase in body temperature, pneumonia, headache, a decrease in the number of white blood cells and lymphocytes, new pulmonary infiltrates visible upon chest radiography, and no noticeable improvement after medication with antibiotics for three days (Zhou et al., 2020) a large number of SARS-related coronaviruses (SARSr-CoVs. The worst outcome of SARS-CoV-2 infection is death.

On March 2, 2020, Indonesia reported the first case of SARS-CoV-2 infection in two Depok residents who had recently interacted with some Japanese citizens (Nuraini, 2020). Furthermore, COVID-19 began to emerge and was identified in various provinces in the territory of Indonesia, resulting in a high mortality rate in the country. As of May 9, 2020, Indonesia has a total of 13,645 confirmed COVID-19 cases, with 959 deaths and 2,607 patients declared cured (Worldometer, 2020; Gugus Tugas Percepatan Penanganan COVID-19, 2020). The wide distribution of COVID-19 allows mutations to occur in the transmission process. The genetic diversity of the virus little by little increases as the virus spreads; certain genome variants have been repeatedly sequenced at different geographical locations and time points (Skums et al., 2020). Phylogenetic trees are usually built to determine the evolutionary relationships between species (Horiike, 2016) it is important to choose the suitable molecule for the phylogenetic analysis of a given lineage. For example, when the evolutionary rate of the gene (or protein. This study aimed to describe the distribution of COVID-19 in Indonesia and also constructed the phylogenetic tree of Indonesian SARS-CoV-2 isolates with isolates from other countries and other CoVs to determine their relationship.

MATERIALS AND METHODS

Distribution of COVID-19 in Indonesia

The COVID-19 case data in Indonesia were collected from the Gugus Tugas Percepatan Penanganan COVID-19 database until May 8, 2020. The data were analyzed descriptively to determine the five provinces with the highest case fatality rate (CFR) and cure rate levels.

SARS-CoV-2 Isolates

The sequences of SARS-CoV-2 isolates were retrieved from the GenBank® (National Center of Biotechnology Information (NCBI), USA) and GISAID EpiCoV™ databases (Table 1).

Molecular Phylogenetic Analysis

The phylogenetic model and tree visualization were constructed based on the SARS-CoV-2 nucleocapsid protein gene using MEGA X with the maximum likelihood method.
The phylogenetic tree was validated by running the analysis on 1000 bootstrapped input datasets using the Tamura-Nei substitution model.

**RESULTS AND DISCUSSION**

**Distribution of COVID-19 in Indonesia**

Fig. 1 shows the distribution of COVID-19 cases in 34 provinces in Indonesia. From the surveillance system report conducted by the acceleration of COVID-19 countermeasures, the confirmed cases of COVID-19 have spread to all provinces. The five provinces with the highest distribution of confirmed cases were DKI Jakarta, Jawa Barat, Jawa Timur, Jawa Tengah, and Sulawesi Selatan. These five provinces also had a relatively high number of deaths. The number of cases cured in the five provinces was also considerably high. The province of Nusa Tenggara Timur (NTT) reported the least number of confirmed cases (n=12). Additionally, no deaths have been reported in NTT. Other provinces that have not reported deaths are Maluku Utara and Jambi.

On May 8, 2020, 13,122 confirmed cases were reported in Indonesia. The cases consisted of 13,091 confirmed cases from 34 provinces and an additional 21 cases that were being verified in the field. The number of reported deaths from 31 provinces was 943 (CFR=7.2%) (Gugus Tugas COVID-19, 2020), which was obtained from a COVID-19 surveillance report conducted in

![Fig. 1. Distribution of COVID-19 cases in 34 provinces of Indonesia](image-url)

**Table 1. SARS-CoV-2 and other CoV isolates from the NCBI and GISAID databases**

<table>
<thead>
<tr>
<th>Code</th>
<th>Origin</th>
<th>Host</th>
<th>Isolation Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>MT188341.1</td>
<td>USA: Minnesota (MN)</td>
<td>Homo sapiens</td>
<td>Nasopharyngeal or oropharyngeal swab</td>
</tr>
<tr>
<td>MN908947.3</td>
<td>China: Wuhan</td>
<td>Homo sapiens</td>
<td>Unknown</td>
</tr>
<tr>
<td>MN985325.1</td>
<td>USA: Washington (WA)</td>
<td>Homo sapiens</td>
<td>Oropharyngeal swab</td>
</tr>
<tr>
<td>MN988713.1</td>
<td>USA: Illinois (IL)</td>
<td>Homo sapiens</td>
<td>Sputum</td>
</tr>
<tr>
<td>MT253701.1</td>
<td>China: Zhejiang, Hangzhou</td>
<td>Homo sapiens</td>
<td>Sputum</td>
</tr>
<tr>
<td>MN994468.1</td>
<td>USA: California (CA)</td>
<td>Homo sapiens</td>
<td>Nasopharyngeal swab</td>
</tr>
<tr>
<td>EPI_ISL_414377</td>
<td>Indonesia: Jakarta</td>
<td>Homo sapiens</td>
<td>Sputum</td>
</tr>
<tr>
<td>EPI_ISL_414376</td>
<td>Indonesia: Jakarta</td>
<td>Homo sapiens</td>
<td>Sputum</td>
</tr>
<tr>
<td>EPI_ISL_414375</td>
<td>Indonesia: Jakarta</td>
<td>Homo sapiens</td>
<td>Sputum</td>
</tr>
<tr>
<td>EPI_ISL_413219</td>
<td>Indonesia: Jakarta</td>
<td>Homo sapiens</td>
<td>Oro-pharyngeal swab</td>
</tr>
</tbody>
</table>
Indonesia. If the surveillance system attributes are examined, namely the completeness of the data, the data only reaches approximately 30% for reporting positive confirmed (35%) and cured cases (39%), and 67.7% for deaths. All confirmed cases (positive, death, or recovered) should have had a reporting date, but many officials did not include a complete report date.

Indonesia has a high mortality rate due to COVID-19. The five provinces in Indonesia with the highest CFR were Banten (11.1%), Jawa Timur (10.7%), DKI Jakarta (8.6%), Jawa Tengah (7%), and Jawa Barat (6.6%). The calculation of mortality was obtained by comparing the number of deaths with the number of confirmed cases (Fig. 2). In addition, DKI Jakarta was the largest contributor to COVID-19 cases in Indonesia and was also the point of entry of SARS-CoV-2 in Indonesia. This was evident from Indonesia’s index case originating from DKI Jakarta. DKI Jakarta has the largest international airport in Indonesia, which is the entrance for foreign tourists. A positive index case occurred after contact with Japanese citizens visiting Indonesia. From DKI Jakarta, the virus spread to the surrounding provinces, especially Jawa Barat. A cluster from Jakarta spread to other areas in the vicinity due to the large number of residents from Jawa Barat who work in Jakarta (DetikNews, 2020). The main coping effort for COVID-19 was the selection of cases. Cases that were identified in the community received treatment. Treatments that have been administered in Indonesia are shown in Fig. 3. Kepulauan Riau, Bali, Aceh, Gorontalo, and DI Yogyakarta were the five provinces that reported the highest cure rates.

The probability of transmission of SARS-CoV-2 is that one positive person can transmit the virus to two to three other people, with an average transmission time of five days (DetikNews, 2020). According to epidemiology theory (Gordis, 2004), this is referred to as the $R_0$ (R-naught). COVID-19 cases in Indonesia occurred by spreading from metropolitan cities, with the main source of transmission being the Jabodetabek cluster, and then to Surabaya, which is the second largest metropolitan city in Indonesia, as well as the capital of Jawa Timur Province (Erfinanto, 2020). Metropolitan cities are considered sources of employment and education. The largest cluster of SARS-CoV-2 transmission in Jawa Timur was the Hajj Cluster in Sukolilo, Surabaya (Kurniawan, 2020). This is the main cause of the acceleration of the spread of cases in Indonesia (KompasTV, 2020; Mantra, 2000).

The mortality rate (CFR) in Indonesia is relatively high. This can be observed from Indonesia’s CFR (CFR=7.2%), which exceeds that of the world (6.9%) (Kementrian Kesehatan, 2020). The diverse character of Indonesian society was the main reason for the difficulty in preventing and controlling the spread of SARS-CoV-2. Not all Indonesian people adhered to local lockdowns or territorial restrictions, which are the main mechanisms for suppressing the spread of SARS-CoV-2. Additionally, there were many people who were sick and not honest when checking themselves into health facilities. This is precisely the source of SARS-CoV-2 transmission to medical personnel (Liputan 6, 2020).

**Phylogenetic Tree Construction of SARS-CoV-2 in Indonesia**

SARS-CoV-2 is thought to have originated from bats that migrated to humans through food in a wholesale market for seafood and wild animals.
near the city of Wuhan, China (Hui et al., 2020; Li et al., 2020). Bats of the genus *Rhinolophus* in southern China are known to have many SARS-like CoVs (Sun et al., 2020a). SARS-CoV-2 belongs to the subgenus *Sarbecovirus*, genus *Betacoronavirus*, and family *Coronaviridae* (Gorbalenya et al., 2020). SARS-CoV-2 has characteristics similar to SARS-CoV, including its genome sequence (Xu et al., 2020). SARS-CoV-2 is also similar to the Middle East respiratory syndrome-related coronavirus (MERS-CoV) (Elfiky et al., 2017). This group of viruses is genetically diverse and shows frequent recombination (Sun et al., 2020b; Zhou et al., 2020).

In the distribution process, SARS-CoV-2 may mutate, as SARS-CoV-2 is an RNA virus. The characteristic of RNA viruses is that they mutate up to 1 million times faster than their host (Pachetti et al., 2020). Phylogenetic analysis is a method to explain the evolutionary history and relationships among a group of organisms (Horiike, 2016) and is usually visualized in the form of phylogenetic trees to make it easier to determine the relationship between species. In this study, phylogenetic trees were constructed using the Indonesian SARS-CoV-2 nucleocapsid gene sequences with sequences from isolates sourced from several other countries, as well as sequences of several other types of CoVs collected from the NCBI and GISAID databases. The N-protein is the most abundant protein in coronavirus (Fig. 1), is encoded by a conserved gene, and is used as a molecular marker (Sino Biological, 2020).

Fig. 4 shows that SARS-CoV Tor2 2003 has the closest relationship with Bat-CoV YN2018C 2018. SARS-CoV-2 2020, which is found in various countries, is also closely related to Pangolin-CoV GX-P5E 2017 and Bat-SL-CoV ZC45 2018. The bat coronavirus (BatCoV RaTG13), previously detected in *Rhinolophus affinis* from Yunnan province, has many similarities with SARS-CoV-2, with an overall genome sequence similarity of 96.2% (Zhou et al., 2020). SARS-CoV-2 2020 is also related to MERS-CoV 2012, Human-CoV HKU1 2004, and Human-CoV OC43 2003; SARS-CoV-2 has a genetic similarity of 50% with MERS-CoV and 79% with SARS-CoV (Rabaan et al., 2020). SARS-CoV and SARS-CoV-2 genome sequences have extremely high levels of nucleotide similarity with only six genomic regions showing differences (Xu et al., 2020).

![Fig. 4. Phylogenetic tree of SARS-CoV-2 and other CoV isolates based on the nucleocapsid protein gene](image-url)
SARS-CoV-2 from in Indonesia was very closely related to SARS-CoV-2 from other countries. This demonstrated that they had very similar genomes, although there were slight genetic changes during the transmission process. The results of the study by Forster et al. (2020) show that the patient’s travel history determines patterns in the SARS-CoV-2 phylogenetic network. Different phylogeographic patterns can be influenced by different migration histories, founder events, and sample sizes. Additionally, patient treatment and increased patient immunity also affects mutation occurrence. Any decrease in COVID-19 cases and/or CFR is likely caused by increased immunity in the human population and is an epidemiological context of viral mutation (Zhang and Holmes, 2020). Monitoring mammalian viruses is very important to prevent the occurrence of virus transmission from animals to humans. Furthermore, monitoring viruses that have infected humans is important to prevent wider transmission and to identify mutations during transmission.

CONCLUSION
Of the 37 provinces in Indonesia, five provinces with the highest CFR were DKI Jakarta, Jawa Barat, Jawa Timur, and Banten, and five provinces with the highest cure rate were Kepulauan Riau, Bali, Aceh, Gorontalo, and DI Yogyakarta. The rapid and widespread distribution of SARS-CoV-2 in Indonesia was caused by the lack of compliance with territorial restrictions and dishonesty with medical personnel. SARS-CoV-2 isolates from Indonesia were closely related to SARS-CoV-2 isolates from other countries. These data revealed that mutations can occur during the transmission process, which can be caused by a history of travel and increased patient immunity.

ACKNOWLEDGMENTS
This study supported by the Directorate General of Higher Education, Ministry of Education and Culture of the Republic of Indonesia.

CONFLICT OF INTEREST
The authors declares that there is no conflict of interest.


